Language Related Research
E-ISSN: 2383-0816
https://lrr.modares.ac.ir
https://doi.org/10.29252/LRR.15.3.10
http://dorl.net/dor/20.1001.1.23223081.1401.0.0.361.3

Tarbiat Modares University

Vol. 15, No. 3
pp. 245-274
July & August 2024

Received: 5 December 2021
Received in revised form: 26 April 2022
Accepted: 10 June 2022

On the Relationship between Sensory Learning Styles and Reading Subskill Profiles: An Application of Fusion Model

Mona Askari[1] *& Hossein Karami*[2]

Abstract

The role of learning styles in academic performance has long been the question of many educationalists. Sensory learning styles, which categorize learners into three groups of visual, auditory, and tactile students, have been said to be likely to play parts in academic performance. The purpose of the present study is twofold. Initially, this study aimed to see what weaknesses Iranian university students have in reading comprehension task. The next step was to seek if possessing different sensory learning styles can lead to a significant difference regarding reading comprehension performance. In this study, Cognitive Diagnostic Assessment was applied to provide us with comprehensive mastery reading subskill profiles of everyone. To do so, a reading comprehension test along with a learning style questionnaire were given to 301 Iranian university students, the responses were all divided into either correct or incorrect responses, and according to examinees' questionnaire, they were categorized into three groups of visual, auditory, and tactile learners. According to the present study, Iranian university students were found to have difficulty dealing with implicitly stated information, understanding difficult vocabulary, and summarizing the textual information. Regarding the second question of this study, visual learners performed significantly better than their auditory counterparts in four skills of Basic Linguistic Knowledge, Implicitly Stated Information, Understanding Difficult Vocabulary, and Understanding Complex Text. However, no significant difference was found between auditory and tactile participants. This result reinforces the prominent role of learning styles in academic and educational settings, to develop efficient instructions and curriculums that best meet learners' needs.

Keywords: cognitive diagnostic assessment, fusion model, learning styles, reading skills

[1] Corresponding Author, M.A, English Department, Faculty of Foreign Languages and Literature, University of Tehran, Tehran, Iran.
ORCID:https://orcid.org/0000-0003-3291-7019;*Email: askarymona91@gmail.com/nsn22@txstate.edu*
[2] Assistant Professor, English Department, Faculty of Foreign Languages and Literature, University of Tehran, Tehran, Iran

1. Introduction

Over recent decades, testing and assessment in educational environments have been of paramount importance. The utmost efforts have been made by the measurement specialists to see whether students have completely learned the contents of the course or what areas of weakness they have. The appropriate response to these questions was followed by the essential instruction for students to meet their needs. However, throughout recent decades, there has been little agreement over the best means of evaluation in classrooms (Taras, 2005). For example, using summative assessment, most of teachers provide their students with a final score, indicating their level of mastery at the end of the course (Taras, 2005). Given this way of assessment, teachers do not have the chance to realize whether pupils have learned the content of the course.

Moreover, learners also fail to find out about their areas of weakness to work on them. Thus, the purpose of assessment within educational environments, which is gathering exhaustive information for decision-making (Bachman, 1990), was not met satisfactorily. Instead, formative assessment, in contrast to summative assessment, is administered in the flow of the course and leads to a better understanding of teachers and students in terms of their achievements (Black & Wiliam, 2009). It is deemed to be a proper alternative to satisfy the need to gather fine-grained information about the problems students may have throughout the course. Thus, summative evaluation is replaced by formative assessment in search of broad information about students' strengths and weaknesses to design the most appropriate instruction (Ranjbaran & Alavi, 2017).

A single score obtained at the end of the course cannot provide teachers with the meticulous information needed to identify learners' weaknesses and then modify the instruction. For a test to be efficacious for examinees' self-learning and instructors' self-reflection, more fine-grained information about learners' performance should be at hand. For instance, reading comprehension consists of various underlying subskills, such as skimming and scanning, necessary for the test takers to master, if attempting to enhance reading comprehension performance. Gaining information about each of these subskills (which subskills have been mastered by the test taker and which of them have not), can make a great contribution to identifying certain reading deficiencies and then planning for more meticulous instruction that focuses on the diagnosed needs (Lee & Sawaki, 2009a). Due to the escalating demand for such fine-grained information, cognitive diagnostic assessment (CDA) is increasingly receiving attention in language assessment (Lee & Sawaki, 2009a). Here, the term "diagnosis" refers to a) Precisely knowing b) Decision making c) Agreement upon the strengths and weaknesses of the students These three phases are followed by remedial instructions (Ruppet al., 2010). As Lee and Sawaki (2009b) argued, the important characteristic of CDA is the combination of cognitive psychology and psychometrics within a single framework to assess the skill mastery of the examinees in a particular domain. Chen and Chen (2016) also defined cognitive diagnostic models (CDM) as follows: "Cognitive Diagnostic Models are psychometric models developed mainly to assess examinees mastery of a given set of skills or attributes within a domain and they can be applied to different assessments for diagnostic purposes" (p. 218).

Regarding reading comprehension assessment, researchers have long been wondering if any other external factors influence the performance of learners and their comprehension (Rogowsky et al., 2015; Sadeghi, et al., 2012; Soemer & Schiefele, 2019; Vaughn et al., 2019). Detecting these intervening factors might contribute to the final interpretation of students' outcomes and can enhance both the reliability and validity of tests. For example, individual differences have long been under scrutiny by educationalists (Rogowsky et al., 2015; Sadeghi et al., 2012). As the focus is on educational environments, this question may arise as to how individual differences can affect the performance of students in academia.

Research shows that there is a correlation between the personality characteristics of students and the way they establish their learning styles (Sadeghi et al., 2012). According to Reid (1995), learning styles are categorized into three main groups, cognitive, sensory, and personality styles each of which has its subcategorizations. For example, sensory learning style is divided into three styles, visual, auditory, and tactile. As Gilakjani (2012) noted, visual learners mostly rely on pictures, visual images, and non-verbal cues such as body language. However, auditory learners are mostly affected by pitch, emphasis, and speed in discovering information and they are less likely to perform successfully in written tasks. Tactile learners are known as interactors with the physical world. Although some studies have been carried out to determine the effects of different learning styles on reading comprehension performance (Pfister, 2000; Rogowsky, et al., 2015; Sadeghi, et al., 2012), no study has examined the impact of different learning styles on mastery of subskills of reading comprehension.

To enhance their performance in reading comprehension, examinees should get familiarized with their weaknesses in reading skills and strategies (by providing them with diagnostic feedback). However, the idea of providing examinees with a detailed mastery profile is rarely seen to be practiced and the examinees eventually end up having no idea about their deficiencies. Identifying the problems and going through them are considered key elements of progression which have been greatly ignored in reading comprehension assessment. Therefore, the present study aimed to diagnose a group of Iranian students' weaknesses in reading comprehension tasks, familiarize them with their deficiencies, and provide them with opportunities to focus on their problems. The problem of lack of diagnostic feedback only accounts for one aspect of this study.

The second purpose of the present study is to see whether individual differences can lead to students performing differently in reading comprehension tasks. Previous research corroborated the idea that there is a relationship between learners' personalities, the way they establish their learning styles, and their academic success (Dunn et al., 2002; Noftle & Robins, 2007; Sadeghi et al., 2012). As regards reading comprehension, there is a possibility that students with different sensory learning styles function differently on various subskills. Different subskills of reading comprehension are assumed to contribute to the final performance of the students. To clarify, as examinees are involved with written English in a reading comprehension task, it is regarded as a skill that requires the readers to have high visual skills. Thus, visual subskills such as word identification and phonological decoding seem to make a greater contribution to the performance and success of the students having visual learning styles (Vellutino et al., 2007). Thus, it is believed that detecting the impact of sensory learning styles can make a great contribution to the goal of CDA, which is providing fine-grained information to classify the students.

1.1 Research Questions

1. What are the strengths and weaknesses of Iranian university students in reading comprehension?

2. Is there any significant difference between the three sensory learning styles about their reading subskill mastery profiles?

[DOR: 20.1001.1.23223081.1401.0.0.144.6]

[DOI: 10.29252/LRR.15.3.10]

2. Literature Review

2.1. Cognitive Diagnostic Assessment

In all academic situations, a chain exists that includes three major components, namely, instruction, learning, and assessment (Lee & Sawaki, 2009b). These three play a complementary role in enhancing the quality of their fellow partners in the chain. The planned content of learning, if taken in by the students properly, demonstrates the appropriateness of the devised instruction. Meanwhile, this is the responsibility of the assessment to detect the flaws in the instruction, if there are any, or to determine if the process of learning has taken place. Therefore, since the very beginning, assessment has been of great importance to teachers and learners. So far, according to what Li and Suen (2013) noted, assessment has been used for two main purposes: a) Accountability; and b) Diagnosis.

In the former, the final unidimensional score, indicating the degree of achievement by the learners, suffices. It mostly focuses on the chance of comparability that it provides to make a comparison between test takers. The assessments based on the latter; however, tend to be diagnostically designed and are in search of rich, detailed, and comprehensive information about the areas of strength and weakness of the students. Traditional means of assessment would not be a wise choice for compiling fine-grained information regarding students' strengths and deficiencies for remedial measurements afterward (Bachman, 1990; Jang, 2005; Lee & sawaki, 2009b; Leighton & Gierl, 2007; Ranjbaran & Alavi, 2017). They mostly rely on the product of the assessment which is a single score and has nothing to do with what actual processes test takers apply to answer the questions. Diagnostically developed assessments, on the other hand, are the ones capable of collecting detailed data on test takers' responses and informing them of weaknesses leading to their poor performances (Rupp et al., 2010).

According to Lee and Sawaki (2009b), CDA consists of two major components. The first component is cognitive analysis to examine each of the items and come up with a list of cognitive attributes necessary for correct responses. In other words, cognitive analysis attempts to establish relationships between the test items and the cognitive attributes. The second critical component is said to be the psychometric modeling of the items. Psychometric modeling comes to play the statistical role of the assessment, checks the precision of cognitive analysis, and eventually informs the students of their areas of strength and weakness.

[DOR: 20.1001.1.23223081.1401.0.0.144.6]
[DOI: 10.29252/LRR.15.3.10]

There are four steps to a CDA. In the first one, *identifying the attribute*, which is scrutinizing the content of the test and items, analysts identify some probable attributes and then specify sets of attributes essential for each item to be correctly answered. The next step is *Q-Matrix construction*. Q-matrix is an item-and-attribute relationship display, indicating mastery of which skills or attributes is a prerequisite for correctly answering the item. According to Jang (2009a), establishing the links between learners' competencies and test items can be regarded as the most distinguished feature of CDA. The constructed Q-matrix is used for *data analysis* which is the third step of CDA. The psychometric aspect of CDA estimates the mastery profiles of test takers, using the data entered according to test takers' performance as well as a psychometric model chosen based on detected relationship in Q-matrix. The last step is when test takers are provided with diagnostic feedback about their cognitive strengths and weaknesses. Thus, test takers eventually have a detailed and comprehensive mastery profile at hand, informing them of their deficiencies and leading them to plan for remedial practices accordingly.

2.2. Fusion Model

The model used in this study is called the "fusion model". The fusion model, also known as a reduced reparametrized unified model (RRUM), was regarded as a non-compensatory model relatively comparable to the general model of G-DINA (Li & Lei, 2016). As the fusion model is a slightly modified version of the unified model, first, the major characteristics of the unified model will be discussed through four main components of the fusion model system which according to Roussos et al. (2007) are as follows:

1- An identifiable and interpretable item response function model, a reparameterization of the foundational Unified Model.

2- A parameter estimation method referred to as Arpeggio, which employs a Markov Chain Monte Carlo (MCMC) algorithm within a Bayesian modeling framework for model estimation, including item parameter estimation and ability distribution parameter estimation.

3- A collection of model checking procedures, including statistical MCMC convergence checking, ability distribution and item parameter estimates with standard errors, model fit statistics, internal validity statistics, and reliability estimation methods.

4- Skills-level score statistics, including mastery/non-mastery estimation, sub-scoring options for assessing mastery/non-mastery, and proficiency scaling statistics that relate test scores to skill mastery.

2.2.1. Reparametrized unified model

The primitive step in any model of CDM is the construction of a Q-matrix indicating which item of $i = 1\ldots..., I$ relates to which skill of $k = 1\ldots\ldots, K$. The fact that skill k is required by item i is shown by $q_{ik} = 1$ and the fact that skill k is not necessary for correctly responding to item i is shown by $q_{ik} = 0$. To delve into the deepest cognitive layers of test takers' mental processing to identify their deficiencies and gaps of knowledge, some parameters have been proposed predicting the test takers' subskill mastery profiles and they differ from model to model. The unified model includes both item and ability parameters which are represented by $P\ (Xij = x|\ \vartheta_j, \beta_i)$, where $X_{ij} = x$ is the response of examinee j to item i, (with $x = 1$ indicating a correct response and $x = 0$ an incorrect response), ϑ_j is a vector of examinee j ability parameters, and β_i is a vector of item i parameters. As regards the formula of the unified model, Roussos et al. (2007, p. 282) noted that:

$$\mathrm{P}\ (X_{i,j}{=}1|\ \alpha_j, \eta_j) = \pi_i^* \prod_{K=1}^{k} r_{ik}^{*(1-\alpha_{jk})\times q_{ik}} P_{\mathrm{ci}}(\eta_{\mathrm{k}})$$

The distinguishing feature of the unified model is that it considers even the existence of higher-order processing skills which may not be considered in Q-matric construction. In other words, the examinee parameter in the unified model is defined by two other parameters of α_Q and α_b. The vector α_Q accounts for all the underlying skills of the Q-matrix, while α_b takes into consideration any skills other than those specified by the Q-matrix.

Parameters in the formula are categorized into two kinds, examinee parameters and item parameters. While α_j, η_j are examinee parameters, three parameters of π_i^* , r_{ik}^*, and c_i are known as item parameters evaluating the quality of test items in terms of difficulty, discrimination, and completeness. To start with examinee parameters, α_j represents attribute mastery only for the attributes specified in the Q-matrix. In addition, η_j refers to probable attributes which are not indicated in the Q-matrix but could be used by the test takers to respond to the items. Item difficulty, π_i^* stands for the probability that an examinee having mastered all the required subskills in the Q-matrix will correctly respond to the item; hence, it shows the unacceptable difficulty

[DOR: 20.1001.1.23223081.1401.0.0.144.6]
[DOI: 10.29252/LRR.15.3.10]

of the item, if an examinee with all the subskills mastered, cannot answer it correctly.

As the item makes no difference between mastering and not mastering the needed subskills, it is better to be omitted. Thus, an acceptable probability value for item difficulty is defined above 0.6 to show good chances of correctly responding, providing that the examinees have mastered the required attributes. The other item parameter, r^*_{ik} is known as the item discrimination index, shows how well the items distinguish masters from non-masters. The value of these parameters is divided into three categories. Values beneath 0.3 show a strong dependence of the item on its associated subskills. It means one should master all the attributes to answer the item correctly. Thus, it has the most discrimination value. Values under 0.5 are just slightly less discriminatory than the previous range, and values under 0.8 offer just sufficient item discrimination between masters and non-masters. However, an item discrimination index above 0.8 is unacceptable and shows little difference between the ones having mastered the corresponding attributes and non-masters.

As for the last item parameter, also known as item completeness, it comes to say whether the specified attributes in the Q-matrix are sufficiently complete to lead test takers to respond correctly. This item completeness ranges from zero to three ($0 \leq c_i \leq 3$) with values close to zero, say, $c_i \leq 1.25$, indicating that the item has some unspecified attributes that are absent in the Q-matrix. On the other hand, values above 1.25 show the completeness of the item in terms of the associated and required attributes. In summary, by reparametrizing the original unified model, most critical examinee parameters are not only estimable but interpretable in terms of their underlying skills, the feature that makes this model superior (Roussos et al., 2007).

2.2.2. Parameters estimation algorithm

In advance of informing the test takers of their detailed mastery profiles, the selected model, here the fusion model, should estimate the value of each parameter to exclude unnecessary items and attributes. By doing so, items and their associated subskills can be analyzed to see if they should be included or not. Thus, the reliability of the assessment and its following results would be guaranteed to desirable extents.

Several methods for parameter estimation have been developed among which the Bayesian framework is one type (Roussos et al., 2007). Unlike other statistical approaches, the Bayesian approach does not treat the parameters as fixed and unknown parameters within a specific continuum in the normal sampling distribution,

[DOR: 20.1001.1.23223081.1401.0.0.144.6]
[DOI: 10.29252/LRR.15.3.10]

but it treats them as probability distributions that are unfixed (Niedermayer, 2008). Bayesian network has been applied for both ability and item parameters and within this framework, Markov Chain Monte Carlo (MCMC), is used. Together with MCMC, Expectation Maximization (EM) is also used for estimation purposes. However, according to what DiBello et al. (2007) noted, compared to MCMC, EM algorithms are more difficult. In addition, they are not fully straightforward (Patz & Junker, 1999). These two negative points of EM algorithms are reason enough to avoid applying this way of estimation and consider the alternative algorithm of MCMC.

As the first step in parameter estimation, Markov Chain estimates all the involved parameters, using simulated values. Following this step, a distribution consisting of several steps or simulated values is provided. According to the MCMC theory stated in DiBello et al. (2007), after a good number of steps or simulated values (called the burn-in phase), the remaining distribution will be closely like the desirable Bayesian posterior distribution of the parameters. The very considerable point here is the number of chains chosen by practitioners for the total length of the burn-in chain. The longer the length of discarded chains, the more reliable our final posterior distribution and the more precise parameter estimation.

2.2.3. Model checking procedure

After applying MCMC to estimate existing parameters in the model, checking procedures are essential to examine whether the parameters are working to their best functions or not. To further clarify, the process of model checking is carried out to evaluate the preciseness and accuracy of item-subskill relationships. It can be followed by some changes in the Q-matrix. As an example, this procedure checks whether the items measure the attributes reported in the Q-matrix. Within the fusion model, several checking methods have been proposed among which are convergence checking, interpretation of model parameters estimation, model fit statistics, and internal validity checks. Each of the mentioned steps for model checking will be fully explained in the next chapter. However, chain plots, as a means of convergence checking, are clarified as an example of a model checking procedure through a tangible example derived from what Li and Suen (2013) conducted. As can be seen in Figure 1 (Li & Suen, 2013), a time-series chain plot (for stability check of parameters) is displayed to check the diagnostic capacity of item 5 regarding skill 1,

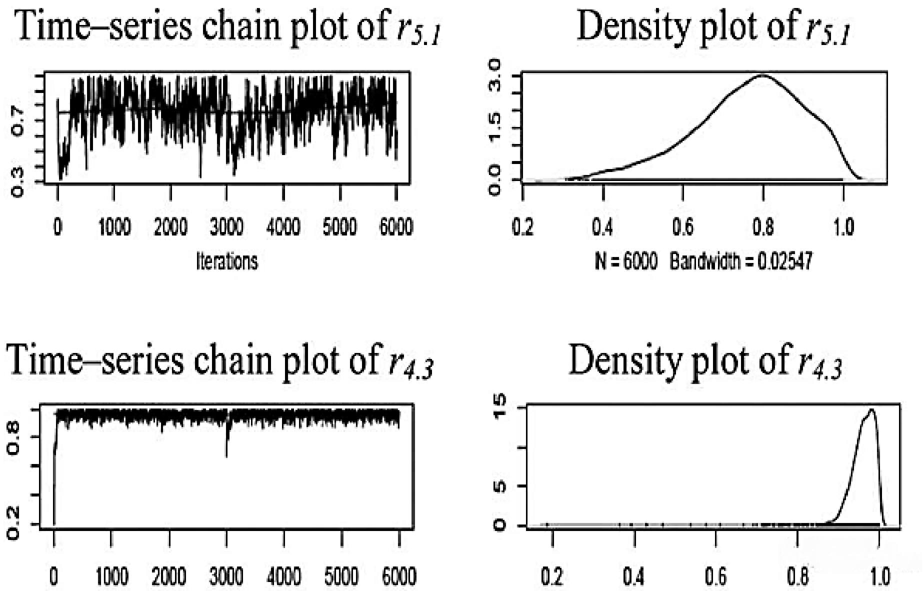
[DOR: 20.1001.1.23223081.1401.0.0.144.6]

[DOI: 10.29252/LRR.15.3.10]

along with diagnostic capacity of item4 regarding skill 3. It intends to check whether skill 1 and skill 3 can be measured by items 5 and 4, respectively.

Figure 1

Convergance Checking Through Time-Series Chain Plot and Density (Adapted from Li & Suen, 2013)



As displayed, there are many fluctuations and almost no convergence for $r_{5.1}$. This indicates that skill 1 should be excluded from the group of attributes for item 5. However, the full convergence seen for $r_{4.3}$ represents the stability of the parameter value, meaning that skill 3 can be measured through item 4. The other MCMC checking graphic is the density plot which is employed to determine if the mean of the parameter has been stabilized. Just like the time-series chain plot for $r_{5.1}$, the density plot does not provide a desirable result, as the distribution of the parameter's mean has a wide range showing a lack of stability. On the other hand, as the distribution for $r_{4.3}$ is narrowed focusing on one spot, it suggests great stability of the parameter.

2.3. Reading Comprehension Assessment

Out of the four existing skills of language, reading comprehension has recently been at the center of attention of many language teachers, instructors, and measurement

[DOR: 20.1001.1.23223081.1401.0.0.144.6]

[DOI: 10.29252/LRR.15.3.10]

specialists (Ranjbaran & Alavi, 2017; Baghaei & Ravand, 2016; Chen & Chen, 2016; Fletcher, 2006; Jang, 2005; Jang, 2009a; Javadianmehr & Anani Sarab, 2019; Li & Suen, 2013; Ravand, 2016; Ravand & Robitzsch, 2015; Soltani & Taghizadeh, 2023). Reading comprehension assessment seems to be a quite demanding and painstaking task, as one should consider different aspects simultaneously while examining it. The questions that should be considered while examining a reading test are as follows: What variables play a role in an acceptable performance of a reading comprehension test, how does reading relate to memory, how much does it depend on text type, how does it relate to other cognitive abilities, how reading comprehension ability differs in a second or foreign language (Alderson, 2000). Nevertheless, the main purpose of all the above-mentioned studies revolves around the processes, strategies, subskills, and attributes that test takers go through while taking a reading comprehension test.

Regarding various categories of DCM, such as compensatory and non-compensatory, the proposed psychometric models that would best fit our intended data and lead to the most reliable and solid cognitive data about test takers are sought. Finding these models does not direct us to applying the sole best DCM ever. The very considerable fact that should be taken into consideration while intending to single out a DCM is the inter-skill relationship existing between the attributes (Ranjbaran & Alavi, 2017; Jang, 2009a; Javadianmehr & Ananisarab, 2019; Ravand, 2016; Ravand & Robitzsch, 2015).

Apart from groups of specific and general models that play their roles in model selection, some researchers have conducted studies to measure the degrees of reliability and meticulousness of DCMs regarding reading comprehension skills. What Ravand and Robitzsch (2018) found out is that in a reading test, the interactions between the specified subskills could be a combination of both compensatory and non-compensatory. Thus, a general DCM, flexible for both compensatory and non-compensatory relationships, can be suggested as the best choice of model. Furthermore, Li and Lei (2016) made a comparison between a general model of GDINA, two compensatory models of DINO and ACDM, and two non-compensatory models of DINA and RRUM. Of these stated models, the general model, G-DINA, was found to have the best model fit and classification results. Then, RRUM, as a non-compensatory model, showed negligible difference compared to the G-DINA model.

[DOR: 20.1001.1.23223081.1401.0.0.144.6]
[DOI: 10.29252/LRR.15.3.10]

2.4. Learning Styles

The word "learning style" is self-explanatory. As the name suggests, learning style could be defined as the way one learns something, and in terms of language learning, it refers to how well learners acquire a language resorting to their preferred ways of learning (Oxford, 2003). The notion of learning styles primarily came to confirm the individual differences among language learners and propose the most tailored learning guidelines for learners having various tastes in acquiring a new language (Oxford, 2003; Sadeghi et al., 2012). As regards individual differences, Mosalli et al. (2022) have emphasized the important role of individual characteristics in language testing contexts, concluding that different personality traits, such as strategy-use behaviors, impact learners' performance in language testing contexts. In addition, it is considerably important to know that according to what Sadeghi et al. (2012) noted, learning styles should never be treated as black-or-white concepts. They are regarded as some continuums that learners may fall somewhere in between the extremes.

According to Reid (1995), learning styles are divided into three major groups: 1- Cognitive learning styles; 2- Sensory learning styles; and 3- Personality learning styles. Based on Sadeghi et al. (2012), different learning styles can be categorized in the following pattern: Cognitive learning styles are of three types: 1- Field-independent vs. Field-dependent; 2- Analytic vs. Global; 3- Reflective vs. Impulsive. Sensory learning styles are categorized into two sub-groups: 1- Perceptual learning styles: Auditory learner, Visual learner, Tactile learner; 2- Environmental learning styles: Physical vs. Sociological learner. In the end, personality learning styles consist of Extroversion vs. Introversion; Sensing vs. Perception; Thinking vs. Feeling; Judging vs. Perceiving; Ambiguity tolerant vs. Ambiguity-intolerant; and Left-brained vs. Right-brained learners. Before moving on, a very brief review of some of these learning styles can give us an insight into the importance of the role they play in the process of learning. For instance, field-independent learners are the ones who tend to notice the details as they do not consider the background or the field they are working on. This is the reason why they are also called analytic learners. On the contrary, field-dependent learners tend to look at issues more holistically. Thus, they are also called synthetic and global learners.

On the other hand, impulsive learners are the ones who come up with some guesses before reaching a solution. This is in marked contrast with reflective learners who weigh up all the probable options in advance of offering any kind of final response. Further explanation of details of these different learning style categories is beyond the purpose of this study. However, it is worth mentioning that regarding these existing individual differences in the way they grasp knowledge, this question may be raised if the efficacy and efficiency of the learning process are tied to these 45 various preferences of students. In other words, these different learning styles have opened possibilities for further research on the potential impacts on the efficiency of learning as well as performance in various domains (Gohar & Sadeghi, 2015; Hsu, 2017; Li et al., 2014).

What this study focuses on is the impact of the perceptual category of sensory learning styles on learners' performance in reading comprehension tasks. To define what sensory learning styles are, Gilakjani (2012) noted that visual learners are the ones whose learning is far more efficient when visual materials are applied. In other words, visual materials facilitate the process of learning and even increase the quality. It is said that visual learners act more efficiently in reading tasks than their auditory or tactile counterparts. Auditory learners, on the other hand, rely on their ears, listen, and learn by interpreting the information using pitch, emphasis, and speed. These learners show preferences for reading out loud in the classrooms. The last group of tactile learners are the ones who favor physical and hands-on activities and learn through interaction with the world.

3. Methodology

3.1. Participants and Sampling

As the very first purpose of this study was to diagnose the areas of strength and weakness of Iranian students, a sample of Iranian university students was needed. The test was sent to different students via email, and they were asked to complete the test. The process of data collection lasted around three months, as finding virtual students was a difficult task. This sample of students consisted of 301 students of different disciplines ranging from pharmacy students to urban management students. Participants were chosen from either state or private universities. As the students were chosen from both bachelor's and master's degrees, the age of the students made a very wide continuum, ranging from 19 to 47, with an average of 24. The considerable point was the students' level of English proficiency. This required us to limit the test takers to students with at least an intermediate level of proficiency in the English language. On the other hand, the advanced students who were able to answer all the questions correctly had to be excluded, as the areas of weakness could not be

[DOR: 20.1001.1.23223081.1401.0.0.144.6]

[DOI: 10.29252/LRR.15.3.10]

diagnosed for students who had correctly responded to the entire set of items.

3.2. Instruments

The following instruments were used in this study.

3.2.1. Reading comprehension test

The reading considered for this study was a 13-item diagnostically designed reading test by a doctoral student in Teaching English (Mesgarshar, 2020). As noted, employing CDA is a very informative, yet delicate, method to provide the students, teachers, and instructors with a detailed profile of examinees' weaknesses and strengths. To access reliable results of cognitive assessments, the retrofitting approach of CDA would never make the best choice, as it is based on an existing non-diagnostic reading test (Liu et al., 2017).

What Mesgarshahr (2020) did was design a reading test that can assess reading performance deficiencies of Iranian students based on the Attribute Hierarchy Method (AHM; Leighton et al., 2004). For doing so, the first step was providing a list of required reading skills obtained out of related literature. Reading skills were also extracted out of non-diagnostically designed tests, and it was considered a serious problem. To solve the problem, a practical way for identifying the essential subskills was applied, such as a think-aloud protocol. After developing the initial Q-matrix, designing the reading test based on the reading subskills was conducted. The reading test was finalized after repetitive rounds of verbal protocol analysis of the designed test and revisions. The hierarchy of selected and diagnosed subskills are as follows: A1: basic linguistic knowledge (BLK); A2: constructing prepositions to achieve local comprehension (CPL); A3: comprehending implicitly stated information (ISI); A4: Understanding text with difficult vocabulary (UDV); A5: understanding complex text (UCT); A6: combining prepositional meaning to achieve more global comprehension (CPG); and A7: summarizing textual information (STI).

3.2.2. Learning style survey

Learning styles are preferable ways through which one learns things. It refers to how well learners acquire a language by resorting to their preferred ways of learning (Oxford, 2003). The concept of "learning style" confirms the individual differences among language learners and recommends the best learning guidelines for learners having various tastes in acquiring a new language (Oxford, 2003; Sadeghi et al., 2012).

As noted, this study focused on the sensory learning styles of Iranian university students and categorized them into three groups, visual, auditory, and tactile. To do so, a learning style questionnaire designed at the University of Texas (Bright Success Center, 2006) was used and the participants were asked to carefully respond to 24 items regarding their preferred ways of learning. The questionnaire is shown in the appendix section.

The questions were translated into Persian for easing the process. After collecting the surveys, the answers were changed into specific numbers (often = 5, sometimes = 3, seldom = 1) and added up to reach a final number dedicated to each group. The highest number signified the group the reader belonged to.

3.2.3. The Q-matrix

The Q-matrix for the reading comprehension test and attributes were created. the final Q-matrix is shown below:

$$Q = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

In the Q-matrix shown above, the rows represent seven subskills studied in this research, with the columns representing 13 test items. What the Q-matrix intends to show is the relationship between the items and the skills. As an example, for an examinee to respond to item 5 correctly, attributes 1, 2, 3, 4, and 5 should be mastered. As seen in the Q-matrix, all the items have attribute 1 as a prerequisite element for a correct response. However, items 10, 11, 12, and 13 are the only items that measure attribute 7. Other relationships within this Q-matrix can be interpreted in the same manner.

[DOR: 20.1001.1.23223081.1401.0.0.144.6]

[DOI: 10.29252/LRR.15.3.10]

In addition to the item-subskill relationship, there is also a within-skill relationship, showing which attributes should be considered as a prerequisite for the other attributes. This is shown within the Reachability matrix (Tatsuoka, 1990). This matrix indicates the relationships between the attributes (1 indicates that the attribute is essential, and 0 indicates no connections between the attributes). In other words, this matrix signifies the interconnectedness between attributes within a test. Take the given matrix as an example. The R-matrix suggests that to master attribute 3, four attributes of 4, 5, 6, and 7 are considered as prerequisites which are shown by 1.

$$R = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

As can be seen in the R matrix, attribute 1 is a prerequisite for all the other subskills. On the other hand, attribute 7 is not a prerequisite for any other subskill shown in the Q-matrix. The other relationships within this Q-matrix can be explained in the same manner.

4. Results

To answer the research questions, we analyzed the data using the R package. The first step we took was to make sure of the validity of the results. As mentioned before, the study used the Fusion model in measuring the variables. To see if the items measure the attributes, the MCMC estimated the posterior distribution for each item and subskill, and the average point of each was calculated and reported. High probabilities could best support this claim that the item measures the associated attributes (Raftery, 1996). Accordingly, the less the standard deviation of the estimated distribution, the more reliable the results would be.

First, it is a prerequisite to assign different success possibilities to each item. Table 1 reports the parameter estimation for item 1, as an example.

[DOR: 20.1001.1.23223081.1401.0.0.144.6]

[DOI: 10.29252/LRR.15.3.10]

Table 1

Item Parameter Estimation for Item 1

Item	Attributes Measured	P	Standard Error
1	1,2	0.9368	0.0172
2	1,2,3	0.9435	0.0346
3	1,2,3,4	0.9999	0.0626
4	1,2,3,5	0.9999	0.0873
5	1,2,3,4,5	0.9999	0.2266
6	1,2,3,6	0.8615	0.0606
7	1,2,3,4,6	0.6771	0.1068
8	1,2,3,5,6	0.8592	0.0627
9	1,2,3,4,5,6	0.9612	0.0820
10	1,2,3,6,7	0.7648	0.1089
11	1,2,3,4,6,7	0.7926	0.1136
12	1,2,3,5,6,7	0.9492	0.0888
13	1,2,3,4,5,6,7	0.8791	0.1185

As we can see in Table 1, the first column represents items of the reading test accompanied by the second column illustrating the attributes that the item was supposed to measure. The third column in Table 1, is the probability of the items measuring the intended subskills. The higher this probability the better, as it shows correct predictions and specification of item-attribute relationship. The probabilities shown in Table 2 are all acceptable and show a good degree of item-skill fit.

Along with the probabilities shown in the previous section which indicates the appropriateness of the Q-matrix, it is of great importance to make sure that the model fits the data. Depending on which package the researcher uses for analyses, many diverse fit statistics show if there is a desirable fit for the model and data. Most fit statistics are reported based on the average difference between the predicted data by the model and the observed data obtained from actual participants. The less this difference, the more accurate model estimations are and the more congruence the model has with the data, meaning that there is a good model-data fit based on which all the results can be reported and relied on.

The package used in this study is the GDINA package in which the Square Root of Mean Square Residuals (SRMSR) has been reported as a proper absolute model-data fit statistic for dichotomous responses (Hansen et al., 2016; Liu et al., 2016).

Maydeu-Olivares (2013, p. 84) noted that:

$$\text{SRMSR} = \sqrt{\sum_{i<j} \frac{(r_{ij}-p_{ij})}{n(n-1)/2}}$$

[DOR: 20.1001.1.23223081.1401.0.0.144.6]

[DOI: 10.29252/LRR.15.3.10]

According to Maydeu-Olivares (2013), SRMSR has been regarded as a very straightforward model-fit statistic. He added that SRMSR ≤ 0.05 can be reported as a cutoff point, meaning that numbers under 0.05 indicate a negligible amount of misfit between the model and data. The package reported SRMSR as being 0.0378 which is a small number and a favorable fit statistic, indicating a desirable fit between the predicted and observed data.

4.1. The First Research Question

Having made sure of the validity of the measurement and the acceptability of the model fit, the data obtained from the Fusion model were used to answer the research questions. As mentioned in previous sections, the reading comprehension test was given to 301 Iranian university students who were asked to response to the items carefully. Depending on how the examinees performed in the reading comprehension test, they were supposed to be assessed in terms of their deficiencies and strengths in seven diverse attributes.

The skill mastery profile of each examinee is also needed in which the mastery probability for every attribute is reported. Having said this, examinees have this opportunity to refer to these profiles and gain a complete understanding of their reading capabilities and deficiencies.

Before reporting the attribute mastery probabilities of the test takers, it is worth mentioning that attribute probability 1 means mastery of the attribute (the examinee has mastered the attribute), and 0 means non-mastery (the examinee has not mastered the attribute). The probabilities close to 1 are regarded as mastery, and the ones close to 0 are considered as not mastered of the attribute. However, how each probability number is attributed to either mastery or non-mastery is discussed in the following part.

There are two different ways of treating the attribute mastery probabilities. In the first one, examinees are categorized into two groups, masters and non-masters. It means that there is a cutoff point of 0.5 and test takers with probabilities above the cutoff point fall in the category of masters and the examinees with probabilities under 0.5, are regarded to be non-masters (Roussos et al., 2007). On the other hand, according to Jang (2005), three ranges can be specified for the probabilities which are as follows:

[DOR: 20.1001.1.23223081.1401.0.0.144.6]
[DOI: 10.29252/LRR.15.3.10]

1- $0 \le p < 0.40$

2- $0.40 \le p < 0.60$

3- $0.60 \le p \le 1.00$

In this study, range 1 is considered as mastery, range 2 is regarded as possible mastery and range 3 is non-mastery. The problem in the former classification is that it is not exactly clear whether an examinee with an attribute probability of 0.5 should be regarded as a master or non-master of the attribute. The latter classification, however, seems to be more precise, as a range has been specified to examinees with possible mastery of the attribute. As an overall look over areas of weakness and strength of 301 participants taking part in this reading comprehension test, Table 2 has been shown.

Table 2

Examinees Mastery Percentage of Each Attribute

Attribute	Level 0	Level 1
A1(BLK)	6.45%	93.55%
A2(CPL)	5.15%	94.85%
A3(ISI)	72.78%	27.22%
A4(UDV)	68.44%	31.56%
A5(UCT)	22.64%	77.36%
A6(CPG)	31.7%	68.3%
A7(STI)	66.88%	33.12%

As can be seen in Table 2, regarding BLK, 93.55% of the examinees are reported to be masters of this attribute with only 6.45% having not mastered BLK. The same results have been reported for CPL, with mastery and non-mastery percentages of 94.85% and 5.15% respectively. It is important to note that ISI is mastered by only 27.22% of the test takers and 72.78% of the examinees were not regarded as masters of this attribute. Other mastery probabilities for attributes are also shown in Table 2 and can be explained in the same manner. According to the table, it is noticeable that high proportions of the examinees have not mastered ISI, UDV, and STI.

[DOR: 20.1001.1.23223081.1401.0.0.144.6]

[DOI: 10.29252/LRR.15.3.10]

Table 3 shows descriptive statistics of the attribute probabilities for 301 participants of the study. Regarding BLK, the minimum mastery probability for this attribute has been reported as 0.0002 and the maximum mastery probability is 1.

Table 3

Descriptive Statistics of Attribute Mastery Probabilities

Skill	N	Min	Max	Mean	SD
BLK	301	0.0002	1	0.935	0.2069
CPL	301	0.0001	1	0.948	0.2001
ISI	301	0.0002	1	0.272	0.4177
UDV	301	0	1	0.315	0.4021
UCT	301	0	0.9945	0.773	0.3301
CPG	301	0	1	0.683	0.3624
STI	301	0.0011	1	0.331	0.3529

According to Table 3, the mean mastery probability of BLK for the 301 examinees is 0.935, indicating that most participants have mastered this attribute. For the other six attributes, mastery probability can be interpreted in the same manner.

4.2. The Second Research Question

As mentioned in previous sections, examinees were categorized into three groups, visual, auditory, and tactile learners. The number of examinees falling in the visual, auditory, and tactile groups were 116, 97, and 28 persons respectively. This must be noted that 60 test-takers in this section were excluded, as they did not show any preferences regarding their learning styles, meaning that they had at least two learning styles simultaneously.

Regarding the presented results, the question of whether visual learners showed significantly better overall reading performance, compared to their auditory and tactile counterparts may be raised. To seek the answer to this question, ANOVA was carried out to see if there was any significant difference regarding the overall performances of the examinees. The results of ANOVA on overall performances were significant ($F = 9.41$, $p = .0024 < .05$) and the post hoc analysis showed that the difference only existed between visual and auditory learners, the former significantly ($MD = .872$, $0 = .007 < .05$) outperforming the latter. To further examine the data, the attributes were compared within the seven skills. The results of ANOVAs are reported in Table 4.

[DOI: 10.29252/LRR.15.3.10]

Table 4

Results of Between-Groups ANOVAs for Seven Skills

Skill	Sum of Squares	df	Mean Square	F	Sig.
A1(BLK)	1.0834	2	0.54171	6.8351	0.001298*
A2(CPL)	0.487	2	0.24363	1.4191	0.244
A3(ISI)	1.58	2	0.78993	4.2976	0.01467 *
A4(UDV)	1.659	2	0.82944	4.6373	0.01057 *
A5(UCT)	0.9015	2	0.45073	3.8468	0.02268 *
A6(CPG)	0.2358	2	0.11788	0.9072	0.405
A7(STI)	0.4711	2	0.23554	1.9954	0.1382

The results suggest a significant difference between the three groups of visual, auditory, and tactile test takers only in four subskills of BLK, ISI, UDV, and UCT. The post hoc comparisons showed that within BLK, visuals significantly outperformed both auditory and tactile learners. Concerning ISI, UDV, and UCT, visual learners significantly outperformed auditory learners. No other significant differences were found.

5. Discussion

The results indicated that most Iranian university students have mastered BLK, CPL, UCT, and CPG. It suggested that most students possess the basic linguistic knowledge (BLK), they can construct propositions to understand local meanings (CPL), most of them can successfully comprehend complex texts (UCT), and they can combine propositions to build a more global understanding of texts. On the other hand, they have difficulty when it comes to implicitly stated information and difficult vocabulary in the text (ISI and UDV). In addition, non-mastery was also found in STI which refers to an attribute that requires test takers to keep pieces of information in mind, summarize them, and make conclusions. In sum, the three skills of ISI, UDV, and STI are regarded to be more difficult for participants to master. Moreover, ISI, with the lowest mean of 0.272 proved to be the most cognitively demanding attribute for the participants to master. It is followed by UDV and STI, with means of 0.315 and 0.331 respectively.

According to what Mesgarshahr (2020) has suggested, learners are expected to show mastery in more of the initial subskills, as they are set depending on their degrees of easiness. In other words, concerning reading comprehension, intermediate students are expected to have mastered prerequisites of reading, including basic

[DOR: 20.1001.1.23223081.1401.0.0.144.6]

[DOI: 10.29252/LRR.15.3.10]

knowledge, proposition construction at an elementary level, and being able to deal with relatively complex texts. The following subskills as they get harder are less likely to be mastered by intermediate students.

Concerning the related literature, the results of the present study are in line with some studies which is to be explained in the next part. The results of Mesgarshahr (2020) indicated that most Iranian university students, almost mastered four skills, BLK, CPL, ISI, and UDV. Additionally, the next three attributes were found to be more difficult to master. In his result, STI was the most cognitively demanding attribute. It is consistent with the result of the present study since STI was found to be among the most cognitively demanding attributes that test takers had not mastered. In addition, just like the result of the present study, BLK and CPL were reported as the easiest attributes most of the students possessed.

It is reported by Jang (2009b) that TIM (Textually Implicit Information), regarded as one of the attributes in her study, was also among the attributes that test takers had difficulty dealing with. This is in line with what we reported for ISI, which was supposed to measure the ability of students to comprehend implicitly stated information. As reported in the present study, like what Jang (2009b) concluded, participants had difficulty mastering implicitly stated information. Regarding implicitly stated information, the same results have been also attained by Ranjbaran and Alavi (2017). The mastery probability of the attributes reported by Ranjbaran and Alavi (2017) also indicated that comprehending text-implicit information dedicated the least amount of mastery probability to itself, confirming the fact that this skill falls in the group of demanding skills for students to master, which is compatible with the results of the present study.

The present results also back up the research conducted by Javadianmehr and Anani Sarab (2019). Most of the participants in their study failed to master the attribute of "connecting and synthesizing". This is compatible with the non-mastery of STI by most of the test takers in the present study. Summarizing textual information comprises the ability to integrate and synthesize different sections of the text which seems to be missing in most Iranian university students in both studies. Additionally, just like the present study, Javadianmehr and Anani Sarab (2019) reported the skill of "local comprehension" as an attribute mastered by a high proportion of the examinees, which was about 70%. This attribute has been called CPL (Comprehending Propositions to Understand Local meanings) in the present study and was mastered by 94.85% of test takers.

The second part of this study concentrated on comparing three groups of learners with their specific learning styles. The result here signified the superiority of the visual learners in four skills of BLK, ISI, UDV, and UCT, and their considerably better overall reading performance than their auditory fellow test takers. In the other attributes, no significant difference was found between the three categories. Since reading comprehension can be regarded as more of a visual skill, the obtained result is in line with the assumption of this study which expected the visual learners to perform better in general and in some attributes specifically.

6. Conclusion

In conclusion, the present study provided practical evidence on the applicability of the Fusion model in analyzing the reading comprehension of the participants with different learning styles, by obtaining a good model-data fit statistic. The results in terms of learners' areas of strength and weakness indicated that most Iranian students have not mastered ISI, UDV, and STI. It means that most Iranian learners have difficulty responding to the items that require them to figure out implicitly stated information, know the meaning of difficult vocabulary, and the items requiring them to combine different sections of the text to come up with a reasonable conclusion. As STI was the most demanding attribute (Mesgarshahr, 2020), it was expected that students would not master this attribute.

In the next section, the difference between three groups of learners, visual, auditory, and tactile learners, was sought about their overall performance in both the reading test and the seven attributes of the reading comprehension test. The findings indicated the superiority of visual learners in both overall reading performance and four subskills of BLK, ISI, UDV, and UCT. This reinforces the prominent role of learning styles and the probable advantages they bring about in academic environments.

This study also found that CDA can play a crucial role in the weakness diagnosis of learners in academia. Also, instructors, teachers, and educationalists can take advantage of this precise means of assessment to grasp a more comprehensive understanding of their educational contexts. This can also be concluded that many variables play parts in constructing an efficient curriculum for the target population of learners. Learning styles of the majority along with precisely diagnosed areas of deficiency can be regarded as influential factors, leading to well-developed

[DOR: 20.1001.1.23223081.1401.0.0.144.6]

[DOI: 10.29252/LRR.15.3.10]

instructions.

It should not be left unmentioned that this research was limited in its scope by contextual features like the participants' ethnicity, educational context, and language proficiency. Other studies may extend the generalizability of the findings by applying the same model in different contexts.

[DOR: 20.1001.1.23223081.1401.0.0.144.6]

[DOI: 10.29252/LRR.15.3.10]

[DOR: 20.1001.1.23223081.1401.0.0.144.6]

[DOI: 10.29252/LRR.15.3.10]

References

Alderson, J. C. (2000). *Assessing reading*. Cambridge University Press.

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford University Press.

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

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 5–31. https://doi.org/10.1007/s11092-008-9068-5

Chen, H., & Chen, J. (2016). Exploring reading comprehension skill relationships through the G-DINA model. *Educational Psychology*, *36*(6), 1049–1064. https://doi.org/10.1080/01443410.2015.1076764

DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques. In P. D. Nichols, D. F. Chipman, & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361–389). Lawrence Erlbaum.

Dunn, R., Beaudry, J. S., & Klavas, A. (2002). Survey of research on learning styles. *California Journal of Science Education*, *2*(2), 75–98. https://www.researchgate.net/publication/242174436

Fletcher, J. M. (2006). Measuring reading comprehension. *Scientific Studies of Reading*, *10*(3), 323–330. https://doi.org/10.1207/s1532799xssr1003_7

Gilakjani, A. P. (2012). Visual, auditory, and kinaesthetic learning styles and their impacts on English language teaching. *Journal of Studies in Education*, *2*(1), 104–113. https://doi.org/10.5296/jse.v2i1.1007

Gohar, M. J., & Sadeghi, N. (2015). The impact of learning style preferences on foreign language achievement: A case study of Iranian EFL students. *Procedia-Social and Behavioral Sciences*, *171*, 754–764. https://doi.org/10.1016/j.sbspro.2015.01.188

Hansen, M., Cai, L., Monroe, S., & Li, Z. (2016). Limited-information goodness-of-fit testing of diagnostic classification item response models. *British Journal of Mathematical and Statistical Psychology*, *69*(3), 225–252.

https://doi.org/10.1111/bmsp.12074

Hsu, T. C. (2017). Learning English with augmented reality: Do learning styles matter? *Computers & Education*, *106*, 137–149. https://doi.org/10.1016/j.compedu.2016.12.007

Jang, E. E. (2005). *A validity narrative: Effects of reading skills diagnosis on teaching and learning in the context of NG TOEFL* [Unpublished doctoral dissertation]. University of Illinois at Urbana-Champaign.

Jang, E. E. (2009a). Demystifying a Q-matrix for making diagnostic inferences about L2 reading skills. *Language Assessment Quarterly*, *6*(3), 210–238. https://doi.org/10.1080/15434300903071817

Jang, E. E. (2009b). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for Fusion Model application to LanguEdge assessment. *Language Testing, 26*(1) 031–073. https://doi.org/10.1177/0265532208097336

Javidanmehr, Z., & Anani Sarab, M. R. (2019). Retrofitting non-diagnostic reading comprehension assessment: Application of the G-DINA model to a high stakes reading comprehension test. *Language Assessment Quarterly*, *16*(3), 294–311. https://doi.org/10.1080/15434303.2019.1654479

Lee, Y. W., & Sawaki, Y. (2009a). Cognitive diagnosis approaches to language assessment: An overview. *Language Assessment Quarterly*, *6*(3), 172–189. https://doi.org/10.1080/15434300902985108

Lee, Y. W., & Sawaki, Y. (2009b). Application of three cognitive diagnosis models to ESL reading and listening assessments. *Language Assessment Quarterly, 6*, 239–263. https://doi.org/10.1080/15434300903079562

Leighton, J. P., Gierl, M. J., & Hunka, S. M. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoka's rule-space approach. *Journal of Educational Measurement*, *41*(3), 205–237. https://doi.org/10.1111/j.1745-3984.2004.tb01163.x

Leighton, J., & Gierl, M. (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.

Li, H., & Suen, H. K. (2013). Detecting native language group differences at the subskills level of reading: A differential skill functioning approach. *Language Testing*, *30*(2), 273–298. https://doi.org/10.1177/0265532212459031

Li, H., Hunter, C. V., & Lei, P. W. (2016). The selection of cognitive diagnostic models for a reading comprehension test. *Language Testing*, *33*(3), 391–409. https://doi.org/10.1177/0265532215590848

Li, Y. S., Yu, W. P., Liu, C. F., Shieh, S. H., & Yang, B. H. (2014). An exploratory study of the relationship between learning styles and academic performance among students in different nursing programs. *Contemporary Nurse*, *48*(2), 229–239. https://doi.org/10.1080/10376178.2014.11081945

Liu, R., Huggins-Manley, A. C., & Bulut, O. (2017). Retrofitting diagnostic classification model to responses from IRT-based assessment forms. *Educational and Psychological Measurement*, *78*(3), 357–383. https://doi.org/10.1177/0013164416685599

Liu, Y., Tian, W., & Xin, T. (2016). An application of M 2 statistic to evaluate the fit of cognitive diagnostic models. *Journal of Educational and Behavioral Statistics*, *41*(1), 3–26. https://doi.org/10.3102/1076998615621293

Maydeu-Olivares, A. (2013). Goodness-of-fit assessment of item response theory models. *Measurement: Interdisciplinary Research and Perspectives*, *11*(3), 71–101. https://doi.org/10.1080/15366367.2013.831680

Mesgarshahr, A. (2020). *Cognitive diagnostic assessment of reading: An application of the attribute hierarchy method* [Unpublished doctoral dissertation]. University of Tehran.

Mosalli, Z., Marandi, S. S., & Tajik, L. (2022). Cognitive and metacognitive strategy use in reading: The case of Iranian EFL students' test performance. *Language Related Research*, *13*(3), 55– 85. http://dx.doi.org/10.52547/LRR.13.3.4

Niedermayer, D. (2008). An introduction to Bayesian networks and their contemporary applications. In D. E. Holmes & L. C. Jain (Eds.), *Innovations in Bayesian networks* (pp. 117–130). Springer.

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology*, *93*(1), 116–130. https://doi.org/10.1037/0022-3514.93.1.116

Oxford, R. L. (2003). Language learning styles and strategies: Concepts and relationships. *IRAL*, *41*(4), 271–278. https://doi.org/10.1515/iral.2003.012

Patz, R. J., & Junker, B. W. (1999). A straightforward approach to Markov chain Monte

Carlo methods for item response models. *Journal of Educational and Behavioral Statistics*, *24*(2), 146–178. https://doi.org/10.3102/10769986024002146

Pfister, A. A. (2001). The effect of personality type of bilingual students on English reading performance in a computer-driven developmental reading laboratory: Implications for educational leaders. *Journal of Reading Behavior*, *8*(3), 335–336.

Ranjbaran, F., & Alavi, S. M. (2017). Developing a reading comprehension test for cognitive diagnostic assessment: A RUM analysis. *Studies in Educational Evaluation*, *55*, 167–179. https://doi.org/10.1016/j.stueduc.2017.10.007

Ravand, H. (2016). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. *Journal of Psychoeducational Assessment*, *34*(8), 782–799. https://doi.org/10.1177/0734282915623053

Ravand, H., & Robitzsch, A. (2015). Cognitive diagnostic modeling using R. *Practical Assessment, Research & Evaluation, 20*(11), 1–12. https://doi.org/10.7275/5g6f-ak15

Ravand, H., & Robitzsch, A. (2018). Cognitive diagnostic model of best choice: A study of reading comprehension. *Educational Psychology*, *38*(10), 1255–1277. https://doi.org/10.1080/01443410.2018.1489524

Reid, J. M. (1995). *Learning styles in the ESL/EFL classroom*. Heinle & Heinle Publishers.

Rogowsky, B. A., Calhoun, B. M., & Tallal, P. (2015). Matching learning style to instructional method: Effects on comprehension. *Journal of Educational Psychology*, *107*(1), 64–78. https://doi.org/10.1037/a0037478

Roussos, L. A., DiBello, L. V., Stout, W., Hartz, S. M., Henson, R. A., & Templin, J. L. (2007). The fusion model skills diagnosis system. In J. Leighton & M. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 275–318). Cambridge University Press.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications.* Guilford Press.

Sadeghi, N., Kasim, Z. M., Tan, B. H., & Abdullah, F. S. (2012). Learning styles, personality types and reading comprehension performance. *English Language Teaching*, *5*(4), 116–123. https://doi.org/10.5539/elt.v5n4p116

Soltani, F., & Taghizadeh, M. (2023). Investigating high and low achieving talented students' strategy use, perceptions, and challenges of reading comprehension. *Language Related Research, 14*(4), 141–171. http://dx.doi.org/10.29252/LRR.14.5.6

Soemer, A., & Schiefele, U. (2019). Text difficulty, topic interest, and mind wandering during reading. *Learning and Instruction*, *61*, 12–22. https://doi.org/10.1016/j.learninstruc.2018.12.006

Taras, M. (2005). Assessment–summative and formative–some theoretical reflections. *British Journal of Educational Studies, 53*(4), 466–478. https://doi.org/10.1111/j.1467-8527.2005.00307.x

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*(4), 345–354. https://www.jstor.org/stable/1434951

Vaughn, S., Roberts, G., Capin, P., Miciak, J., Cho, E., & Fletcher, J. M. (2019). How initial word reading and language skills affect reading comprehension outcomes for students with reading difficulties. *Exceptional Children*, *85*(2), 180–196. https://doi.org/10.1177/0014402918782618

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skills model of reading development. *Scientific Studies of Reading*, *11*(1), 3–32.

About the Authors

Mona Askari is currently a Ph.D. student at the Department of Counseling, Leadership, Adult Education & School Psychology, Texas State University, majoring in Adult, Professional, and Community Education. She has completed her master's degree in Applied Linguistics/TESOL at the English Department, Faculty of Foreign Languages and Literatures, University of Tehran, Iran. Her research interests are applied linguistics/TESOL and teacher education.

Hossein Karami is a lecturer in Applied Linguistics/TESOL at the English Department, Faculty of Foreign Languages and Literatures, University of Tehran, Iran. His areas of interest include validity and fairness, especially in the context of language testing. His research has been published in various international scholarly journals including *Language Testing, Educational Research and Evaluation, International Journal of Bilingual Education and Bilingualism, RELC Journal, Journal of Multilingual and Multicultural Development, Language Sciences, Current Psychology, Psychological Test and Assessment Modeling, TESOL Journal,* and *Asia-Pacific Education Review*.