Psychological Test and Assessment Modeling, Volume 63, 2021 (1), 13–44

Cognitive Components of Writing in a Second Language: An Analysis with the Linear Logistic Test Model

Farshad Effatpanah[1] & *Purya Baghaei*[2]

Abstract

Writing in a second/foreign language (L2) is a demanding task for L2 writers because it calls for multiple language abilities and (meta)cognitive knowledge. Research investigating the (meta)cognitive processes involved in composing in L2 have emphasized the complex and multidimensional nature of L2 writing with many underlying (meta)cognitive components. However, it is still unclear what factors or components are involved in composing in L2. Employing correlational and qualitative approaches and through the modeling of L2 writing proficiency, previous studies could not offer adequate evidence for the exact nature of such components. This study aimed at examining the underlying cognitive operations of L2 writing performance using an IRT-based cognitive processing model known as linear logistic test model (LLTM). To achieve this, the performance of 500 English as a foreign language (EFL) students on a writing task was analyzed. Five cognitive processes underlying L2 writing were postulated: content fulfillment, organizational effectiveness, grammatical knowledge, vocabulary use, and mechanics. The results of the likelihood ratio test showed that the Rasch model fits significantly better than the LLTM. The correlation coefficient between LLTM and Rasch model item parameters was .85 indicating that about 72 % of variance in item difficulties can be explained by the five postulated cognitive operations. LLTM analyses also revealed that vocabulary and content are the most difficult processes to use and grammar is the easiest. More importantly, the results showed that it is possible to envisage a model for L2 writing with reference to a set of subskills or attributes.

Keywords: L2 Writing Attributes, Q-matrix, Linear Logistic Test Model (LLTM)

[1]*Correspondence concerning this article should be addressed to:* Farshad Effatpanah, English Department, Islamic Azad University of Mashhad, Ostad Yusofi St., Mashhad, 91871, Iran; E-Mail: farshadefp@gmail.com

[2]English Department, Islamic Azad University, Mashhad Branch, Mashhad, Iran

Introduction

Writing is considered as an essential form of communication and meaning-making process (Zamel, 1982). Consequently, learning to write well is an important educational goal and ones need to develop a complex variety of skills in order to be able to successfully involve in a meaning-making process. Generally speaking, writing in one's native language is known to be a challenging task because it requires the coordination of several linguistic abilities and (meta)cognitive knowledge. When it comes to a second/foreign language (L2), writing becomes even more taxing. It has been well-established that L2 writing is a multifaceted process which contains many (meta)cognitive components and variables. However, the precise nature of these constituents are not still clear. In fact, it is not manifest what operations are involved in composing and to what extent they might contribute to the performance of students.

To better understand what constitutes writing ability, numerous researchers have developed different models to demonstrate factors and their interplay in the writing process. The proposed models have focused on writing processes (Alamargot & Chanquoy, 2001; Flower & Hayes, 1980, 1983; Hayes, 1996; Sasaki, 2002; Van den Bergh & Rijlaarsdam, 1996) and writing proficiency development or sub-skills (Abbott & Berninger, 1993; Bereiter & Scardamalia, 1987; Grabe & Kaplan, 1996; Schoonen, Snellings, Stevenson, & van Gelderen, 2009; Schoonen et al., 2003). Still, due to the complicated nature of writing, the evidence they provided for the existence of sub-skills is rather inconclusive and contradictory.

In addition to these models, a large number of researchers have qualitatively and quantitatively examined (meta)cognitive factors that could explain second/foreign language learners' writing ability. These factors embrace first language (L1) writing ability (Kobayashi & Rinnert, 1992; Manchón & Roca de Larios, 2007; Silva, 1993), writing strategies in L1 and L2 (Krapels, 1990; Pennington & So, 1993; Roca de Larios, Murphy, & Marin, 2002), writing experience (Cumming, 1989; Matsuda, 1997), educational background (Bereiter & Scardamalia, 1987), linguistic knowledge (Abbott & Berninger, 1993; Laufer & Nation, 1995; Ortega, 2003), metacognitive knowledge (Schoonen & De Glopper, 1996; Victori, 1999), fluency or processing speed and working memory (Bereiter & Scardamalia, 1987; Chenoweth & Hayes, 2001; Kellogg & Whiteford, 2009; Lea & Levy, 1999; McCutchen, 1996, 2000; Ransdell & Levy, 1999), and L2 language knowledge and proficiency (Cumming, 1989; Raimes, 1985; Reid, 1984, 1990). Because most of these investigations examined the construct only one factor at a time, they failed to offer a satisfactory explanation of the basic processes of writing. In order to achieve a more comprehensive understanding of the factors underlying the construct, several studies have analyzed a cluster of variables all together. In his multivariate analyses of French students' English L2 compositions, Cumming (1989) reported that writing expertise and second language proficiency distinctly account for large portions of the variance in the quality of L2 writing.

In much the same vein, Hirose and Sasaki (1994) examined explanatory variables for nineteen Japanese university students' writing ability. Quantitative analysis showed that L2 proficiency and first language writing ability significantly explicates 74.5 % of the students' L2 composition score variance. In their qualitative analysis, the results revealed that writing competence is associated with the use of good writing strategies and writing fluency. In their next study, Sasaki and Hirose (1996) investigated the performance of first-year Japanese students in terms of their L2 proficiency, first language writing ability, writing strategies in L1 and L2, metaknowledge of L2 writing, past writing experience, and educational background. Their quantitative analysis, using regression, indicated that L2 language proficiency (52 %), L1 writing ability (18 %), and metaknowledge (11 %) are the major predictors of L2 writing ability. Among these variables, unique contribution of L2 proficiency explained 32.6 % of the total score variance followed by metaknowledge and L1 writing ability with .3 % and 1.5 %, respectively.

Schoonen et al. (2003) studied the performance of Dutch secondary school students learning English as a foreign language (EFL) to identify the importance of linguistic knowledge (grammar, vocabulary, and orthography), metacognitive knowledge, and speed of language processing in both L1 and L2. The results of structural equation modeling analysis showed a substantial correlation of .93 between L1 writing ability and English L2 writing proficiency. The analysis of correlations between components of linguistic knowledge and L2 writing proficiency showed that there is a high correlation between English L2 writing proficiency and orthography, grammar, and vocabulary with values of .85, .84, and .63, respectively. It was also turned out that the speed of accessibility to linguistic knowledge is correlated with overall writing performance in both L1 and L2. However, it was found that the fluency makes no unique contribution to the prediction of writing proficiency in L1 and L2 writing performance. They concluded that L2 writing is more dependent on L2 linguistic knowledge relative to L1 writing and metacognitive knowledge.

Yun (2005), using confirmatory factor analysis (CFA), found the correlation between L1 writing (Korean) and L2 writing ability (English) to be moderate (.42). The results of structural equation modeling indicated that L2 language knowledge is the major predictor for L2 writing performance so that all the five variables (e.g., L1 writing ability, L2 language knowledge, L2 writing experience, L2 reading experience, and test preparedness) could explain 70.4 % of the variance in L2 writing performance.

Y. Lu (2010) also applied multiple regression to analyze five factors that might influence Chinese EFL learners' writing product. The factors included L2 language proficiency, L1 writing ability, genre knowledge, writing strategies, and working memory capacity. Y. Lu (2010) found that the five factors can explain about 30 % of the variation, with L2 language proficiency being the most important predictor, accounting for 20.4 % of the variance, and genre knowledge and strategy use add slightly 7.2 % and 2.1 % variance, respectively.

Schoonen, van Gelderen, Stoel, Hulstijn, and de Glopper (2011) conducted a longitudinal study to model the development of EFL writing proficiency of Dutch secondary school students. Results of structural equation modeling showed that compared to L1 writing proficiency, EFL writing ability has a strong relationship with linguistic knowledge and linguistic fluency. They found that students' EFL writing ability improved to a great extent rather than L1 writing proficiency after two years. Modeling the writing proficiency of L1 and EFL also showed a high correlation between L1 and EFL writing proficiency, with metacognitive knowledge and general fluency intervening this interaction.

Finally, Sparks, Humbach, Patton, and Ganschow's (2011) factor analytic study identified four factors for predicting oral and written L2 proficiency: Language Analysis (e.g., L1 and L2 language comprehension, grammar, vocabulary, and inductive language learning measures), Phonology/Orthography, IQ/Memory, and Self-Perceptions of Language Skills. It turned out that the four factors explained 76 % of the variance in oral and written language proficiency. The results of regression analysis showed that Language analysis account for 12 % of the variance followed by Phonology/Orthogrpahy (10 %), Self-Perceptions of Language Skills (9 %), and Intelligence/Memory (2 %).

Based on the above-mentioned studies, it seems that the general idea among L2 writing researchers is that L2 writing ability is composed of several distinct and interconnected components so that second/foreign language-related knowledge is the foremost explanatory variable for writing proficiency. Despite of the fact that all the previous studies provided invaluable insight into the factors contributing to L2 writing proficiency, it is yet unknown what linguistic factors are involved and to what degree they might influence second/foreign language writing development (M. Kim & Crossley, 2018). Little is known about desirable linguistic knowledge resources or component skills which are of value to a successful writing performance (Schoonen et al., 2003).

These empirical studies also suffer from a major drawback with respect to the appropriacy of their methodology. Most of the studies used a qualitative or correlational approach to account for underlying writing processes. Gorin (2005) argued that the problem with correlational approach is that a high correlation between item components (i.e., difficulty, discrimination, reaction time) and cognitive operations does not necessarily explicate a causal relationship. Furthermore, correlational methods are extremely sensitive to the variance of item difficulties within the sample. Sonnleitner (2008) demonstrated that a pool of very difficult items would result in low correlations with cognitive processes, whereas items with a wide range of difficulties would lead to high correlations. Respecting items of nearly equal difficulties, the cognitive processes in question would account for the small differences in item difficulties. Multiple regression studies have also received considerable debate over their analysis and interpretation. These studies are very liable to collinearity and suppression between some of the predictor variables. They also offer only information about group performance because regression analysis involves mean performances on the dependent variables; it cannot ascertain the difficulty or easiness of the variables. Particularly relevant to this study is applying item scores to predict item difficulty on the basis of the attributes required to give a correct response to the items (Buck, Tatsuoka, & Kostin, 1997) in these studies. Buck et al. (1997) indicated that this practice is limited in that there is a disparity between predictors and items, that is, there are too many predictors with too few items coded for each subskill/process. Factor analytic studies, on the other hand, though very robust statistical methods, have provided unclear results on the nature of the factor structure and the number of attributes involved in the performance. Weir and Khalifa (2008) argued that the factorial approach cannot appropriately specify the actual processes or subskills assumed to be involved in answering a set of given items or tasks. They only focus on the divisibility of the components based on their factor loadings on items. Therefore, with regard to the methodological limitations of correlational approaches, it is evident that more powerful statistical methods are required to identify cognitive operations underlying a given skill. Item response theory (IRT)-based cognitive processing models could be effective alternatives which avoid these problems. These models have the potential to model cognitive complexity and estimate item difficulty parameters in a non-correlative manner.

IRT-based cognitive processing models such as Linear Logistic Test Model (LLTM; Fischer, 1973), component latent trait model (CLTM; Embretson (Whitely), 1984), and Cognitive Diagnostic Models (CDMs) are able to identify, model, and parameterize different mental steps or cognitive processes that underlie task performance (Sonnleitner, 2008). An important characteristic of such models is that they decompose items or tasks into various strategies, processes, and knowledge required to accomplish a task or item. The idea behind this dates back to Atkinson and Paulson (1972) who emphasized the necessity of dividing learning concepts into smaller units or quantitative difficulty parameters, known as learning quanta, which can help teachers to replace students' faulty strategies and design remedial instructional materials for an individual learner. A psychometric model which has been developed with respect to these properties is Fischer's (1973) linear logistic test model (LLTM) which can provide a sensible basis for explaining cognitive operations involved in completing a set of test items or tasks. Although the model can provide more precise information for realizing cognitive components underlying a skill, its application has been disregarded in second/foreign language contexts. Only handful studies have applied the LLTM to empirically examine sources of item difficulty in L2 educational contexts (Baghaei & Ravand, 2015; Embretson & Wetzel, 1987; Ghahramanlou, Zohoorian, & Baghaei, 2017; Gorin, 2005; Sonnleitner, 2008).

Literature Review

The Linear Logistic Test Model

The linear logistic test model (LLTM; Fischer, 1973) is an extension of the Rasch model (RM) (Rasch, 1960/1980) which assumes that the item parameters $\beta_i (i = 1, 2, \ldots, k)$ of the RM can be decomposed into a linear combination of certain basic parameters η_j. In fact, item difficulty is postulated to involve certain hypothesized elementary parameters that have to be taken to accomplish a task or test item, each of which has a difficulty parameter (Fischer, 1995). For the standard Rasch model, the probability that person v

gives a correct response to item i, given his/her ability θ_v and the item difficulty β_i is defined by:

$$P(x_{vi} = 1 \mid \theta_v, \beta_i) = \frac{\exp(\theta_v - \beta_i)}{1 + \exp(\theta_v - \beta_i)} \tag{1}$$

The linear decomposition of the item parameters into the basic parameters which is imposed by the LLTM is expressed as:

$$\beta_i = \sum_{j}^{p} q_{ij}\eta_j + c \tag{2}$$

where β_i is the difficulty parameter of item $i(i = 1, 2, \dots, n)$, q_{ij} is the weight of the hypothesized cognitive operation or basic parameter j on item i, η_j is the estimated difficulty of the cognitive operation $j(j = 1, 2, \dots, m)$, c is a normalization constant, which can easily be eliminated. It is defined as

$$c = \frac{-\sum_{i=1}^{n}\sum_{j=1}^{m} q_{ij}\eta_j}{n} \tag{3}$$

The decomposition of item difficulty is an advantage for the LLTM because only p parameters have to be estimated instead of k parameters, which results in optimal gain from the data's information (Kubinger, Reif, & Yanagida, 2011).

Inserting Equation (2) into Equation (1) gives the model equation of the LLTM:

$$P(X_{vi} = 1 \mid \theta_v, \boldsymbol{\eta}, \boldsymbol{q}) = \frac{e^{\theta_v - \sum_j w_{ij}\eta_j}}{1 + e^{\theta_v - \sum_j w_{ij}\eta_j}} \tag{4}$$

To assess the fit of the LLTM or the validity of the hypothesized model, there are two common approaches mainly employed by researchers. As the LLTM is a linear constrain of the RM, the two models are considered as nested models and the likelihood ratio test (LRT) can be used to compare the fit of the LLTM against the RM. Let $L^{(RM)}$ be the maximum likelihood function of the Rasch model and $L^{(LLTM)}$ be the maximum of the conditional likelihood function of the LLTM. In this case:

$$X^2 = -2\ln\frac{L^{(LLTM)}}{L^{(RM)}} \tag{5}$$

is asymptotically χ^2-distributed with $k - p$ degrees of freedom $(df = k - p)$. If the null hypothesis (e.g., there is no difference in data's likelihood for the models) is confirmed, it can be concluded that the LLTM and the RM can describe a given data equally well. The problem with this method is that LRT mostly turns out to be significant and thus LLTM is rejected (Fischer & Formann, 1982) and the usefulness of the hypothesized model for explaining item difficulties becomes dubious. Fischer and Formann (1982) argued that reaching a good fit for the model is difficult or rather, it is attained only when a test is studiously constructed with an eye to a cognitive model. Alternatively, Alexandrowicz (2011) suggested that in cases where the LLTM does not describe the data equally well as the RM, the person parameters of the RM and the LLTM virtually provide the same precise information regarding the persons tested. In fact, model deviations have not greatly affected the person parameters and individuals can be assessed with sufficient preciseness.

The second approach for checking the fit of the model is to examine the correlation between the Rasch model item parameters and the difficulty parameters reproduced by the LLTM. A high correlation between the Rasch model and LLTM parameter estimates can be generally considered as an indicator of a sufficient explanatory value of the model. In other words, one can infer the validity of the cognitive model, defined in terms of the hypothesized basic parameters in the Q-matrix, to the extent that, except for random error, the Rasch model and LLTM produce the same item parameter results (Baghaei & Hohensinn, 2017). In a simulation study, Baghaei and Hohensinn (2017) proposed a method for evaluating the weight matrix in the LLTM. They suggested a kind of parallel analysis for Q-matrix validation in LLTM. They simulated 1000 random Q-matrixes with the same specifications as their empirical Q-matrix. They argued that if the empirical Q-matrix is correctly specified, the correlation between the RM item parameters and those recovered by the LLTM using the empirical weight matrix should be higher than the correlation yielded by simulated weight matrices. Their findings showed that the mean of correlations from completely fake simulated matrixes is around .50 and their 95 % percentile is .76. Therefore, they concluded that a correctly designed Q-matrix should produce a correlation coefficient of at least .76, that is, the correlation coefficient of .76 can be set as a lower bound for the plausibility of the cognitive model to explain variance in item parameters.

Before wrapping up this section, it should be pointed out that although LLTM is an extension of the Rasch model, it maintains all the important assumptions of the Rasch model such as unidimensionality and parameter separability. One well-known property of the Rasch model is parameter separation through which item parameters can be estimated without estimating person parameters and person parameters can be estimated without estimating item parameters. This assumption allows for using the conditional maximum likelihood estimation (CML), as a means of providing specific objective parameter estimates (Scheiblechner, 2009). The other critical assumption or requirement of the Rasch model is unidimensionality. This assumption implies that a set of items forming an instrument should measure one construct or dimension at a time. However, this assumption is very difficult to satisfy in educational tests because performance on any language task entails a variety of mental processes. As Bejar (1983) explained, "unidimensionality does not imply that performance on items is due to a single psychological process. In fact, a variety of psychological processes are involved in responding to a set of test items. However, as long as they are involved in unison-that is, performance on each item is affected by the same process and in the same form-unidimensionality will hold" (p. 31). Therefore, although cognitive operations in the LLTM are regarded as distinct features, they are interconnected and complementary, and one requires to use them simultaneously to answer given items or tasks.

Review of the Related Literature on the LLTM

To date, the linear logistic test model (LLTM) has been primarily employed to either identify construct-related processes which focus on explaining postulated cognitive operations involved in solving/processing test items (Adroher & Tennant, 2019; Baghaei & Ravand, 2015; Embretson & Wetzel, 1987; Fischer, 1973; Gorin, 2005; Gorin & Embretson, 2006; Kubinger, 1979, 1980; Kubinger, Hohensinn, Holocher-Ertl, & Heuberger, 2011; Poinstingl, 2009; Sheehan & Mislevy, 1990; Sonnleitner, 2008; Whitely & Schneider, 1981; Zeuch, Holling, & Kuhn, 2011) or detect construct-irrelevant processes such as content-specific learning, item position effects, speeded item presentation, and item response format (Hohensinn & Baghaei, 2017; Hohensinn & Kubinger, 2009; Hohensinn et al., 2008; Hohensinn, Kubinger, Reif, Schleicher, & Khorramdel, 2011; Kubinger, 2008, 2009; Kubinger, Reif, & Yanagida, 2011) which might affect item difficulty. Unlike the vast practical applications of the latent component cognitive processing models such as the LLTM, very little attention has been devoted to the utility of the LLTM in second/foreign language contexts. Very few studies have been sporadically applied the model to explicate cognitive operations that underlie language skills. Embretson and Wetzel (1987) used the LLTM to develop a processing model to explain sources of difficulty in multiple-choice paragraph comprehension items. Their model suggested that two independent processes are involved in solving a reading comprehension item, namely, a text representation process and a response decisions process. Their analysis revealed that decision processes have more substantial impact on item difficulty than text representation processes.

Gorin (2005) examined the extent to which variations in item difficulty of reading comprehension items can be changed by experimentally manipulating certain item characteristics such as propositional density and syntax, passive voice and negative wording, order of information, and response alternatives. Results of LLTM estimation showed that manipulation of negative wording considerably increases item difficulty in some cases; however, other factors, e.g., propositional density, information order, and response alternatives, did not significantly affect item difficulty, but they had effects on reaction time.

In a similar study to Embretson and Wetzel (1987), Gorin and Embretson (2006) investigated generative components of the Graduate Record Examination (GRE) paragraph comprehension items. The results of their study indicated that decision variables contribute to item difficulty greatly more than text representation variables. Specifically, the vocabulary level and reasoning level of the correct responses affect remarkably processing difficulty.

Sonnleitner (2008) analyzed the effect of input- and response-related factors on item difficulties of a German reading comprehension test with LLTM. Results of his study showed that the two components explain item difficulties with a greater effect of response-related radicals on the difficulty of items. Similar to Embretson and Wetzel (1987) and Gorin and Embretson (2006), he concluded that both input- and response-related components should be considered when modeling complexity of reading comprehension items.

Baghaei and Ravand (2015) applied the LLTM to a high-stakes reading comprehension test to investigate underlying cognitive components and processes of foreign language reading comprehension. They derived five attributes underlying reading performance: reading for details, making inferences, reading for main idea, syntax, and vocabulary. The analysis of LLTM showed that making inferences is the most difficult process and vocabulary is the easiest. They sustained that the poorer fit of the LLTM compared to the Rasch model in their study may be due to the presence of more construct-relevant processes and construct-irrelevant processes such as response decision.

More recently, Ghahramanlou et al. (2017) examined the cognitive processes underlying the listening section of the International English Language Testing System (IELTS) using the LLTM. Six cognitive operations were postulated to be involved in underlying the test. The likelihood ratio test showed a poorer fit of the LLTM compared to the Rasch model. The correlation between the LLTM-reconstructed item parameters and the RM-based item parameters was .85, indicating that the six basic parameters can explain 72 % of the variance in item difficulty of the Rasch model. They also found that keeping up with the pace of the speaker and understanding reduced forms are the most difficult operations for the candidates of the IELTS. These studies clearly witness that most of the LLTM application has been on receptive skills, in particular on reading, and there is scant research on the use of the LLTM for explaining cognitive operations that underlie L2 writing performance. With this need in mind, the present study attempts to offer an explanation of the cognitive processes underlying L2 writing ability using LLTM.

Method

Data

The data analyzed in this study was taken from Effatpanah, Baghaei, and Boori (2019). In a recent study, they applied the additive cognitive diagnostic model (ACDM) to diagnose English as a foreign language (EFL) students' L2 writing ability. The data includes the performance of 500 Iranian EFL students on a writing task. The task required a 60-min timed composition in which participants were asked to write at least a 350-word essay in their L2 writing courses in response to the following prompt:

> How to be a first-year student in college? Write about the experience you have had. Make a guide for students who might be in a similar situation. *Describe* how to make new friends, how to escape homesickness, how to be successful in studying, etc.

The partcipants were Junior (42.4 %), Senior (30.4 %), and (27.2 %) postgraduate students. There were 349 female (69.8 %) and 151 male (30.2 %) students between the ages of 19 and 58 years old (M = 24.89 years, SD = 6.30). As the research involved human participants, the study was reviewed and approved by the review board of the Islamic Azad University of Mashhad with a written consent (n. 62594) prior to beginning the study. The participants provided their written informed consent to take part in this study and were reassured that their information would remain confidential and anonymous.

The written texts were marked by three experienced college-level teachers along with the first author of the paper. The raters first served as content experts in the Q-matrix construction stage to stipulate the conceptual relationship between each descriptor and its sub-skills and later as raters to rate the essays. The sample of raters included two females and two males between the ages of 25 to 52 years old. They were all non-native speakers of English, knowing Persian as their first language and English as their foreign language. The three raters were full-time EFL teachers in the university and held a Ph.D. degree in English language teaching (ELT). They had a mean of 14/66 years of experience in teaching and assessing L2 writing. The other rater had acquaired an M.A. in Teaching English as a Foreign Language (TEFL) and had adequate competency in L2 writing. They used an adaptation of a diagnostic assessment scale called the Empirically-derived Descriptor-based Diagnostic (EDD) checklist (Y.-H. Kim, 2011). The EDD checklist was designed to assess the writing of non-native English-speaking students in an academic context. It comprises 35 dichotomous (Yes, No) descriptors or items assessing five writing attributes, namely, content fulfilment (CON), organizational effectiveness (ORG), grammatical knowledge (GRM), vocabulary use (VOC), and mechanics (MCH). Each sub-skill is defined as

1. Content fulfillment assesses a writer's ability to address a given question by presenting unity and relevance of supporting sentences, information, and examples;
2. Organizational effectiveness assesses a student's ability to develop and organize ideas and supporting sentences cohesively and coherently within and between paragraphs;
3. Grammatical knowledge assesses a student's ability to demonstrate syntactic variety and complexity accurately.
4. Vocabulary use assesses a student's ability to use a wide range of lexical items accurately and appropriately.
5. Mechanics assesses a student's ability to follow the conventions of English writing such as margins and indentation, punctuation, spelling, and capitalization.

As the EDD checklist was initially developed based on the independent essay section of Test of English as a Foreign Language^{TM} Internet-based Test (TOEFL iBT), Ravand, Effatpanah, and Baghaei (under review) recently modified some descriptors of the EDD checklist to adapt it for descriptive essays (See Appendix A, for more information about the adaptation of the checklist, see Ravand et al., under review).

As rater inconsistency is a common problem and considered as a major threat of construct validity (Baldwin, Fowles, & Livingston, 2005), all raters underwent a two-hour training session prior to scoring the essays. During the session, the interpretation of each descriptor was meticulously reviewed and discussed. Researchers have shown that training allows raters to interact, pose their questions, review various facets of writing prompt and scoring scheme, and get feedback on their scoring. Following the scoring guidelines of Weigle (2007), the judges were provided with a set of essays to try out the rating scale in order to acquaint with the scale and exemplify certain features of the scheme. After training, the scripts were randomly packaged and assigned to the raters. Each rater receive 125 essays and copies of the adapted checklist. Thirty five common essays were inserted in each package to be rated by all the four judges. The adapted checklist had a good internal consistency with a value of .88 Cronbach alpha (α) coefficient. To check agreement between the judges, Cohen's kappa was estimated. The value of Kappa measure of agreement was .62 indicating a substantial agreement. According to Peat (2001), values ≤ 0 show no agreement, .01–.20 none to slight, .21–.40 fair, .41–.60 moderate, .61–.80 substantial, and .81–1.00 almost perfect agreement. The Pearson correlation was also computed to specify the magnitude of inter-rater reliability. The value of .82 was obtained indicating an acceptable agreement across the two markings.

Q-matrix Construction

A central component in the linear logistic test model is a weight or Q-matrix, in which each row corresponds to an item and each column corresponds to an attribute. The Q-matrix specifies which operation contributes to which item and thus is an embodiment of substantive theory. In a test that targets J operations, each of the I items requires a set of relevant operations to be responded correctly. These specific item-attribute associations are collected into a $J \times I$ matrix, Q $= \{q_{ji}\}$, where $j = 1, \ldots, J$ and $i = 1, \ldots, I$. The matrix indicates whether or not the ith item requires the jth attribute. Baker (1993) demonstrated that a small degree of misspecification in the Q-matrix can lead to faulty estimates of the basic parameters. Therefore, correct identification of attributes underlying performance and their associations with test items improves the quality of information obtained from a cognitive processing model. Various methods have been suggested to define attributes involved in a test including test specifications, theories of content domain, content analysis of the test items, think-aloud protocol analysis, and eye-tracking (Embretson, 1991). To define the attributes that writers should possess in order to perform successfully on each descriptor (or item), in the present study, the raters and the researchers, as domain experts, mutually specified the conceptual relationships between the descriptors and the attributes. Before coding the descriptors, the raters were firstly trained how to interpret and code the attributes for items. All the descriptors and their attribute associations were discussed one by one and the reseachers mediated discussions wherever a disagreement was observed between the coders. An initial Q-matrix for the 35-item adapted checklist was constructed by achieving a consensus between the coders. Appendix B presents the initial Q-matrix. The Q-matrix contains five sub-skills which are assumed to be involved in composing in L2: content fulfillment (CON), organizational effectiveness (ORG), grammatical knowledge (GRM), vocabulary use (VOC), and mechanics (MCH). Each descriptor appeared to measure either one or two L2 writing attributes. As shown in the Q-matrix, the number 1 indicates that the skill is required for a given descriptor, whereas 0 indicates that the skill is not necessary for the descriptor. For instance, a successful performance on descriptor 2 (The essay is written clearly enough to be read without having to guess what the writer is trying to say) requires the mastery of ORG and GRM. Based on the Q-matrix, CON is associated with nine descriptors or items, ORG with twelve items, GRM with fifteen items, VOC with five items, and MCH with seven items.

LLTM Analysis

The package *eRm* version 1.0-1 (Mair, Hatzinger, & Maier, 2020) in the R statistical software (R Core Team, 2013) was used to estimate the parameters of the RM and the LLTM. An important requirement of LLTM estimation is that the superior model, e.g., the Rasch model, must first hold for the data (Fischer, 1995). If the RM does not fit the data at least approximately, it makes no sense to break down the item parameter (β) because then the basic parameter and its estimator would lack an empirical meaning (Fischer, 2005). To check the fit of the Rasch model, the stringent Andersen's (1973) likelihood ratio (LR) test with the mean and median of raw scores as the partitioning criteria were analyzed. The Andersen's LR test divides the data sets into subsamples based on mean and median as split criteria, for example, students with high and low scores. Consistent item parameter estimates (invariance) is expected from a sample of any subgroup of population if the model holds. The results showed that the 35 descriptors of the checklist do not have adequate fit to the model, $\chi^2 = 152.393$, $df = 34$, $p < .001$ (mean), $\chi^2 = 136.094$, $df = 34$, $p < .001$ (Median). The Wald-test-based chi-square-Test was implemented to diagnose causes of the model misfit. As shown in Table 1, 14 descriptors (e.g., 4, 5 ,6, 7, 9, 10, 19, 21, 25, 26, 28, 30, 31, 34) had low p-value $(p > .05)$ and large z-statistics. The fit of the model also was graphically checked. It illustrated that the 14 descriptors are far away from the 45° line. In graphical model check, three steps are required: (1) all items are individually calibrated within two subsamples, (2) items are placed on a common scale, and (3) they are cross-plotted against each other. If the data fit the Rasch model, items should scatter around the 45° line. In order to construct a Rasch model-fitting measurement instrument (A. W. Glas & Verhelst, 1995), the misfitting items, e.g., those items which were far from the line, were deleted.

By excluding the misfit descriptors, the fit of the RM to the data was examined again. The results of Andersen's LR test showed that the 21 remaining descriptors fit the Rasch model with mean and median of raw scores as splitting criteria, $\chi^2 = 26.177$, $df = 20$, $p = .16$ (mean), $\chi^2 = 14.016$, $df = 20$, $p = .83$ (median), respectively. The Martin-Löf-test (C. A. W. Glas & Verhelst, 1995) with the mean and median as the item splitting criteria was also checked. "Median" and "mean" divide items in two groups one the basis of their items' raw scores. The results indicated the sufficient global Rasch model fit (MLöef $= 80.614$, $df = 103$, $p = .95$ (mean), MLöef $= 115.246$, $df = 109$, $p = .323$

Table 1
Wald Test Estimation

Item Difficulty	z-statistic	p-value
1	−0.187	0.851
2	−1.511	0.131
3	−1.843	0.065
4	−2.994*	0.003*
5	−2.865*	0.004*
6	−2.281*	0.023*
7	−2.488*	0.013*
8	−1.533	0.125
9	−2.217*	0.027*
10	−2.549*	0.011*
11	−0.852	0.394
12	−0.287	0.774
13	−1.127	0.260
14	−1.002	0.316
15	−0.553	0.580
16	1.337	0.181
17	−0.046	0.963
18	0.580	0.562
19	4.046*	0.000*
20	−1.320	0.187
21	2.147*	0.032*
22	0.396	0.692
23	0.305	0.761
24	−0.534	0.593
25	4.632*	0.000*
26	2.651*	0.008*
27	0.566	0.572
28	−2.263*	0.024*
29	0.967	0.334
30	3.959*	0.000*
31	3.323*	0.001*
32	0.787	0.431
33	0.296	0.767
34	1.997*	0.048*
35	0.092	0.926

Note. * denotes deleted items

(median)). Table 2 also provides the difficulty parameters of the 21 descriptors and their standard errors as well as their infit and outfit mean square statistics. According to Linacre (2002), infit mean square is an inlier-sensitive or information-weighted fit statistic which captures unexpected patterns of observations by test takers (or test items) on items (or test takers) that are roughly targeted on them. Outfit mean square, on the other hand, is more sensitive to unexpected patterns of observations by test items (or test takers) distant from test takers' ability (or items' difficulty). The ideal range for infit and outfit values is between .70 and 1.30 (Bond & Fox, 2015). As shown in Table 2, the items fit the model adequately with average infit and outfit mean squares values ranging between .81 and 1.24, suggesting that the patterns of item difficulties accord with the expectations of the Rasch measurement model and that there is no construct-irrelevant variance in the data (Baghaei, 2008).

Table 2
Item Difficulty Parameters, Standard Error, and Fit Values of the Rasch Model

Items	Estimate	SE	Outfit MNSQ	Infit MNSQ
1	0.905	0.104	0.950	0.989
2	−0.802	0.105	0.911	1.029
3	−0.461	0.101	1.014	0.979
8	0.244	0.099	1.008	0.964
11	−0.115	0.100	0.915	0.972
12	−0.282	0.100	0.863	0.938
13	0.244	0.099	0.910	0.952
14	0.973	0.105	0.874	0.919
15	0.389	0.100	0.943	0.983
16	−1.826	0.125	1.017	1.088
17	−1.085	0.109	0.911	0.920
18	−0.678	0.103	0.995	1.042
20	−0.157	0.100	0.819	0.904
22	−0.504	0.102	1.187	0.977
23	−0.836	0.105	1.242	1.061
24	−0.302	0.100	1.089	1.004
27	1.773	0.120	0.940	0.974
29	−0.043	0.099	1.067	1.064
32	0.152	0.099	1.050	1.009
33	−0.177	0.100	0.987	1.004
35	2.589	0.149	0.879	0.949

Table 3 further provides the result of re-estimation of the Wald test based on chi-square test. The descriptors have large p-value and low z-statistics indicating an ideal range. Graphical model check, as illustrated in Figure 1, also showed that the 21 remaining descriptors do not lie far from the 45° line and thus the data fit the Rasch model. Reliability coefficients of the checklist with 21 descriptors were estimated using Cronbach α Analysis and a value of .84 was revealed which is acceptable.

As the number of deleted descriptors was high for fitting the RM and it may impair the

Table 3
Wald Test Reestimation

Item	z-statistic	p-value
1	0.057	0.955
2	−0.963	0.335
3	−1.822	0.068
8	−0.184	0.854
11	0.404	0.686
12	−0.559	0.576
13	0.229	0.819
14	−0.540	0.589
15	0.023	0.981
16	1.919	0.055
17	−1.093	0.274
18	1.265	0.206
20	−0.718	0.473
22	0.252	0.801
23	−0.035	0.972
24	0.034	0.973
27	−0.570	0.569
29	1.444	0.149
32	0.290	0.772
33	0.297	0.767
35	0.279	0.780

overlap between the descriptors in terms of cognitive operations (Baghaei & Kubinger, 2015), it was essential to modify or reformulate the structure of the association between each item and its requisite attributes (Kubinger, 2008). Green and Smith (1987) noted that when misfit items are omitted, it is judicious to identify items with alternative processes or strategies to make a connection between the elements of the weight matrix. Similarly, Baker (1993) showed that increasing the number of the operations involved in each item of a test tends to reduce the impact of misspecification. In this respect, the researchers, as the content experts, inspected the Q-matrix and decided to substantively change two cells of the Q-matrix. For descriptor 14, in addition to organizational effectiveness (ORG) and mechanics (MCH), vocabulary use (VOC) was specified and grammatical use (GRM) was added to descriptor 18 to make a connection between the descriptors. Table 4 presents the final Q-matrix for estimating LLTM.

In the next step, we subjected the data to the LLTM analysis with the Q-matrix including 21 descriptors associated with five cognitive processes. Table 5 gives the difficulty of the five basic parameters (CON to MCH), their standard errors, and their 95 % confidence intervals. Cognitive components with positive difficulty parameters makes items or descriptors easier and negative difficulty parameters makes the descriptors harder.

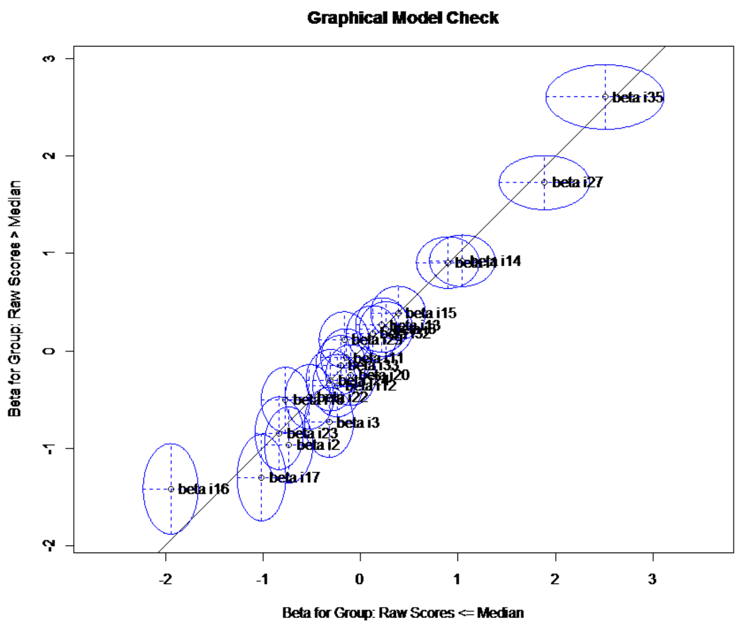

Figure 1
Graphical Model Check

Likewise, the 95 % confidence interval revealed that none of the processes deviate significantly from zero $(p < 0.05)$ (parameters whose CI do not include zero are significant), that is, an item or a descriptor does not significantly change its difficulty when these operations are contained in it (Fischer, 1973). As can be seen from Table 5, vocabulary (VOC) is the most difficult attribute and Grammar (GRM) is the easiest. Content (CON) is the second hardest attribute to use followed by organization (ORG) and mechanics (MCH).

To compare the fit of the LLTM and the RM in terms of their item parameters, the correlation coefficient between LLTM-reconstructed and RM-based difficulty parameters was computed and the value amounts to .85. This value shows that 72.25 % $(.85^2 \times 100)$ of the variance in RM item parameters can be explained by the five postulated operations. According to Baghaei and Hohensinn (2017), values above .76 are acceptable for having a meaningful weight matrix. The graphical analysis (Figure 2) of the agreement between LLTM easiness parameters and RM easiness parameters does not lie too far from the 45° line. The plot displays that there is a concurrence between the two models.

A likelihood ratio test was also employed to compare the fit of the LLTM against the RM. Results showed that just like many other studies, the resulting value of the asymptotically chi-square distributed statistic is much greater than the critical value and thus the null

Table 4
The Final Q-matrix

Item	CON	ORG	GRM	VOC	MCH
1	1	1	0	0	0
2	0	1	1	0	0
3	1	0	1	0	0
8	1	0	0	0	0
11	1	1	0	0	0
12	0	1	0	0	0
13	1	1	0	0	0
14	0	1	0	1	1
15	0	0	1	0	0
16	0	0	1	0	0
17	0	0	1	0	1
18	0	0	1	0	1
20	0	0	1	0	0
22	0	0	1	0	0
23	0	0	1	0	0
24	0	0	1	0	0
27	0	0	0	1	0
29	0	0	1	1	0
32	0	0	0	0	1
33	0	0	0	0	1
35	0	0	0	1	0

Note. CON = Content fullfiment, ORG = Organizational effectiveness, GRM = Grammatical effectiveness, VOC = Vocabulary use, MCH = Mechanics

Table 5
Basic Parameter Estimates

Basic Parameters	Difficulty	SE	Lower CI (95 %)	Upper CI (95 %)
CON	0.082	0.070	−0.058	0.222
ORG	−0.330	0.060	−0.450	−0.210
GRM	−1.035	0.063	−1.161	−0.909
VOC	1.156	0.069	1.018	1.294
MCH	−0.395	0.062	−0.519	−0.271

Note. SE: Standard Error; CI: Confidence Intervals

hypothesis is overly rejected. In other words, the RM fits significantly better than the LLTM, $\chi^2 = 24.995\,79$, $df = 15$, $p < .01$. This means that the hypothesized cognitive

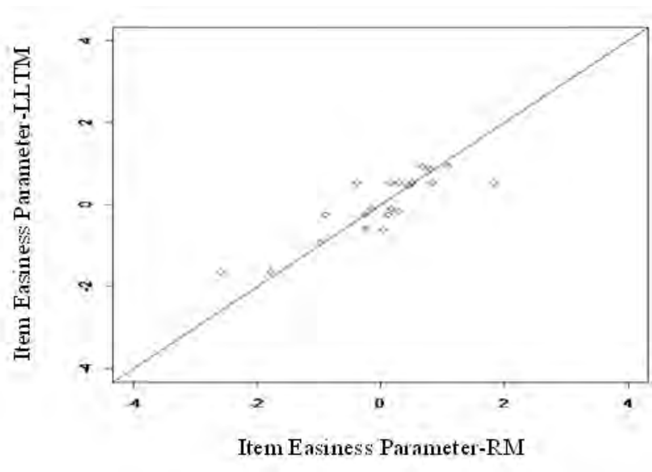

Figure 2
Graphical Comparison of RM and LLTM Parameter Estimates

model cannot account for all the item difficulties of the RM. Following Alexandrowicz's (2011) suggestion upon considering the person parameters when the LRT strongly deprecates the LLTM, we examined to what extent the models share a perfect relation. The Pearson correlation coefficient of the person parameters of two models was surprisingly perfect (1.00), higher than the value of .991 Alexandrowicz obtained in his study. It shows that the person parameter estimates of the RM and LLTM are highly analogous, e.g., the ability parameter of individuals would not change whichever of the two models is applied.

Discussion

This study was designed to demonstrate the application of the linear logistic test model (LLTM) for explaining the cognitive processes underlying L2 writing performance. Five cognitive operations were postulated to be involved in composing in L2, e.g., content fulfillment (CON), organizational effectiveness (ORG), grammatical knowledge (GRM), vocabulary use (VOC), and mechanics (MCH). To check the validity of the hypothesized theory defined in the weight matrix, the fit of the LLTM was compared with the fit of the Rasch model. In common with many previous studies on the application of the LLTM, likelihood ratio test showed a significantly better fit of the RM relative to the LLTM, indicating that the LLTM and the Rasch models do not describe the data equally well and the hypothesized model seems to fail in explaining (all) the item parameters. Fischer and Formann (1982) stated that "… such statistical significances ought not to be over-rated, because in many cases relatively large samples of data were used for testing hypotheses about only a few parameters, i.e., the tests were rather powerful; moreover, ultimately any significance criterion is arbitrary" (p. 412). Rather, the correlation between the RM item parameters and those reconstructed by the LLTM was estimated and the value of .85 was obtained, indicating that about 72 % of item difficulty variance can be satisfactorily accounted for by the five hypothesized basic parameters. The value shows that the postulated cognitive model and the construct theory reflected in the Q-matrix can account for a significant amount of variance in item difficulty. This result consolidates the view among L2 researchers believing that L2 writing is more conditional on L2 linguistic knowledge (Sasaki & Hirose, 1996; Schoonen et al., 2003; Schoonen et al., 2011). The results can also be considered as an indication of the feasibility of envisaging a model for L2 writing with regard to subskills. However, the assumption that the cognitive operations can completely explicate variance in item difficulty is very simplistic. There are indeed more complicated factors or construct-relevant processes which might influence L2 writing performance. Studies on L2 writing revealed that in addition to L2 linguistic knowledge, the complexity of text composition requires metacognitive knowledge, working memory capacity, and writing strategies (Chenoweth & Hayes, 2001; McCutchen, 2000; Roca de Larios et al., 2002; Schoonen & De Glopper, 1996). Moreover, the effect of some random person and random item variables, beyond the basic parameters should not be neglected. In the LLTM, the ability of persons (person effects) is considered random whereas the item contribution, which is decomposed into a set of stimulus features, is treated as fixed effect. Different models have been developed to extend the original version of the LLTM to analyze to what extent the inclusion of random item and person effects may influence item difficulty and response processes. Van den Noortgate, De Boeck, and Meulders (2003) and Janssen, Schepers, and Peres (2004) developed a cross-classification multilevel model and random effects LLTM (RE-LLTM), respectively, which subsume a random error which takes into account an estimate of variance in item difficulty. Rijmen, De Boeck, and Leuven (2002) also proposed the random weights linear logistic test model (RWLLTM), as individual-differences extension of the LLTM, in which some or all of the item stimulus features are regarded as having random coefficients.

The results of estimating the basic parameters showed that vocabulary and content are the most difficult processes for L2 writers to employ and the grammar is the easiest one followed by mechanics and organization. These results are in agreement with English as a second/foreign langauge (ESL/EFL) writing research which highlighted the importance of vocabulary and content as two essnetial components of high-level essays. These findings also converge with those of Y.-H. Kim (2011), Xie (2016), Effatpanah et al. (2019) who investigated the usefulness of cognitive diagnostic models (CDMs), as a kind of IRT-based cognitive processing models, in providing fine-grained information about the mastery/non-mastery of L2 writers on the L2 writing cognitive components. The studies echoed the same result that the hardest skills for writers to master in second/foreign language are vocabulary and content and the simplest skills are mechanics and grammar.

A closer examination of the estimates of the basic operations provides evidence for the presence of a hierarchy of cognitive processes in composing in L2. According to Wilson, Olinghouse, McCoach, Santangelo, and Andrada (2016) and McCutchen (2011), writing consists of lower-level (e.g., mechanics and grammar) and higher-level skills (e.g., content, vocabulary, and organization). As both vocabulary and content are higher-level skills and require more cognitively advanced operations, they made up the most difficult constructs of writing in the current study. Studies have shown that coordinating these low and high-level skills strains working memory and may affect the quality of text construction (Flower & Hayes, 1980; McCutchen, 2011). Fluent or rapid access to lower-level linguistic knowledge resources and retrieving appropriate structures can lower the cognitive processing load and leave little of writers' attention and therefore may increase the cognitive capacity for higher-level processes of writing and, in turn, the quality of writing performance (Schoonen et al., 2003; Torrance & Galbraith, 2006). To illustrate, cohesion and coherence are two important properties in writing. The former refers to the linkage of linguistic elements at the surface level that holds the text together and the latter is the connection of ideas for establishing a mental representation of the text in the mind of the reader (Halliday & Hasan, 1976). Creating cohesion and coherence in a composition requires the knowledge of sentence construction and mechanical conventions. If one is fluent and efficient in retrieving appropriate grammatical structures or sentence frames, sufficient cognitive capacity will leave for expressing ideas coherently and cohesively. Whalen and Ménard (1995) found that L2 writers plan, evaluate, and revise mostly at the lower-level compared to higher-level. In fact, L2 writers relatively frequently attend to lower-level processes and burden working memory capacity and save little or no space for higher-level processing. Benton, Kraft, Glover, and Plake (1984) also reported significant differences between less proficient and proficient writers' cognitive capacities. Less proficient writers are more occupied with lower-level processes while proficient writers are engaged with developing appropriate text structure and features beyond the sentence level (Schoonen & De Glopper, 1996). In this study, it appears that the participants might be so much involved in the lower-level processes of finding grammatical structures and sentence frames which might require too much conscious attention and take up little or no working memory capacity to take heed of higher-level processes of writing. This may be the reason why higher-level processes, e.g., content, vocabulary, and organization, were the most difficult attributes and lower-level processes, e.g., mechanics and grammar, the easiest.

Conclusion

As one of the most challenging and essential language skills, writing in a second/foreign language entails the development of several linguistic abilities as well as (meta)cognitive knowledge. Because of the multidimensional nature of writing, it is thus far not clear what components or processes are involved in completing a writing task. We believed that the use of IRT-based cognitive processing models such as linear logistic test model

(LLTM) can be effective ways for long-lasting explaining operations underlying L2 writing performance. Using the LLTM, the results of the current study showed that five postulated processes are involved in L2 text composition. Vocabulary and content were the most difficult processes and had the highest impact on item difficulty whereas grammar was the easiest process succeeded by mechanics and organization. Generally speaking, our findings underline the importance and viability of envisioning a model for writing in terms of its sub-skills. Identifying and decomposing the underlying processes of L2 writing is crucial for not only explicating the nature of writing but also helping teachers and all stakeholders to develop effective methods and materials for struggling students. If difficult and problematic areas of writing are identified, students themselves can receive sufficient and immediate feedback on their performance. As a consequence, they will be able to adopt some strategies to eliminate or remedy their weaknesses. Specification of the processes also is useful for construct validation and item construction. Messick (1989) noted that understanding substantively mental or psychological processes students use to perform successfully on a set of given test items or tasks is a core feature of construct validation. Understanding and specifying the components that underlie a particular cognitive domain and parameterize their difficulties allow language testers to realize the exact nature of the trait, measured by the item, and its underlying processes, construct theoretically sound items with a priori known item difficulties, which is critical for item banking and adaptive testing (Embretson, 1999).

Finally, as in any research endeavor, a number of potential limitations in the present study need to be considered. Regarding to the process of Q-matrix development, a potential problem is that the Q-matrix construction process is typically developed by domain experts and no standardized method of Q-matrix validation have already been developed for the LLTM. Because of that, the process of Q-matrix development is starkly subjective in nature. This might give rise to the existence of Q-matrix misspecifications that result in misinterpretation of the relative contribution of the basic parameters on the item difficulty (Baker, 1993). The original version of the LLTM does not incorporate an error term to determine any possible misspecifications or imperfect predictions in the Q-matrix. De Boeck et al. (2011) stated that the LLTM shares some characteristics similar to those of a regression model that accounts for all variance, and it is therefore almost always rejected. As mentioned earlier, various extended LLTM have been developed (Janssen et al., 2004; Rijmen et al., 2002; Van den Noortgate et al., 2003) which inserts an error term to the model. Subsequent research can employ such models to obtain a more accurate picture of the cognitive operations that underlie L2 writing performance.

Green and Smith (1987) further provided multiple debatable drawbacks of the original LLTM. First, almost no model typically includes all the features or basic parameters of task performance or item solving. The model is more likely to jeopardize the user to erroneously focus on observable components of the items rather than to focus on the processes and strategies that students use to respond to the given items/tasks. Second, the model assumes that item difficulty is an additive function of the difficulties of the components. This assumption may be unjustified because the model may not accurately represent psychological knowledge about the cognitive operations required for given items. Third, the model assumes uniform response processes or task structure for all examinees, that is, respondents adopt the same processes or strategies to accomplish a set of test items. However, this is not the case in practice. Different students may utilize different ways for giving a correct response to items. Researchers have recently developed models which address situations where respondents can use multiple processes and strategies during the test as well as misconceptions they possess (Kuo, Chen, & de la Torre, 2018; Ma & Guo, 2019).

Another area for further investigation is the application of the LLTM on different academic tasks or genres. Previous studies showed that task types have a great impact on L2 writers' performance including composing processes and language commands they adopt to successfully accomplish a given task (Grabe & Kaplan, 1996; Leki, Cumming, & Silva, 2008; X. Lu, 2011; Plakans, 2008). Thus, it would be interesting for future studies to examine what components are involved in composing in different writing tasks or genres and the extent to which they might contribute the performance of L2 writers.

Future studies also can experimentally investigate the contributions of cognitive processes to item difficulty. Mislevy (1994) suggested that in order to reach optimal fit, the model should be constructed first and then items are developed to fit the model. In this study, we tried to provide at least a list of variables or processes which are regarded as possible sources of item difficulty for L2 writing item or task construction. Overall, what is important is that IRT-based cognitive processing models, in our study the LLTM, have great promise for offering information about the mental processes or components which contribute significantly to the difficulty of items/tasks. These models have shift the attention from explaining consequent relationships to explaining performance with regard to a set of underlying processes or components (Embretson (Whitely), 1983). Consequently, more consideration should be given to developing tests and constructing items in second/foreign language contexts according to cognitive models. Such an endeavor requires the cooperation of practitioners from different fields of study (e.g., subject matter, measurement, pedagogy).

References

Abbott, R. D., & Berninger, V. W. (1993). Structural equation modeling of relationships among developmental skills and writing skills in primary- and intermediate-grade writers. *Journal of Educational Psychology*, *85*(3), 478–508. https://doi.org/10.1037/0022-0663.85.3.478

Adroher, N. D., & Tennant, A. (2019). Supporting construct validity of the Evaluation of Daily Activity Questionnaire using Linear Logistic Test Models. *Quality of Life Research*, *28*(6), 1627. https://doi.org/10.1007/s11136-019-02146-4

Alamargot, D., & Chanquoy, L. (2001). *Through the models of writing*. London: Kluwer Academics.

Alexandrowicz, R. W. (2011). Statistical and practical significance of the Likelihood Ratio Test of the linear logistic test model versus the Rasch model. *Educational Research and Evaluation*, *17*(5), 335–350. https://doi.org/10.1080/13803611.2011.630522

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika*, *38*(1), 123–140. https://doi.org/10.1007/BF02291180

Atkinson, R. C., & Paulson, J. A. (1972). An approach to the psychology of instruction. *Psychological Bulletin*, *78*(1), 49–61. https://doi.org/10.1037/h0033080

Baghaei, P. (2008). The Rasch model as a construct validation tool. *Rasch Measurement Transactions*, *22*(1), 1145–1146. Retrieved from https://www.rasch.org/rmt/rmt221a.htm

Baghaei, P., & Hohensinn, C. (2017). A method of Q-matrix validation for the linear logistic test model. *Frontiers in Psychology*, *8*, 897. https://doi.org/10.3389/fpsyg.2017.00897

Baghaei, P., & Kubinger, K. D. (2015). Linear logistic test modeling with R. *Practical Assessment, Research & Evaluation*, *20*(1), 1–11. Retrieved from http://pareonline.net/getvn.asp?v=20&n=1

Baghaei, P., & Ravand, H. (2015). A cognitive processing model of reading comprehension in English as a foreign language using the linear logistic test model. *Learning and Individual Differences*, *43*, 100–105. https://doi.org/10.1016/j.lindif.2015.09.001

Baker, F. B. (1993). Sensitivity of the Linear Logistic Test Model to Misspecification of the Weight Matrix. *Applied Psychological Measurement*, *17*(3), 201–210. https://doi.org/10.1177/014662169301700301

Baldwin, D., Fowles, M., & Livingston, S. (2005). *Guidelines for constructed-response and other performance assessments*. Princeton, NJ: Educational Testing Service.

Bejar, I. I. (1983). *Achievement testing: Recent advances*. Beverly Hills: CA Sage.

Benton, S. L., Kraft, R. G., Glover, J. A., & Plake, B. S. (1984). Cognitive capacity differences among writers. *Journal of Educational Psychology*, *76*(5), 820–834. https://doi.org/10.1037/0022-0663.76.5.820

Bereiter, C., & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, NJ: Lawrence Erlbaum.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York: Routledge.

Buck, G., Tatsuoka, K., & Kostin, I. (1997). The subskills of reading: Rule-space analysis of a multiple-choice test of second language reading comprehension. *Language Learning*, *47*(3), 423–466. https://doi.org/10.1111/0023-8333.00016

Chenoweth, N. A., & Hayes, J. R. (2001). Fluency in writing: Generating text in L1 and L2. *Written Communication*, *18*(1), 80–98. https://doi.org/10.1177/0741088301018001004

Cumming, A. (1989). Writing expertise and second-language proficiency. *Language Learning*, *39*(1), 81–135.

De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., & Partchev, I. (2011). The estimation of item response models with the lmer function from the lme4 package in R. *Journal of Statistical Software*, *39*(12), 1–28.

Effatpanah, F., Baghaei, P., & Boori, A. A. (2019). Diagnosing EFL learners' writing ability: A diagnostic classification modeling analysis. *Language Testing in Asia*, *9*(12), 1–23. https://doi.org/10.1186/s40468-019-0090-y

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika*, *56*, 495–515. https://doi.org/10.1007/BF02294487

Embretson, S. E. (1999). Generating items during testing: Psychometric issues and models. *Psychometrika*, *64*, 407–433. https://doi.org/10.1007/BF02294564

Embretson, S. E., & Wetzel, C. D. (1987). Component latent trait models for paragraph comprehension tests. *Applied Psychological Measurement*, *11*(2), 175–193. https://doi.org/10.1177/014662168701100207

Embretson (Whitely), S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, *93*(1), 179–197. https://doi.org/10.1037/0033-2909.93.1.179

Embretson (Whitely), S. E. (1984). A general latent trait model for response processes. *Psychometrika*, *49*(2), 175–186. https://doi.org/10.1007/BF02294171

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, *37*(6), 359–374. https://doi.org/10.1016/0001-6918(73)90003-6

Fischer, G. H. (1995). Linear logistic models for change. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments and applications* (pp. 157–180). New York, NY: Springer.

Fischer, G. H. (2005). Linear logistic test models. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (2nd ed., pp. 505–514). Amsterdam: Elsevier.

Fischer, G. H., & Formann, A. K. (1982). Some applications of logistic latent trait models with linear constraints on the parameters. *Applied Psychological Measurement*, *6*(4), 397–416. https://doi.org/10.1177/014662168200600403

Flower, L. S., & Hayes, J. R. (1980). The dynamics of composing: Making plans and juggling constraints. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 31–50). Hillsdale, NJ: Lawrence Erlbaum.

Flower, L. S., & Hayes, J. R. (1983). *A cognitive model of the writing process in adults* [Final Report]. ERIC Document Reproduction Service No. ED 240608. Pittsburgh, PA: Carnegie Mellon University.

Ghahramanlou, M., Zohoorian, Z., & Baghaei, P. (2017). Understanding the cognitive processes underlying performance in the IELTS listening comprehension test. *International Journal of Language Testing*, *7*(1), 62–72. Retrieved from http://ijlte-1.ov2.com/2019/07/01/vol-7-no-1

Glas, A. W., & Verhelst, N. D. (1995). Testing the Rasch model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments and applications* (pp. 69–95). New York: Springer.

Glas, C. A. W., & Verhelst, N. D. (1995). Tests of fit for polytomous Rasch models. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments and applications* (pp. 325–352). https://doi.org/10.1007/978-1-4612-4230-7_18

Gorin, J. S. (2005). Manipulating processing difficulty of reading comprehension questions: The feasibility of verbal item generation. *Journal of Educational Measurement*, *42*(4), 351–373. https://doi.org/10.1111/j.1745-3984.2005.00020.x

Gorin, J. S., & Embretson, S. E. (2006). Item difficulty modeling of paragraph comprehension items. *Applied Psychological Measurement*, *30*(5), 394–411. https://doi.org/10.1177/0146621606288554

Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing: An applied linguistic perspective*. London: Longman.

Green, K. E., & Smith, R. M. (1987). A comparison of two methods of decomposing item difficulties. *Journal of Educational and Behavioral Statistics*, *12*(4), 369–381. https://doi.org/10.2307/1165055

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in english*. London: Longman.

Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 1–27). Mahwah, NJ: Lawrence Erlbaum.

Hirose, K., & Sasaki, M. (1994). Explanatory variables for Japanese students' expository writing in English: An exploratory study. *Journal of Second Language Writing*, *3*(3), 203–229. https://doi.org/10.1016/1060-3743(94)90017-5

Hohensinn, C., & Baghaei, P. (2017). Does the position of response options in multiple-choice tests matter? *Psicológica*, *38*(1), 93–109. Retrieved from https://www.uv.es/psicologica/paraARCHIVES/2017.html

Hohensinn, C., & Kubinger, K. D. (2009). On varying item difficulty by changing the response format for a mathematical competence test. *Austrian Journal of Statistics*, *38*(4), 231–239. https://doi.org/10.17713/ajs.v38i4.276

Hohensinn, C., Kubinger, K. D., Reif, M., Holocher-Ertl, S., Khorramdel, L., & Frebort, M. (2008). Examining item-position effects in large-scale assessment using the Linear Logistic Test Model. *Psychology Science Quarterly*, *50*(3), 391–402. Retrieved from https://doaj.org/article/e5b6c958e97e48308ad3c6b15cd29a26

Hohensinn, C., Kubinger, K. D., Reif, M., Schleicher, E., & Khorramdel, L. (2011). Analysing item position effects due to test booklet design within large-scale assessment. *Educational Research and Evaluation*, *17*(6), 497–509. https://doi.org/10.1080/13803611.2011.632668

Janssen, R., Schepers, J., & Peres, D. (2004). Models with item and item group predictors. In P. de Boeck & M. Wilson (Eds.), *Explanatory item response models: A generalized linear and nonlinear approach* (pp. 189–210). New York: Springer.

Kellogg, R. T., & Whiteford, A. P. (2009). Training advanced writing skills: The case for deliberative practice. *Educational Psychologist*, *44*(4), 250–266. https://doi.org/10.1080/00461520903213600

Kim, M., & Crossley, A. S. (2018). Modeling second language writing quality: A structural equation investigation of lexical, syntactic, and cohesive features in source-based and independent writing. *Assessing Writing*, *37*, 39–56. https://doi.org/10.1016/j.asw.2018.03.002

Kim, Y.-H. (2011). Diagnosing EAP writing ability using the Reduced Reparameterized Unified Model. *Language Testing*, *28*(4), 509–541. https://doi.org/10.1177/0265532211400860

Kobayashi, H., & Rinnert, C. (1992). Effects of first language on second language writing: Translation versus direct composition. *Language Learning*, *42*(2), 83–209. https://doi.org/10.1111/j.1467-1770.1992.tb00707.x

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.

Kubinger, K. D. (1979). Das Problemlöseverhalten bei der statistischen Auswertung psychologischer Experimente. Ein Beispiel hochschuldidaktischer Forschung [Students' problem solving behavior at statistitical tasks. An example of didactics research in higher education]. *Zeitschrift für Experimentelle und Angewandte Psychologie*, *26*, 467–495.

Kubinger, K. D. (1980). Die Bestimmung der Effektivitat universitärer Lehre unter Verwendung des Linearen Logistischen Testmodells von Fischer. Neue Ergebnisse [Evaluation of university teaching's efficiency using Fischer's linear logistic test model—Some new results]. *Archiv für Psychologie*, *133*, 69–79.

Kubinger, K. D. (2008). On the revival of the Rasch model-based LLTM: From constructing tests using item generating rules to measuring item administration effects. *Psychology Science Quarterly*, *50*, 311–327.

Kubinger, K. D. (2009). Applications of the Linear Logistic Test Model in psychometric research. *Educational and Psychological Measurement*, *69*, 232–244.

Kubinger, K. D., Hohensinn, C., Holocher-Ertl, S., & Heuberger, N. (2011). Applying the LLTM for the determination of children's cognitive age-acceleration function. *Psychological Test and Assessment Modeling*, *53*, 183–191.

Kubinger, K. D., Reif, M., & Yanagida, T. (2011). Branched adaptive testing with a Rasch-model-calibrated test: Analysing item presentation's sequence effects using the Rasch-model-based LLTM. *Educational Research and Evaluation*, *17*(5), 373–385. https://doi.org/10.1080/13803611.2011.630549

Kuo, B. C., Chen, C. H., & de la Torre, J. (2018). A cognitive diagnosis model for identifying coexisting skills and misconceptions. *Applied Psychological Measurement*, *42*(3), 179–91. https://doi.org/10.1177/0146621617722791

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

Lea, J., & Levy, C. M. (1999). Working memory as a resource in the writing proves. In M. Torrance & G. C. Jeffery (Eds.), *The cognitive demands of writing: Pro-*

cessing capacity and working memory in text production (pp. 63–82). Amsterdam: Amsterdam University Press.

Leki, I., Cumming, A., & Silva, T. (2008). *A synthesis of research on second language writing in English*. London: Routledge.

Linacre, M. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, *16*(2). Retrieved from https://www.rasch.org/rmt/rmt162f.htm

Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL Quarterly*, *45*(1), 36–62. https://doi.org/10.5054/tq.2011.240859

Lu, Y. (2010). *Cognitive factors contributing to Chinese EFL learners' L2 writing performance in timed essay writing* [Unpublished doctoral dissertation] (Georgia State University). Retrieved from https://scholarworks.gsu.edu/alesl_diss/13

Ma, W., & Guo, W. (2019). Cognitive diagnosis models for multiple strategies. *British Journal of Mathematical and Statistical Psychology*, *72*(2), 370–92. https://doi.org/10.1111/bmsp.12155

Mair, P., Hatzinger, R., & Maier, M. J. (2020). *eRm: Extended Rasch Modeling*. 1.0-1. Retrieved from https://cran.r-project.org/package=eRm

Manchón, R. M., & Roca de Larios, J. (2007). On the temporal nature of planning in L1 and L2 composing. *Language Learning*, *57*(4), 549–593. https://doi.org/10.1111/j.1467-9922.2007.00428.x

Matsuda, P. (1997). Contrastive rhetoric in context: A dynamic model of L2 writing. *Journal of Second Language Writing*, *6*(1), 45–60. https://doi.org/10.1016/S1060-3743(97)90005-9

McCutchen, D. (1996). A capacity theory of writing: Working memory in composition. *Educational Psychology Review*, *8*(3), 299–325. https://doi.org/10.1007/BF01464076

McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational Psychologist*, *35*(1), 13–23. https://doi.org/10.1207/S15326985EP3501_3

McCutchen, D. (2011). From novice to expert: Language and memory processes in the development of writing skill. *Journal of Writing Research*, *3*(1), 51–68. https://doi.org/10.17239/jowr-2011.03.01.3

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 1–103). New York: American Council on Education/Macmillan.

Mislevy, R. J. (1994). Evidence and inference in educational assessment. *Psychometrika*, *59*(4), 439–438. https://doi.org/10.1007/BF02294388

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics*, *24*(4), 492–518. https://doi.org/10.1093/applin/24.4.492

Peat, J. (2001). *Health science research: A handbook of quantitative methods*. Sydney: Allen & Unwin.

Pennington, M. C., & So, S. (1993). Comparing writing process and product across two languages: A study of six Singaporean university student writers. *Journal of Second Language Writing*, *2*(1), 41–63. https://doi.org/10.1016/1060-3743(93)90005-N

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

Poinstingl, H. (2009). The Linear Logistic Test Model (LLTM) as the methodological foundation of item generating rules for a new verbal reasoning test. *Psychology Science Quarterly*, *51*(2), 123–134. Retrieved from https://doaj.org/article/a6f81f85aa8046a19debc96f889b0752

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Raimes, A. (1985). What unskilled ESL students do as they write: A classroom study of composing. *TESOL Quarterly*, *19*(2), 229–258. https://doi.org/10.2307/3586828

Ransdell, S., & Levy, C. M. (1999). Writing, reading, and speaking memory spans and the importance of resource flexibility. In M. Torrance & G. C. Jeffery (Eds.), *The cognitive demands of writing: Processing, capacity and working memory in text production* (pp. 99–113). Amsterdam: Amsterdam University Press.

Rasch, G. (1960/1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

Ravand, H., Effatpanah, F., & Baghaei, P. (under review). Looking for a cognitive diagnostic model for foreign language writing ability.

Reid, J. (1984). Comments on Vivian Zamel's "The composing processes of advanced ESL students: Six case studies". *TESOL Quarterly*, *18*(1), 149–153. https://doi.org/10.2307/3586344

Reid, J. (1990). Responding to different topic types: A quantitative analysis from a contrastive rhetoric perspective. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 191–210). Cambridge: Cambridge University Press.

Rijmen, F., De Boeck, P., & Leuven, K. U. (2002). The random weights linear logistic test model. *Applied Psychological Measurement*, *26*(3), 271–285. https://doi.org/10.1177/0146621602026003003

Roca de Larios, J., Murphy, L., & Marin, J. (2002). A critical examination of L2 writing process research. In S. Ransdell & M.-L. Barbier (Eds.), *New directions for research in L2 writing* (pp. 11–47). Dordrecht, the Netherlands: Kluwer Academic Publishers.

Sasaki, M. (2002). Building an empirically-based model of EFL learners' writing processes. In S. Ransdell & M.-L. Barbier (Eds.), *New directions for research in L2 writing* (pp. 49–80). Dordrecht: Kluwer Academic Publishers.

Sasaki, M., & Hirose, K. (1996). Explanatory variables for EFL students' expository writing. *Language Learning*, *46*(1), 137–174. https://doi.org/10.1111/j.1467-1770.1996.tb00643.x

Scheiblechner, H. H. (2009). Rasch and pseudo-Rasch models: Suitableness for practical test applications. *Psychology Science Quarterly*, *51*, 181–194.

Schoonen, R., & De Glopper, K. (1996). Writing performance and knowledge about writing. In G. Rijlaarsdam, H. van den Bergh, & M. Couzijn (Eds.), *Theories, models and methodology in writing research* (pp. 87–107). Amsterdam: Amsterdam University Press.

Schoonen, R., Snellings, P., Stevenson, M., & van Gelderen, A. (2009). Towards a blueprint of the foreign language writer: The linguistic and cognitive demands of foreign language writing. In R. M. Manchón (Ed.), *Writing in foreign language: Contexts learning, teaching, and research* (pp. 49–76). Clevedon, UK: Multilingual Matters.

Schoonen, R., van Gelderen, A., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning*, *53*(1), 165–202. https://doi.org/10.1111/1467-9922.00213

Schoonen, R., van Gelderen, A., Stoel, R. D., Hulstijn, J., & de Glopper, K. (2011). Modeling the development of L1 and EFL writing proficiency of secondary school students. *Language Learning*, *61*(1), 31–79. https://doi.org/10.1111/j.1467-9922.2010.00590.x

Sheehan, K., & Mislevy, R. (1990). Integrating cognitive and psychometric models to measure document literacy. *Journal of Educational Measurement*, *27*(3), 255–272. Retrieved from http://www.jstor.org/stable/1434930

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly*, *27*(4), 657–677. https://doi.org/10.2307/3587400

Sonnleitner, P. (2008). Using the LLTM to evaluate an item generating system for reading comprehension. *Psychology Science Quarterly*, *50*, 345–362.

Sparks, R. L., Humbach, N., Patton, J., & Ganschow, L. (2011). Subcomponents of second-language aptitude and second-language proficiency. *The Modern Language Journal*, *95*(2), 253–273. https://doi.org/10.1111/j.1540-4781.2011.01176.x

Torrance, M., & Galbraith, D. (2006). *The processing demands of writing* (C. A. MacArthur, S. Graham, & J. Fitzgerald, Eds.). New York: Guilford.

Van den Bergh, H., & Rijlaarsdam, G. (1996). The dynamics of composing: Modeling writing process data. In C. M. Levy & S. Ransdell (Eds.), *The science of writing: Theories, methods, individual differences and applications* (pp. 207–232). Mahwah, NJ: Erlbaum.

Van den Noortgate, W., De Boeck, P., & Meulders, M. (2003). Cross-classification multilevel logistic models in psychometrics. *Journal of Educational and Behavioral Statistics*, *28*(4), 369–386. https://doi.org/10.3102/10769986028004369

Victori, M. (1999). An analysis of writing knowledge in EFL composing: A case study of two effective and two less effective writers. *System*, *27*(4), 537–555. https://doi.org/10.1016/S0346-251X(99)00049-4

Weigle, S. C. (2007). Teaching writing teachers about assessment. *Journal of Second Language Writing*, *16*(3), 194–209. https://doi.org/10.1016/j.jslw.2007.07.004

Weir, C. J., & Khalifa, H. (2008). *A cognitive processing approach towards defining reading comprehension. Research Notes* [UCLES 2008]. Cambridge, UK: Cambridge English Language Assessment.

Whalen, K., & Ménard, N. (1995). L1 and L2 writers' strategic and linguistic knowledge: A model of multiple-level discourse processing. *Language Learning*, *45*(3), 381–418. https://doi.org/10.1111/j.1467-1770.1995.tb00447.x

Whitely, S. E., & Schneider, L. M. (1981). Information structure for Geometric analogies: A test Theory Approach. *Applied Psychological Measurement*, *5*(3), 383–397. https://doi.org/10.1177/014662168100500312

Wilson, J., Olinghouse, N. G., McCoach, D. B., Santangelo, T., & Andrada, G. N. (2016). Comparing the accuracy of different scoring methods for identifying sixth graders at risk of failing a state writing assessment. *Assessing Writing*, *27*, 11–23. https://doi.org/10.1016/j.asw.2015.06.003

Xie, Q. (2016). Diagnosing university students' academic writing in English: Is cognitive diagnostic modelling the way forward? *Educational Psychology*, *37*(1), 26–47. https://doi.org/10.1080/01443410.2016.1202900

Yun, Y. (2005). *Factors explaining EFL learners' performance in a timed essay writing test: A structural equation modeling approach* [Unpublished doctoral dissertation] (University of Illinois at Urbana-Champaign, Urbana).

Zamel, V. (1982). Writing: The Process of Discovering Meaning. *TESOL Quarterly*, *16*(2), 195–209. https://doi.org/10.2307/3586792

Zeuch, N., Holling, H., & Kuhn, J. T. (2011). Analysis of the Latin square task with linear logistic test models. *Learning and Individual Differences*, *21*(5), 629–632. https://doi.org/10.1016/j.lindif.2011.03.004

Appendix A: The Adapted EDD Checklist

		Yes	No
1	This essay answers the question.		
2	This essay is written clearly enough to be read without having to guess what the writer is trying to say.		
3	This essay is concisely written and contains few redundant ideas or linguistic expressions.		
4	This essay provides clear and precise pictures of details using description and sensory details.		
5	The writer skillfully uses logical description with purpose.		
6	There are enough supporting ideas, details, and examples in this essay.		
7	The supporting details in this essay are appropriate and relevant.		
8	This essay includes details that appeals to all five senses (e.g., touch, smell, taste, sound, and sight).		

		Yes	No
9	The ideas are organized into paragraphs and include an introduction, a body, and a conclusion.		
10	Each body paragraph has a clear topic sentence tied to supporting sentences.		
11	Each paragraph presents one distinct and unified idea		
12	Each paragraph is connected to the rest of the essay.		
13	Ideas are developed or expanded well throughout each paragraph		
14	Transition devices are used effectively		
15	This essay demonstrates syntactic variety, including simple, compound, and complex sentence structures.		
16	This essay demonstrates an understanding of English word order.		
17	This essay contains few sentence fragments.		
18	This essay contains few run-on sentences or comma splices		
19	Grammatical or linguistic errors in this essay do not impede comprehension		
20	Verb tenses are used appropriately.		
21	There is consistent subject-verb agreement.		
22	Singular and plural nouns are used appropriately		
23	Prepositions are used appropriately.		
24	Articles are used appropriately.		
25	Pronouns agree with referents.		
26	Sophisticated or advanced vocabulary is used.		
27	A wide range of vocabulary is used.		
28	Vocabulary choices, sensory languages, and figurative languages are appropriate for conveying the intended meaning.		
29	This essay demonstrates facility with appropriate collocations.		
30	Word forms (noun, verb, adjective, adverb, etc) are used appropriately		
31	Words are spelled correctly		
32	Punctuation marks are used appropriately		
33	Capital letters are used appropriately		
34	This essay contains appropriate indentation.		
35	Appropriate tone and register are used throughout the essay		

Appendix B: The Initial Q-matrix

	Content	Organization	Grammar	Vocabulary	Mechanics
1	1	1	0	0	0
2	0	1	1	0	0
3	1	0	1	0	0
4	1	1	0	0	0
5	1	1	0	0	0
6	1	0	0	0	0
7	1	1	0	0	0
8	1	0	0	0	0
9	0	1	0	0	0
10	0	1	1	0	0
11	1	1	0	0	0
12	0	1	0	0	0
13	1	1	0	0	0
14	0	1	0	0	1
15	0	0	1	0	0
16	0	0	1	0	0
17	0	0	1	0	1
18	0	0	0	0	1
19	0	0	1	0	0
20	0	0	1	0	0
21	0	0	1	0	0
22	0	0	1	0	0
23	0	0	1	0	0
24	0	0	1	0	0
25	0	0	1	0	0
26	0	0	0	1	0
27	0	0	0	1	0
28	0	0	0	1	0
29	0	0	1	1	0
30	0	0	1	0	0
31	0	0	0	0	1
32	0	0	0	0	1
33	0	0	0	0	1
34	0	1	0	0	1
35	0	0	0	1	0