

*Critical Studies in Texts & Programs of Human Sciences,*  
Institute for Humanities and Cultural Studies (IHCS)  
Monthly Journal, Vol. 21, No. 10, Winter 2021-2022, 1-26  
Doi: 10.30465/CRTLS.2021.31987.1926

**A Critical Review of the Book**  
**“*Quantitative Data Analysis for Language Assessment***  
***(Fundamental Techniques and Advanced Methods)*”**

**Rajab Esfandiari\***

**Abstract**

Given the significant role of quantitative data in theory building in language assessment, the need for a book covering data analysis methods is strongly felt. The two-volume book, *Quantitative Data Analysis for Language Assessment (Fundamental Techniques and Advanced Methods)*, edited by Vahid Aryadoust and Michelle Raquel and published by Routledge fills this gap. The first volume entitled *Quantitative Data Analysis for Language Assessment: Fundamental Techniques* was published in 2019, and the second volume by the title of *Quantitative Data Analysis for Language Assessment: Advanced methods* was published in 2020. The present study is critical in goal, qualitative in methodology, and emergent in analysis. The book is comprehensive in scope and content, includes high-quality, well-written chapters follows a coherent style of writing, and is rich in the number of quantitative data analysis methods. The book has some limitations, including inexplicitness of criteria for selection of quantitative data methods and failure to include more common traditional methods. The book can generate further research in this subfield of applied linguistics, expand the subfield, and contribute to theory building in

\* Associate Professor of English Language Teaching, Imam Khomeini International University, Qazvin, Iran, [esfandiari@hum.ikiu.ac.ir](mailto:esfandiari@hum.ikiu.ac.ir)

Date received: 22/08/2021, Date of acceptance: 27/11/2021

Copyright © 2010, IHCS (Institute for Humanities and Cultural Studies). This is an Open Access article. This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

۲ پژوهش‌نامه انتقادی متون و برنامه‌های علوم انسانی، سال ۲۱، شماره ۱۰، دی ۱۴۰۰

language assessment. Postgraduate students, L2 assessment practitioners, and researchers are strongly advised to read this invaluable edited book.

**Keywords:** Data Analysis, Quantitative Methods, Language Assessment, Language Testing.



## نقد و بررسی کتاب

### *Quantitative data Analysis for language assessment (Fundamental techniques and advanced methods)*

### تحلیل داده‌های کمی در سنجش زبان

### (تکنیک‌های بنیادین و روش‌های پیشرفته)

رجب اسفندیاری\*

#### چکیده

باتوجه به اهمیت و نقش داده‌های کمی در سنجش زبان، چاپ کتابی که بتواند روش‌های تحلیل داده‌ها را معرفی کند حیاتی بنظر می‌رسد. کتاب دوجلدی *تحلیل داده‌های کمی در سنجش زبان (تکنیک‌های بنیادین و روش‌های پیشرفته)* به‌منظور پرکردن این خلاء توسط وحید آریادوست (Vahid Aryadoust) و میشل راکوئل (Michelle Raquel) گردآوری شده است و انتشارات راتلج (Routledge) آن را منتشر کرده‌است. جلد اول آن با عنوان *تحلیل داده‌های کمی در سنجش زبان: تکنیک‌های بنیادین* در سال ۲۰۱۹ چاپ شد، و جلد دوم آن تحت عنوان *تحلیل داده‌های کمی در سنجش زبان: روش‌های پیشرفته* در سال ۲۰۲۰ منتشر شد. در این نوشتار، نویسنده ضمن معرفی کتاب، آن را مورد نقد هم قرار می‌دهد. مطالعه حاضر از نظر هدف انتقادی، به‌لحاظ رویکرد کیفی، و از حیث تحلیل محتوایی است. کتاب امتیازات بسیار زیادی دارد از جمله کامل بودن فصل‌ها از نظر ساختاری و محتوایی، انسجام و پیوستگی مطالب، و تنوع روش‌های تحلیل داده‌های کمی، اگرچه کاستی‌هایی هم دارد که می‌توان به‌عدم شفافیت در انتخاب روش‌ها اشاره کرد.

\* دانشیار آموزش زبان انگلیسی، عضو هیأت علمی دانشگاه بین‌المللی امام خمینی<sup>(د)</sup>، قزوین، ایران،  
esfandiari@hum.ikiu.ac.ir

تاریخ دریافت: ۱۴۰۰/۰۵/۳۱، تاریخ پذیرش: ۱۴۰۰/۰۹/۰۶

کتاب حاضر می‌تواند منشاء تولید پژوهش‌های سودمند در این حوزه باشد، به‌گسترش آن کمک شایانی بکند، و منجر به تولید علم در سنجش زبان بشود. لذا مطالعه آن به‌جامعه آزمون‌سازی و سنجش زبان موکدا توصیه می‌شود.

**کلیدواژه‌ها:** تحلیل داده‌ها، روش‌های کمی، سنجش زبان، آزمون‌سازی.

## ۱. مقدمه

داده‌ها نقش کلیدی را در علوم امروزی بازی می‌کنند، منجر به نظریه‌پردازی و باعث تأیید و یا رد ادعای صاحب‌نظران می‌شوند، و به‌گسترش رشته‌ها کمک می‌کنند. مکنلی (McKinley, 2020) داده‌ها را به سه دسته کلی در زبان‌شناسی کاربردی تقسیم‌بندی می‌کند: داده‌های کمی، داده‌های کیفی، و داده‌های ترکیبی. نقش داده‌های کمی در سنجش زبان به‌عنوان یکی از زیرشاخه‌های زبان‌شناسی کاربردی خیلی پررنگ بوده‌است. تاجایی که حتی امروزه داده‌های کمی مهمترین نوع داده در این زیرشاخه محسوب می‌شود. علت این اهمیت نیز ماهیت علمی این زیر شاخه است بخاطر اینکه با طراحی انواع آزمون‌ها و خصوصیات روانسجی آنها نظیر پایایی و درجه سختی و راحتی سؤالات، سروکار دارد (Green, 2019) و این نیز مستلزم گردآوری داده از نوع کمی آن است. به‌عقیده مکنامرا (McNamara, 2011)، ویژگی بارز سنجش زبان این است که به‌شدت تحت تأثیر دو رشته تخصصی زبان‌شناسی کاربردی و اندازه‌گیری بوده‌است، و لذا آمار و ارقام به‌عنوان بهترین نوع داده کمی در سنجش زبان مطرح می‌شود (Ockey & Green, 2020).

اهمیت داده‌های کمی در سنجش زبان بر متخصصان آن پوشیده نیست، اما داده‌های محض ارزش خاصی ندارد مگر اینکه بادقت تحلیل شوند بخاطر اینکه داده ماهیتی خام دارد و مفهوم خاصی ندارد. به‌همین علت است که همواره روش‌های آماری جدید و متنوعی متناسب با نوع داده‌های کمی وارد این زیرشاخه شده‌اند تا بتوانند نیاز آن را برآورده کنند و معایب روش‌های قبلی را مرتفع سازند. تحلیل درست و دقیق داده‌ها با استفاده از روش‌های آماری متناسب با داده منجر به‌ارائه نتایج معتبر، بحث و نتیجه‌گیری صحیح می‌شود (Ary et al. 2019).

علی‌رغم اهمیت تحلیل داده‌های کمی در سنجش زبان، شوربختانه توجه خاصی به آن‌ها نشده‌است. روش‌های تحلیل داده‌ها در سنجش زبان یا همان روش‌های رایج در

زبان‌شناسی کاربردی هستند و یا به صورت خیلی کوتاه و گذرا در قالب یک یا چند فصل در کتب زبان‌شناسی کاربردی مطرح می‌شوند و حق مطلب در مورد آن‌ها ادا نمی‌شود. حتی در اولین راهنمانامه سنجش زبان دوم توسط تاساگاری و بنرجی (Tsagari & Banerjee, 2017) و در ویرایش سوم کتاب ۶۰۰ صفحه‌ای آزمون سازی و سنجش زبان که توسط شوهایمی و اور (Shohamy & Or, 2017) منتشر شد، یک فصل هم در مورد تحلیل داده‌های کمی سنجش زبان یافت نمی‌شود. در تنها دایره‌المعارف تخصصی چهارجلدی *The Companion to Language Assessment* که توسط کونان (Kunnan, 2014) به سنجش زبان اختصاص داده شده است، فقط ۹ فصل از ۱۴۰ فصل آن این روش‌ها را توضیح می‌دهد و فصل‌ها اساساً مفاهیم نظری را در بر می‌گیرند و از جنبه‌های کاربردی و ارائه داده برای کاربرد این روش‌ها غافل‌اند. بنابراین، کتابی که به صورت منسجم و هدف‌مند بتواند روش‌های رایج تحلیل داده‌های کمی را در سنجش زبان مطرح کند و توضیح بدهد بیش از پیش احساس می‌شود. کتاب دوجلدی تحلیل داده‌های کمی در سنجش زبان (تکنیک‌های بنیادین و روش‌های پیشرفته) به همین منظور وارد بازار شده است تا بتواند خلأ موجود را پر کند. این اولین کتابی است که به صورت منسجم به بررسی داده‌های زبانی می‌پردازد، و با توجه به جدید بودن هنوز به فارسی ترجمه نشده است. لذا نقد و بررسی آن می‌تواند فرصتی برای معرفی کردن کتاب به جامعه سنجش زبان باشد.

## ۲. نگاهی به پیشینه آموزشی و پژوهشی نویسندگان

وحید آریادوست (Vahid Aryadoust) استادیار سنجش زبان در موسسه ملی تعلیم و تربیت در دانشگاه فناوریانه نانیانگ سنگاپور است. ایشان سردبیر مهمان مجله بین‌المللی شنیداری (*International Journal of Listening*) در سال ۲۰۱۶، و مجله روان‌شناسی تربیتی (*educational psychology*) در سال ۲۰۱۷ بود و عضو هیأت تحریریه مجلاتی چون *Educational Assessment* است. دارای بیش از ۵۰ مقاله چاپ شده در مجلات معتبر بین‌المللی نظیر مجلات زیر است: فصل‌نامه سنجش زبان (*Language Assessment Quarterly*)، آزمون‌سازی (*Language Testing*)، فصل‌نامه آموزش زبان انگلیسی به غیر انگلیسی‌زبانان (*TESOL Quarterly*)، و یادگیری زبان با استفاده از رایانه (*Computer Assisted Language Learning*)، و روان‌شناسی تربیتی.

این پژوهش‌گر ایرانی‌الاصول صاحب‌کتاب تألیفی مانند فرهنگ جامعه‌شناسی زبان، منظورشناسی، و زبان‌ها و نظریه‌پردازی روایی برای ساخت آزمون شنیداری با اهداف علمی است. فصل‌های متعددی را در زمینه‌های مختلفی مانند سنجش زبان، کاربرد انگاره راش در سنجش زبان، طراحی آزمون‌های شنیداری، و اعتبارسنجی سازه آزمون‌های زبانی نوشته‌است که در کتب زبان‌شناسی کاربردی توسط انتشارات معتبر بین‌المللی هم‌چون انتشارات دانشگاه کمبریج، انتشارات بلک وایلی، و انتشارات راتلج به‌چاپ رسیده است. در سال ۲۰۱۹ به‌پاس خدمات ارزنده‌ی علمی-تحقیقاتی در جامعه بین‌الملل، موفق به‌کسب جایزه‌ی بین‌المللی Intercontinental Academia Fellowship شد.

میشل راکوئل (Michelle Raquel) دانشیار مرکز مطالعات انگلیسی کاربردی در دانشگاه هنگ‌کنگ است. او بیش از ۱۵ سال سابقه‌ی تدریس دروس زبان‌شناسی کاربردی در مقاطع مختلف تحصیلات تکمیلی در دانشگاه‌های مختلف هنگ‌کنگ دارد. خانم راکوئل از سال ۲۰۱۸ سرپرست پروژه‌ی دروس فشرده‌ی ارتباطات است و سابقه‌ی طولانی در مدیریت طراحی آزمون‌های مؤسسات زبان دارد. مقالات مختلفی در مجلات معتبر بین‌المللی نظیر فصل‌نامه سنجش زبان به‌چاپ رسانده است، و نویسنده چندین فصل کتاب در کتب زبان‌شناسی کاربردی بوده‌است که توسط انتشارات معتبر بین‌المللی مانند انتشارات فرهیختگان کمبریج (Cambridge Scholars Publishing) چاپ شده‌است.

### ۳. روش‌شناسی پژوهش

مطالعه حاضر، از حیث هدف، در دسته پژوهش‌های انتقادی و، از حیث رویکرد، در زمره پژوهش‌های کیفی است. لذا از روش تحلیل محتوا (content analysis) از نوع طرح پیدایشی (emergent design) که از روش‌های جدید کیفی است استفاده می‌شود تا تحلیل دقیق‌تری از داده‌ها ارائه کند (Ary et al. 2019). این روش بیشتر برای بررسی محتوای نوشتاری استفاده می‌شود و ویژگی‌های محتوایی را توصیف می‌کند تا معنی نهفته در محتوا نمایان شود. بنابراین، تمام ۲۱ فصل این کتاب دو جلدی مورد تحلیل محتوا قرار می‌گیرد تا تصویر درست‌تر و منسجم‌تری از داده‌ها ترسیم شود.

#### ۴. معرفی اجمالی کتاب تحلیل داده‌های کمی در سنجش زبان (تکنیک‌های بنیادین و روش‌های پیشرفته)

این کتاب دوجلدی از سه بخش اصلی تشکیل شده است، و هر دو جلد یک مقدمه مشترک دارد. جلد اول تحت عنوان تحلیل داده‌های کمی در سنجش زبان: تکنیک‌های بنیادین متشکل از ۱۱ فصل نسبتاً طولانی است که در ۲۷۲ صفحه تنظیم شده است. جلد دوم کتاب با عنوان تحلیل داده‌های کمی در سنجش زبان: روش‌های پیشرفته نیز از سه بخش اصلی و ۱۰ فصل نسبتاً مفصل تشکیل شده است و ۲۴۲ صفحه است.

#### ۱.۴ جلد اول: تحلیل داده‌های کمی در سنجش زبان: تکنیک‌های بنیادین

##### ۱.۱.۴ بخش اول: آزمون‌سازی، پایایی آزمون، و تعمیم‌پذیری نمرات آزمون

نویسنده فصل اول تحت عنوان، تحلیل سؤالات آزمون‌های زبانی، ریتا گرین (Rita Green) استاد دانشگاه لنکستر انگلستان است. در این فصل، نویسنده با ۳۰ سال تجربه در پروژه‌های آزمون‌سازی در حوزه تعلیم و تربیت، به بررسی سه خصوصیت مهم سؤالات آزمون‌های زبانی می‌پردازد. گرین معتقد است که بررسی کیفیت سؤالات هر آزمون جدیدی باید قبل از اجرا شدن انجام شود، امری ضروری که به گفته‌ی وی معمولاً اتفاق نمی‌افتد. ضرورت این کار، به اعتقاد گرین، به این دلیل است که این سؤالات در واقع مبین خصوصیات یک آزمون خوب است که عبارتند از: پایایی، سطح دشواری، قدرت تمایز، و توزیع گزینه‌ها. گرین با استفاده از نرم‌افزار پژوهشی SPSS به تحلیل این سؤالات می‌پردازد، و به آن‌ها جواب می‌دهد. گرین معتقد است که این روش می‌تواند به متولیان برگزاری آزمون‌ها کمک کند تا آزمون بی‌عیب و نقصی را طراحی و اجرا کنند.

فصل دوم کتاب به نظریه تعمیم‌پذیری یک‌متغیری (Univariate Generalizability Theory) در سنجش زبان اختصاص داده شده است. نویسندگان این فصل خانم یاسوا ساواکی (Yasuyo Sawaki)، استاد تمام زبان‌شناسی کاربردی دانشگاه واسیدا ژاپن، و خانم زیومینگ زی (Xiaoming Xi)، مدیر اجرایی آموزش جهانی در سازمان جهانی آزمون‌سازی ETS، هستند. نویسندگان این فصل معتقدند که پایایی و تعمیم‌پذیری هر آزمونی به اندازه‌ی خصوصیات روانسنجی آن اهمیت دارد. لذا با استفاده از این نظریه به

بررسی پایایی آزمون و تعمیم پذیری نمرات می‌پردازند. نویسندگان با استفاده از نرم‌افزار GENOVA به تحلیل یک فعالیت خلاصه‌نویسی می‌پردازند و نتایج را بیان می‌کنند و آن‌ها را مورد بحث و بررسی قرار می‌دهند. نویسندگان نتیجه می‌گیرند نقاط قوت این نظریه نسبت به نظریه کلاسیک اندازگیری نمره حقیقی قابل مقایسه نیست بخاطر این‌که نظریه مذکور می‌تواند هم‌زمان چند عامل تأثیرگذار بر روی نمرات زبان‌آموزان را تحلیل کند و نتایج معتبرتری از توانایی‌های آنها به ما بدهد.

فصل سوم در ارتباط با نظریه تعمیم‌پذیری چندمتغیره (Multivariate) در سنجش زبان است. کایری سی گرابوسکی (Kirby C. Grabowski)، استادیار زبان‌شناسی دانشکده تعلیم و تربیت دانشگاه کلمبیا در ایالات متحده آمریکا، و ران چان لین (Rongchan Lin)، استادیار دانشگاه فناوریانه نان‌یانگ سنگاپور، هستند که تجربه‌های ارزنده‌ای در کاربردی کردن این نظریه دارند. نویسندگان از آزمون شنیداری-گفتاری استفاده کردند تا بتوانند داده‌ها را در سنجش زبان با استفاده از نرم‌افزار mGENOVA تحلیل کنند، نتایج را ارائه بدهند، و آنها را بحث و بررسی کنند. اگرچه نویسندگان معتقدند که این نظریه مشکلات نظریه تعمیم‌پذیری یک‌متغیری را مرتفع می‌کند، نظریه بی‌عیب و نقص نیست. مشکل اساسی نظریه، که نویسندگان نیز در نتیجه‌گیری به آن اشاره می‌کنند، این است که تحلیل کلی از متغیرها را فراهم می‌کنند و قادر به تحلیل در سطوح جزئی‌تر نیست. لذا توصیه می‌کنند محققان از روش‌های تکمیلی دیگری نظیر انگاره چندوجهی راش (Many-Facet Rasch Measurement) استفاده کنند.

#### ۲.۱.۴ بخش دوم: انگاره یک‌بعدی راش (Unidimensional Rasch Model)

فصل چهارم در این بخش به استفاده از انگاره راش در سنجش زبان: یک‌بعدی بودن (Unidimensionality) و استقلال موضعی (Local Independence) اختصاص یافته است. در این فصل، جیسون فان (Jason Fan)، استاد پژوهشی مرکز مطالعات سنجش زبان دانشگاه ملبورن استرالیا، و ترور باند (Trevor Bond)، استاد سرشناس سنجش و اندازه‌گیری دانشگاه جیمز کوک استرالیا، انگاره راش را تشریح می‌کنند. فان و باند معتقدند که برای فهم این انگاره، باید دو الزام مهم آن یک‌بعدی بودن و استقلال موضعی که، به نقل قول از نویسندگان، "در پیشینه تحقیقات قبلی سنجش زبان معمولاً نادیده گرفته شده است" (ص. ۹۹) رعایت شود. در ادامه فصل، نویسندگان با استفاده از انگاره امتیازدهی پاره‌ای



راش (Partial Credit Rasch Model) و نرم‌افزار Winsteps به بررسی یک‌بعدی بودن و استقلال موضعی آزمون شنیداری که بخشی از آزمون بسندگی انگلیسی بود و در دانشگاه فودان چین ساخته و برگزار شد، می‌پردازند. توصیه نویسندگان این است که برای به حداقل رساندن عوامل غیرمرتبط بر روایی سازه (Construct Validity) هر آزمونی و کمک به خوانندگان برای فهم بهتر نتایج، بهتر است محققان نتایج مربوط به این دو الزام انگاره راش را گزارش بدهند.

در فصل پنجم، میشل راکوئل، دانشیار مرکز مطالعات انگلیسی کاربردی دانشگاه هنگ‌کنگ، استفاده از انگاره‌ی راش را در تحلیل کارکرد افتراقی سؤال (Differential Item Functioning) در تحقیقات مربوط به سنجش زبان مورد بحث قرار می‌دهد. این فصل در اصل ادامه‌دهنده‌ی بحث مربوط به ویژگی‌های انگاره‌ی راش است که در فصل قبلی معرفی شد. نویسنده کارکرد افتراقی سؤالات قسمت شنیداری آزمون دلتا (DELTA) را با استفاده از انگاره‌ی راش بررسی می‌کند. نتایج بررسی مشخص می‌کند که اگرچه تعدادی از سؤالات این آزمون تشخیصی کارکرد افتراقی داشتند، نتایج آزمون تحت تأثیر این سؤالات معیوب قرار نگرفتند.

در دو فصل قبلی، انگاره‌ی ساده راش، ویژگی‌های این انگاره، و دو الزام مهم آن مورد بررسی قرار گرفت. انگاره‌ی ساده یکی از انگاره‌های خانواده راش است. در این فصل، ایکیو چوی (Ikkyu Choi) کاربرد دو انگاره‌ی دیگر از خانواده راش به نام‌های انگاره‌ی مقیاس نمره‌دهی راش (Rasch Rating Scale Model) و انگاره‌ی امتیازدهی پاره‌ای راش (Partial Credit Rasch Model) را در مطالعات سنجش زبان معرفی می‌کند و با هم مقایسه می‌کند. نویسنده با استفاده از یک پرسشنامه انگیزشی، به بررسی ساختار پرسشنامه با استفاده از این دو انگاره می‌پردازد. با استفاده از نمودارهای گرافیکی، جداول، و ارائه آمارهای فنی، نویسنده نشان می‌دهد که انگاره‌ی امتیازدهی پاره‌ای راش برای بررسی ساختار پرسشنامه مناسب‌تر است. در خاتمه، نویسنده به یک جمع‌بندی خیلی جالب می‌رسد، و جمله زیبایی دارد، اینکه با توجه به قدمت این دو انگاره، "می‌توان باطمینان گفت که آنها بخش جدایی‌ناپذیر روش‌های تحلیل داده‌های چندحالتی در مطالعات سنجش زبان شده‌اند" (ص. ۱۴۹).

فصل آخر در این بخش به‌انگاره‌ی چندوجهی راش و کاربردهای آن در سنجش زبان تعلق دارد، انگاره‌ای که از آن به‌عنوان پرکاربردترین انگاره‌ی خانوادگی راش (Aryadoust, Ng, & Sayama, 2020) نام برده می‌شود و مکنامرا (McNamara, 2011: 436) پیدایش آن را به "جهش کوانتومی" (Quantum Leap) تعبیر می‌کند. تامس اکیز (Thomas Eckes)، با تجربه‌ی چندین ساله‌ی عملی و تحقیقاتی خود در این زمینه، به‌معرفی این انگاره می‌پردازد. نویسنده بعد از معرفی و توضیح آن و مرور مطالعات مرتبط با سنجش زبان، تلاش می‌کند تا با استفاده از داده‌های آماده‌ی یک آزمون نوشتاری کاربرد این انگاره را نشان بدهد. برای این کار از نرم‌افزار FACETS که به‌همین منظور طراحی شده است، استفاده کرد. نویسنده نشان می‌دهد که چگونه می‌توان با استفاده از این انگاره به تحلیل سختگیری مصححان درحین نمره‌دهی پرداخت و یکپارچگی (Consistency) نمره‌دهی آنها را بررسی کرد. دو کاربردی که با انگاره‌های قبلی خانواده راش امکان پذیر نیست. در صفحات پایانی فصل، اکیز به پیشرفت‌هایی که در این زمینه صورت گرفته است و کاربردهای عملی که این انگاره دارد اشاره می‌کند. او همچنین به‌موارد بحث‌برانگیز استفاده از این انگاره گریزی می‌زند و این فصل را با معرفی نرم‌افزارهای دیگری همچون ConQuest جهت تحلیل داده‌های چندحالتی خاتمه می‌دهد.

#### ۳.۱.۴ بخش سوم: تحلیل آماری یک‌متغیری و چندمتغیری

در فصل هشتم از بخش سوم، تفاوت بین گروه‌ها با استفاده از آزمون تی ( $t$ .test) و تحلیل واریانس (Analysis of Variance) در سنجش زبان بحث می‌شود. تقبا آلیف تبرک (Tuğba Elif Toprak)، نویسنده‌ی این فصل، این دو آزمون رایج و پرکاربرد را با استفاده از مثال‌های متعدد و دو مجموعه داده (Dataset) توضیح می‌دهد. نویسنده ابتدا آزمون تی و سپس تحلیل واریانس را معرفی می‌کند، پیش‌فرض‌های آنها را توضیح می‌دهد، مفاهیم آماری مرتبط با این دو آزمون را تشریح می‌کند، به آزمون‌های غیرپارامتریک (Non-parametric Tests) مانند Mann-Whitney U Test, Kruskal-Wallis Test, و Friedman's Test اشاره می‌کند که باید استفاده شود اگر پیش‌فرض‌های آزمون تی و تحلیل واریانس نقض شوند، و مطالعاتی را که از این دو آزمون را در سنجش زبان استفاده کرده‌اند مرور می‌کند. توضیحات مبسوط نویسنده باعث می‌شود شبهاتی را که خواننده در مورد این دو آزمون دارد برطرف شود. نویسنده در قسمت دوم مقاله با استفاده از این دو آزمون در

دو مجموعه داده‌ی متفاوت مرحله به مرحله خواننده را با کاربرد آنها آشنا می‌کند. با استفاده از یک آزمون تلفظ، تأثیر آهنگ‌های انگلیسی را بر روی یادگیری تلفظ زبان‌آموزان ترکی بررسی می‌کند تا بتواند کاربرد آزمون تی را نشان بدهد، و با استفاده از یک آزمون درک مطلب سطح توانایی درک مطلب را در میان دانشجویان ترکیه‌ای سال اول، سال دوم، سال سوم، و سال چهارم دانشگاه بررسی می‌کند تا کاربرد تحلیل واریانس را به‌خواننده توضیح بدهد. اگرچه اکثر خوانندگان ممکن است با این دو آزمونی که در این فصل مطرح شد آشنایی نسبی داشته باشند، نحوه‌ی ارائه آنها، بکارگیری زبان ساده، و استفاده از سناریوهای آشنا باعث فهم بهتر آنها برای خواننده می‌شود.

ژی لی (Zhi Li) و میشل وای چن (Michelle Y. Chen) نویسندگان فصل نهم هستند و تلاش می‌کنند کاربرد تحلیل کوواریانس (Analysis of Covariance) و تحلیل کوواریانس چندمتغیره (Multivariate Analysis of Covariance) را در سنجش زبان توضیح بدهند. نویسندگان استدلال می‌کنند که هدف از هر دوی این دو آزمون این است که تأثیر متغیرهای کمکی را کنترل کنند، خصوصیتی که در تحلیل واریانس وجود ندارد. نویسندگان ابتدا پیش‌فرض‌های نسبتاً زیاد این دو آزمون را توضیح می‌دهند. با استفاده از بخش خواندن آزمون PISA (برنامه‌ی سنجش دانش زبانی دانش‌آموزان بین‌المللی)، نویسندگان با استفاده از نرم‌افزار محبوب و پرکاربرد SPSS مراحل اجرای این دو آزمون آماری را برای خوانندگان توصیف می‌کنند. بر اساس پیشینه‌ی تحقیق و تحلیلی که در این فصل انجام دادند، لی و چن نتیجه می‌گیرند که در طرح‌های تحقیقاتی شبه‌تجربی و هم‌بستگی (Quasi-experimental and Correlational Research Plans)، بهتر است بجای استفاده از تحلیل واریانس از این دو آزمون استفاده شود چون در این نوع طرح‌ها معمولاً متغیرهای کمکی وجود دارند و روی متغیر وابسته تأثیر می‌گذارند و نتایج را ممکن است تغییر بدهند. علی‌رغم اهمیت و مفید بودن این دو آزمون، به اعتقاد نویسندگان، در مقایسه با آزمون تی و تحلیل واریانس کمتر از آنها برای مقایسه میانگین نمرات گروه‌ها استفاده می‌شود.

فصل دهم کاربرد رگرسیون خطی (Linear Regression) در سنجش زبان را بررسی می‌کند. این فصل را داریانگ سیو (Daeryong Seo) و حسین تاهربای (Husein Taherbhai) نوشته‌اند. در دو فصل قبلی در این بخش، تمرکز اصلی بر روی تفاوت میانگین نمرات گروه‌ها بود، اما در این فصل هدف این است تا تأثیر پیش‌بینی‌کننده‌ی (Predictor)

یک یا چند متغیر را بر روی متغیر وابسته مورد بررسی قرار بگیرد. به‌همین دلیل، دو نوع مهم رگرسیون خطی ساده (Simple Linear Regression) و رگرسیون خطی چندمتغیری (Multiple Linear Regression) مطرح می‌شود و تفاوت‌های بین آن دو و پیش‌فرض‌های آن‌ها از جمله خطی بودن (Linearity)، همبستگی متغیرهای مستقل (Multicollinearity)، همپراکنشی (Homoscedasticity)، توزیع نرمال داده‌ها (Normality) توضیح داده می‌شود. نمونه‌ی داده برای بررسی رگرسیون خطی که در این فصل استفاده شد بخشی از داده‌های آزمون بسندگی (Proficiency) بود که بر روی ۵۰۰ زبان‌آموز اجرا شده بود. بعد از تحلیل داده‌ها، ارائه و بحث نتایج، نویسندگان به‌نکات جالبی در استفاده از رگرسیون خطی اشاره می‌کنند که معمولاً در کتب تحقیق زبان‌شناسی کاربردی مطرح نمی‌شوند. نکته اول، در استفاده از رگرسیون مطلبی که بیش از جنبه‌های ریاضی آن اهمیت دارد دانش عمیقی است که محقق باید در انتخاب متغیرها داشته باشد. دوم اینکه واقعاً نیازی نیست پژوهش‌گران نگران نقض برخی از پیش‌فرض‌های رگرسیون خطی باشند اگر صرفاً از این آزمون‌ها جهت پیش‌بینی استفاده شود چون این آزمون آماری به‌اندازه‌ی کافی استوار (Robust) است که نقض برخی از پیش‌فرض‌ها نتواند ضرایب رگرسیون (Coefficients) را تحت تأثیر قرار بدهد.

در فصل پایانی جلد اول، نویسندگان کاربرد تحلیل عاملی اکتشافی (Exploratory Factor Analysis) را در سنجش زبان تشریح می‌کنند. لیمی ژانگ (Limei Zhang) و ونشو لو (Wenshu Luo) نویسندگی این فصل را برعهده دارند. در ابتدای فصل، معرفی خیلی کوتاهی از تحلیل عاملی به‌عنوان روش آماری چندمتغیری و یکی از روش‌های داده‌کاهی (Data Reduction) و دو نوع مهم آن—تحلیل عاملی اکتشافی و تحلیل عاملی تأییدی (Confirmatory) ارائه می‌شود. سپس، تحلیل عاملی اکتشافی و پنج مرحله‌ی مهم در اجرای آن با جزئیات کامل توضیح داده می‌شود. نویسندگان در ادامه توضیح می‌دهند که غایت اصلی تحلیل عاملی اکتشافی شناسایی عوامل اساسی است که برای سنجش صفت مکنون (Latent Trait) در پرسشنامه و یا آزمون استفاده می‌شود، در واقع روایی سازه‌ی هر ابزاری که برای سنجش آزمودنی‌ها (Testees) مورد استفاده قرار بگیرد. به‌همین منظور، مؤلفین با استفاده از پرسشنامه‌ی راهبردهای خواندن (Reading Test Strategy Use Questionnaire) نسبت به‌شناسایی عوامل اساسی راهبردهای خواندن دانشجویان چینی می‌پردازند. نتایج تحلیل عاملی هفت راهبرد اساسی شامل راهبردهای برنامه‌ریزی،

ارزیابی، نظارتی، تلفیقی، استنباطی، تصاعدی، و شناسایی اطلاعات را نشان می‌دهد. قسمت پایانی فصل به‌استفاده درست تحلیل عاملی اکتشافی اختصاص دارد که به پژوهش‌گران توصیه می‌شود در کارگاه‌های مرتبط شرکت کنند و مقالات جدید را مطالعه کنند تا بهتر بتوانند داده‌ها را با استفاده از این آزمون آماری پیچیده تحلیل کنند.

## ۲.۴ جلد دوم: تحلیل داده‌های کمی در سنجش زبان: روش‌های پیشرفته

### ۱.۲.۴ بخش اول: انگاره‌های پیشرفته سؤال - پاسخ در سنجش زبان

آغازکننده‌ی بخش اول از جلد دوم فصلی است درباره‌ی استفاده از انگاره‌ی ترکیبی راش (Mixed Rasch Model) برای سنجش خواندن و درک مطلب. پوریا بقایی (Purya Baghaei) کریستوف جی کمپر (Christoph J. Kemper)، ساموئل گریف (Samuel Greif)، و مونیک رایهارت (Monique Reichert) در مورد این انگاره توضیح می‌دهند. قسمت اول مقاله با بررسی مطالعات خیلی کمی که در سنجش زبان با استفاده از این انگاره صورت گرفته است به‌پایان می‌رسد. قسمت دوم فصل به تحلیل مجدد داده‌های بخش خواندن و درک مطلب آزمون سراسری (کنکور) اختصاص داده شده است که داده‌های آن در سال ۲۰۱۱ گردآوری شده بود. نویسندگان با استفاده از انگاره‌ی ترکیبی راش دو طبقه پنهان آزمودنی شناسایی می‌کنند که سطح دشواری سؤالات برای آنها متفاوت است. نویسندگان ادعا می‌کنند که این انگاره طبقات پنهانی را شناسایی می‌کند که آزمودنی‌ها به آنها تعلق دارند ادعایی که نتایج تحلیل داده‌ها آن را اثبات می‌کند. نویسندگان اینطور نتیجه می‌گیرند که می‌توان از این انگاره برای شناسایی چندبعدی بودن (Multidimensionality) آزمون‌ها استفاده کرد، البته ایراد بزرگی که به این انگاره وارد است، و آنها نیز به آن اشاره می‌کنند، این است که به داده‌های آماری زیادی برای تحلیل نیاز دارد.

انگاره‌های چندبعدی (Multidimensional) راش در آزمون‌های شنیداری زبان اول عنوان فصل دوم است. کریستین اسپودان (Christian Spoden) و جانز فلایشر (Fleischer Jens) در مورد این انگاره توضیح می‌دهند. در ابتدای فصل، در مورد پیدایش انگاره‌های چندبعدی در سنجش زبان صحبت می‌شود و علت ورود آن به سنجش زبان را ناکارآمدی انگاره‌ی یک‌بعدی راش می‌دانند بخاطر اینکه زبان ماهیتی چندبعدی و پیچیده دارد و اعمال این انگاره‌های ساده در سنجش عملکرد آزمودنی‌ها ممکن است

عوامل زبانی دیگری که توانایی آزمودنی را تحت تأثیر قرار می‌دهد، نادیده بگیرد و استدلال می‌کنند که روش‌های چندبعدی مناسب‌تر هستند و بهتر می‌توانند متغیرهای پیچیده زبانی را تبیین کنند. با استفاده از رویکرد تعلیمی (Didactic)، به توضیح در مورد این انگاره پرداخته می‌شود، و کاربرد آن در ساختار روان‌سنجی یک آزمون شنیداری تشریح می‌شود. نویسندگان به صورت مبسوط مزایای استفاده از آن را در ساختار این آزمون توصیف می‌کنند و توضیح می‌دهند که چطور از آن می‌شود برای محاسبه‌ی دقیق‌تر پایایی آزمون شنیداری، و سایر آزمون‌ها، استفاده کرد. این فصل با معرفی کوتاه سایر انگاره‌های چندبعدی راش و استفاده‌ی آنها در سنجش زبان به پایان می‌رسد.

در فصل سوم، یکی دیگر از انگاره‌های چندبعدی مطرح می‌شود. انگاره‌ی شناختی تشخیصی لگاریتم خطی (Log-linear Cognitive Diagnosis Modeling) به‌نوعی از انگاره‌ی چندبعدی غیر راش اطلاق می‌شود که ویژگی‌های دو علم روان‌سنجی و روان‌شناسی را باهم تلفیق می‌کند و فرق آن با انگاره‌های چندبعدی راش در این است که این انگاره تسلط آزمودنی‌ها به مهارت‌های فرعی (Subskills) را اندازه‌گیری می‌کند. بعد از بحث مفصلی در باره‌ی ویژگی‌های ریاضی و مبانی نظری انگاره، نویسندگان مطالعاتی که با استفاده از آن در سنجش زبان انجام شده است را مرور می‌کنند. در ادامه، برای فهم بیشتر، با استفاده از یک آزمون شنیداری کاربردها و خصوصیات آن را نشان می‌دهند و بر اساس تحلیل داده‌ها، ارائه و بحث نتایج، ادعا می‌کنند که یکی از مهم‌ترین کاربردهای آن ارائه اطلاعات تشخیصی در مورد توانایی سؤالات آزمون است. پایان‌بخش این فصل محدودیت‌های این انگاره است که به آنها به صورت گذرا اشاره می‌شود.

حمدالله راوند (Hamdollah Ravand) کاربرد انگاره‌ی سلسله‌مراتبی تشخیصی - دسته‌بندی (Hierarchical Diagnostic Classification Model) را در سنجش خواندن و درک مطلب توضیح می‌دهد. مانند انگاره‌ی شناختی تشخیصی لگاریتم خطی در فصل قبلی، انگاره‌ی سلسله‌مراتبی تشخیصی - دسته‌بندی یکی دیگر از روش‌های شناختی تشخیصی است که در سال‌های اخیر در سنجش زبان دوم اهمیت پیدا کرده است. راوند پس از معرفی انواع روش‌های تشخیصی دسته‌بندی و استفاده آنها در سنجش زبان دوم، به معرفی و توضیح مفاهیمی اساسی نظیر انتخاب و برازش مدل می‌پردازد. نویسنده در ادامه بحث می‌کند که این انگاره برای بررسی آزمون‌های هنجاری (Norm-referenced) مانند

آزمون کارشناسی ارشد آموزش زبان انگلیسی که هر ساله در ایران برگزار می‌شود، بسیار مناسب است بخاطر اینکه می‌تواند روابط بین متغیرهایی که عملکرد آزمودنی را توضیح می‌دهد شناسایی کند و هم اینکه وابستگی متغیرها نسبت به همدیگر را کشف کند. مقاله با ارائه راه‌کارهای کاربردی برای انتخاب مدل جهت تحلیل داده‌ها با استفاده از این انگاره به پایان می‌رسد.

#### ۲.۲.۴ بخش دوم: روش‌های آماری پیشرفته در سنجش زبان

در این بخش، چهار روش آماری پیشرفته در سنجش زبان دوم بحث می‌شوند که یک خصوصیت مهم مشترک دارند—ایجاد رابطه علی-معلولی در میان متغیرها. به‌عنوان مثال، ژولیان ژو (Xuelian Zhu)، میشل راکوئل، و وحید آریادوست مدل‌سازی معادلات ساختاری (Structural Equation Modelling) را در پیش‌بینی عملکرد آزمودنی‌ها در آزمون‌های بسندگی زبان انگلیسی بررسی می‌کنند. با استفاده از این روش، نویسندگان به بررسی دو آزمون مهم یعنی DELTA و IELTS (آزمون‌های مهم بسندگی زبان انگلیسی) می‌پردازند و می‌خواهند به این پرسش پاسخ بدهند که آیا نمراتی که آزمودنی‌ها از آزمون DELTA کسب می‌کنند می‌تواند نمراتی که آنها در آزمون IELTS می‌گیرند، پیش‌بینی کند. علاوه بر کاربرد پیش‌بینی‌کنندگی، نویسندگان معتقدند که مدل‌سازی معادلات ساختاری را می‌توان برای اعتبارسنجی روایی سازه‌ی آزمون‌ها و تأیید فرضیه‌های قبلی استفاده کرد.

فصل ششم روشی را برای مدل‌سازی رشد عمومی معرفی می‌کند که اصطلاحاً چارک رشد عمومی دانش‌آموزان (Student Growth Percentile) و یا SGP نام دارد و برای محاسبه‌ی آن از روش رگرسیون چارکی (Quantile Regression) استفاده می‌شود. ویژگی بارز این روش این است که توانایی آزمودنی‌ها را با توانایی آن دسته از آزمودنی‌هایی که عملکرد یکسانی در آزمون داشته‌اند، مقایسه می‌کند. به این معنی که حتی زمانی که نمرات آزمون فعلی برای دو آزمودنی با سوابق آزمون‌ی متفاوت یکسان باشد، نمرات واقعی SGP آنها به هم فرق دارد (نویسندگان در صفحه ۱۲۸ مثالی کاربردی ارائه می‌دهند که به علت کمبود فضا از آن صرف‌نظر می‌شود). ویژگی بارز دیگر این روش که آن را از روش‌های مشابه مانند منحنی رشد عمومی نهفته (Latent Growth Curve) و مدل‌سازی چندسطحی (Multilevel Modelling) متمایز می‌کند این است که نیاز به هم‌ارزسازی آزمون

(Test Equating) ندارد، که خود آن یک فرایند وقت‌گیر است. سیو و تاهربای چارک رشد عمومی دانش‌آموزان را در ارزشیابی تکوینی (Formative Assessment) آزمون بسندگی زبان انگلیسی با استفاده از روش چارک رگرسیونی بررسی می‌کنند. برای این کار، از داده‌های آزمون بسندگی که بین سال‌های ۲۰۰۵ و ۲۰۰۹ بر روی دانش‌آموزان ابتدایی آمریکایی برگزار شده بود استفاده کردند و نشان دادند که چگونه می‌توان گرایش آزمودنی‌ها برای کسب نمره‌ی GSP که لازمه گرفتن نمره برش است، را اندازه گرفت. نویسندگان براین باورند که این روش، روش خوبی برای بررسی تغییراتی است که در چهار مهارت شنیداری، گفتاری، خواندن، و نوشتن در طی بازدهی طولانی‌مدت صورت می‌گیرد.

در فصل هفتم، یو اینامی (Yo In'nam) و خالد بارکاوی (Khaled Barkaoui) مدل‌سازی چندسطحی را برای بررسی منابع تغییرپذیری نمرات آزمون زبان انگلیسی معرفی می‌کنند. نویسندگان در مقدمه‌ای کوتاه با استفاده از یک نمودار گرافیکی سعی می‌کنند تا به این روش ورد پیدا کنند. سپس، این روش را در سنجش زبان دوم تشریح می‌کنند و گله‌مند هستند از اینکه از این روش به‌ندرت در این حوزه استفاده شده است. در ادامه، منطق ریاضی این روش توضیح داده می‌شود. در قسمت دوم فصل، با استفاده از این روش به تحلیل مجدد بخشی از داده‌های مطالعه کوزامی (Koizumi) و اینامی می‌پردازند تا رابطه بین واژگان و توانایی گفتاری را بررسی کنند. در ادامه، به ملاحظاتی در استفاده از این روش نظیر حجم داده‌ها، داده‌های گم‌شده (Missing Data)، و نوع نرم‌افزار اشاره می‌کنند. در خاتمه، رهنمودهایی برای استفاده از این روش برای مطالعات آتی در حوزه سنجش زبان دوم ارائه می‌دهند.

فصل پایانی به‌نوعی ادامه دهنده‌ی فصل قبلی در این بخش است بخاطر اینکه از روش مدل‌سازی چندسطحی برای تحلیل داده‌ها استفاده می‌شود با این تفاوت که نوع داده‌ها در این فصل متفاوت است و نویسندگان داده‌های طولی را (Longitudinal Data)، داده‌هایی که در بلند مدت گردآوری می‌شود، با این روش بررسی می‌کنند. به‌همین خاطر، بارکاوی و اینامی، نویسندگان این فصل، مجموعه داده‌های گردآوری‌شده از آزمون انگلیسی پیرسون (Pearson Test of English) را در فاصله سال‌های بین آوریل ۲۰۱۴ و می ۲۰۱۵ بررسی کردند تا تغییراتی که در نمرات آزمودنی‌ها ایجاد شده بود مورد ارزیابی قرار بگیرد.



با استفاده از نرم افزار HLM6 داده‌ها تحلیل می‌شوند، نتایج مطالعه با جزئیات ارائه می‌شوند، و در انتهای فصل به محدودیت‌های مطالعه اشاره می‌کنند. این فصل در حالی به پایان می‌رسد که نویسندگان سؤالاتی را مطرح می‌کنند که می‌تواند مبنای تحقیقات بعدی قرار بگیرد.

### ۳.۲.۴ بخش سوم: روش‌های داده‌کاوی الهام‌گرفته از طبیعت ( Nature-Inspired Data-Mining Methods) در سنجش زبان

جلد دوم کتاب در حالی به پایان می‌رسد که در این بخش پایانی دو روش پیشرفته داده‌کاوی معرفی می‌شوند که در سنجش زبان بسیار تازگی دارند و به ندرت می‌توان در مورد آن‌ها در مجلات سنجش زبان و یا حتی کتب زبان‌شناسی کاربردی مطلبی پیدا کرد. داده‌کاوی به گروهی از روش‌های آماری پیشرفته اطلاق می‌شود که برای شناسایی الگوهای پنهان (Hidden Patterns) در داده‌ها نظیر روابط خطی و غیرخطی بکار می‌رود (Aryadoust & Goh, 2020). کاربردهای مهم داده‌کاوی عبارتند از خوشه‌بندی (Clustering)، طبقه‌بندی (Classification)، پیش‌بینی، و خلاصه‌سازی (Winne & Baker, 2013). در این بخش، دو روش داده‌کاوی به‌خواننده معرفی می‌شود.

در فصل نهم، آریادوست و گه روش درخت رگرسیونی و طبقه‌بندی (Classification and Regression Trees) را برای بررسی پیش‌بینی سطح دشواری سؤالات آزمون شنیداری به‌جامعه‌ی سنجش زبان معرفی می‌کنند. به‌همین منظور، ابتدا مقدمه‌ی کوتاهی بیان می‌کنند و مفاهیمی چون داده‌کاوی، طبقه‌بندی، و رگرسیون را توضیح می‌دهند. بعد از این مقدمه‌ی کوتاه، با استفاده از یک مثال فرضی به‌تیبین این روش می‌پردازند و مفاهیم بنیادی آن از جمله اعتبارسنجی متقابل (Cross-validation)، برازش، و شاخص اهمیت متغیر (Variable Importance Index) را بحث می‌کنند. در ادامه تنها دو مطالعه‌ی که در این زمینه در سنجش زبان انجام شده‌است را خلاصه می‌کنند. سپس، با استفاده از این روش سطح دشواری ۳۲۱ سؤال آزمون شنیداری را بررسی می‌کنند. با استفاده از تحلیل داده‌ها به‌مجموعه‌ای از حکم‌های اگر-آن‌گاه (If-Then rules) دست می‌یابند که سطح دشواری را به‌ویژگی‌های زبانی مرتبط می‌سازد. به‌علت جدید بودن آن در سنجش زبان، نویسندگان پیشنهاد می‌کنند تا از آن در محیط‌های دیگر استفاده شود تا بتوان به‌نتایج قابل تعمیم دست پیدا کرد.

پایان‌بخش این بخش از جلد دوم کتاب فصلی است با عنوان رگرسیون نمادین مبتنی بر الگوریتم تکاملی (Evolutionary Algorithm-Based Symbolic Regression) جهت تعیین رابطه‌ی بین خواندن و دانش دستوری-واژگانی. آریادوست، نویسنده‌ی این فصل، مبحثی را مطرح می‌کند که فقط خود او توانسته است در حوزه‌ی سنجش زبان در مجلات غیرزبان‌شناختی مقاله چاپ کند. از این رو، این فصل می‌تواند شروع خیلی خوبی برای تحقیقاتی از این دست در سنجش زبان باشد. به دلیل جدید بودن آن، آریادوست تمام تلاش خود را بکار می‌گیرد تا مفاهیم و اصطلاحات فنی این روش را با زبان ساده‌ای برای خواننده توضیح بدهد و در این مسیر رنج بسیار زیادی هم متحمل می‌شود. نویسنده با استفاده از تحلیل داده‌های یک آزمون بسندگی با این روش و مقایسه‌ی آن با یک روش رگرسیون خطی نشان می‌دهد که این روش مناسب‌تری برای پیش‌بینی توانایی خواندن آزمودنی‌ها است و نتایج دقیق‌تری را ارائه می‌کند. درخاتمه، نویسنده هشدار می‌دهد که نباید مجذوب معادلات پیچیده‌ی چنین روش‌هایی شد و عاقلانه‌تر این است که پس از برازش داده‌ها با این روش، آن را با روش‌های مرسوم رگرسیون خطی مقایسه کرد تا تصویر دقیق‌تری از داده‌ها بدست آوریم.

## ۵. امتیازات کتاب

این کتاب دوجلدی ویژگی‌های منحصر به فردی دارد که در قسمت‌های بعدی توضیحات مبسوطی در مورد آنها داده می‌شود.

## ۱.۵ جامع بودن کتاب

همانطور که در قسمت‌های پیشین نیز به آن اشاره شد، کتاب حاضر اولین کتابی است که در زمینه تحلیل داده‌های کمی در سنجش زبان دوم نوشته شده است. چاپ این کتاب دوجلدی در سال ۲۰۱۹ و ۲۰۲۰ از این حیث مهم است که سنجش زبان در حال حاضر به‌عنوان یکی از زیرشاخه‌های مهم زبان‌شناسی کاربردی مطرح است که از نظر موضوعی، روش‌شناسی، و تحلیل آماری متفاوت از سایر زیرشاخه‌های زبان‌شناسی کاربردی است. علاوه بر این، متغیرهای که در سال‌های اخیر در این زیرشاخه مورد تحلیل قرار گرفته‌اند به مراتب پیچیده‌تر از قبل است. بنابراین، چاپ این کتاب، فرصتی طلایی را برای پژوهشگران

فراهم می‌کند تا خود را با انواع آزمون‌های آماری رایج (اعم از ساده و پیشرفته) آشنا کنند تا بتوانند داده‌های سنجش زبان را با آگاهی بیشتر و بهتر تحلیل کنند.

چاپ این کتاب باعث می‌شود تا دانشجویان تحصیلات تکمیلی اعم از کارشناسی ارشد و دکتری در حوزه‌ی زبان‌شناسی کاربردی منبع مستقلی در مورد روش‌های تحلیل داده‌های کمی در سنجش زبان داشته باشند تا با این روش‌ها آشنایی پیدا کنند و در موقع لزوم به‌عنوان مرجعی از آن استفاده کنند. این کتاب تقریباً تمامی روش‌های تحلیل انواع داده‌های کمی را پوشش می‌دهد. از طرفی، روش‌هایی مطرح می‌شوند که داده‌های رایج و کلاسیک شامل داده‌های اسمی، ترتیبی، و فاصله‌ای را تحلیل می‌کنند. از طرفی دیگر، ما شاهد روش‌هایی هستیم که داده‌های پویاتری مانند داده‌های طولی، آشیانه‌ای، و سلسله‌مراتبی را آنالیز می‌کنند. روش‌هایی توضیح داده می‌شوند که قادر به تحلیل داده‌های آزمون‌های پرمخاطره (High-stakes) بین‌المللی نظیر IELTS هستند. روش‌هایی که در کتاب مطرح شده‌اند به تحلیل روابط علی-معلولی داده‌ها، همبستگی بین متغیرها، معادلات ساختاری داده‌ها، پیش‌بینی توانایی آزمودنی‌ها، روابط خطی و غیرخطی متغیرها، خوشه‌بندی داده‌ها، کارکرد افتراقی سؤالات، تحلیل مسیر، و تحلیل اکتشافی و عاملی داده‌ها می‌پردازند.

نویسندگان هم روش‌های مرسوم تحلیل داده‌ها نظیر تحلیل واریانس و هم روش‌های پیشرفته‌ی آماری مانند مدل‌سازی معادلات ساختاری را با وسواس خاصی برای خوانندگان توضیح می‌دهند. شاید مهم‌ترین بخش کتاب دو فصل آخر جلد دوم باشد که نویسندگان تمام همت خود را بکار بسته‌اند تا پیچیده‌ترین و پیشرفته‌ترین روش‌های تحلیل داده‌ها نظیر روش‌های بهینه‌سازی را توضیح بدهند، روش‌هایی که بیشتر در هوش‌های مصنوعی، علوم کامپیوتری، و علم ژنتیک کاربرد دارند. اما نویسندگان با جسارت و خلاقیت توانسته‌اند آنها را در سنجش زبان مطرح کنند تا هم مبنای تحقیقات بعدی در این حوزه قرار بگیرد و هم بتواند داده‌های جدیدتر و متنوع‌تری را در سنجش زبان تحلیل کنند.

## ۲.۵ کامل بودن بخش‌ها و فصل‌های کتاب به لحاظ ساختاری و محتوایی

گردآورندگان این کتاب دو جلدی هر جلد را به سه بخش کلی تقسیم‌بندی کرده‌اند که عبارتند از نظریه‌های اندازه‌گیری، روش‌های آماری، و داده‌کاوی. این سه بخش در عین

کلی‌بودن کامل نیز هستند بخاطر اینکه گردآورندگان را قادر ساخته‌است تا بتوانند از طرفی انواع روش‌های تحلیل داده را زیر این سه بخش مدیریت کنند و از طرفی دیگر این امکان را فراهم ساخته‌است تا به‌نوعی پوشش‌دهنده‌ی روش‌های تحلیل داده در سنجش زبان باشند. هر فصل تصویر کاملی از روش تحلیل داده ارائه می‌کند. هر فصل با مقدمه‌ی نسبتاً کوتاهی برای ورود به بحث آغاز می‌شود و با توضیح مفاهیم نظری مرتبط با روش ادامه پیدا می‌کند. سپس، مطالعات مرتبط با آن روش در سنجش زبان با دقت بررسی می‌شود تا زمینه برای ورود به مطالعه تجربی فراهم شود. در هر فصل با استفاده از روش به‌کاررفته یا به تحلیل مجدد داده‌های آماده می‌پردازد و یا داده‌هایی که صرفاً برای آن روش گردآوری شده است را تحلیل می‌کند. این تحلیل تمام بخش‌های یک مقاله کامل نظیر روش‌شناسی، نتایج، بحث و نتیجه‌گیری را دارد.

در واقع، هر فصل را از نظر ساختاری می‌توان به دو بخش نظری و عملی تقسیم کرد. این ساختار دو ویژگی خاص دارد. از یک طرف به‌نویسنده این امکان را می‌دهد تا بتواند در مورد مفاهیم نظری، متغیرها، و مطالعات انجام شده، نرم‌افزارهای مرتبط، تارنماهای مربوطه به‌صورت مفصل صحبت کند و از طرف دیگر هم با استفاده از تحلیل مجموعه داده‌ها بتواند روش مورد نظر را به‌صورت عملی به‌خواننده بشناساند، شیوه‌ای که معمولاً در موارد مشابه اتفاق نمی‌افتد. بنابراین، فصل‌های کتاب هم به‌لحاظ ساختاری و هم محتوایی کامل و جامع هستند.

### ۳.۵ انسجام و پیوستگی مطالب

تقسیم‌بندی کتاب به سه بخش نظریه‌های اندازه‌گیری، روش‌های آماری، و داده‌کاوی به‌لحاظ موضوعی باعث شده‌است تا نظم خاصی بر آن حاکم شود. درست است که در هر فصلی یک روش متفاوت تحلیل داده‌ها مطرح شده است، ولی تنوع روش‌ها در ارائه از انسجام منطقی برخوردار است. روش‌ها از ساده به پیشرفته سازماندهی شده‌اند، این روش‌ها از پشتوانه نظری برخوردار می‌شوند، و با استفاده از مثال‌های متعدد و مجموعه داده‌ها برای خواننده کاربردی می‌شود. هر فصل ضمن مستقل بودن انسجام خود را به‌لحاظ ساختاری با فصل‌های قبل و بعد از خود با ارجاعات متعدد حفظ می‌کند.

روش‌ها در فصل‌ها طوری انتخاب شده‌اند که به‌لحاظ دشواری پیش‌نیاز روش‌های بعدی در فصل‌های بعدی هستند. به‌عنوان نمونه، در بخش دوم از جلد اول، چهار فصل گنجانده شده در آن به ترتیب عبارتند از آزمون تی و تحلیل واریانس؛ تحلیل کوواریانس و تحلیل چندمتغیری کوواریانس؛ رگرسیون خطی؛ و تحلیل عاملی اکتشافی. خواننده تا آزمون تی و تحلیل واریانس را متوجه نشود نمی‌تواند تحلیل عاملی را درک کند به‌خاطر این‌که در تحلیل عاملی هم بحث تأثیر مطرح است و هم بحث همبستگی. لذا لازمه درک بهتر آن مطالعه سه روش قبلی است. این انتخاب آگاهانه در انتخاب روش‌ها و مرتب کردن آن‌ها باعث انسجام، پیوستگی، و هدفمندی مطالب شده‌است.

#### ۴.۵ تنوع روش‌های کمی و نرم‌افزارهای تحلیل داده‌ها

در مجموع، ۲۷ انگاره، نظریه‌ی اندازه‌گیری، و روش آماری متفاوت در این کتاب مطرح می‌شود. این تنوع تقریباً تمام روش‌های موجود در سنجش زبان را پوشش می‌دهد و این امکان را برای خوانندگان فراهم می‌کند تا درک بهتر و عمیق‌تری نسبت به روش‌های مرسوم در سنجش زبان داشته باشند و بتوانند تصمیم‌های سنجیده‌تر و آگاهانه‌تری برای انتخاب آن‌ها جهت انجام طرح‌های پژوهشی داشته باشند. تنوع در روش باعث شده‌است تا نویسندگان از نرم‌افزارهای متناسب با این روش‌ها در تحلیل داده‌ها استفاده کنند. از نرم‌افزارهای محبوب و پرکاربردی چون SPSS تا نرم‌افزارهای پیچیده‌تر و جدیدتری مانند Facets، AMOS، Eureka، و HLM6 و ده‌ها نرم‌افزار دیگر برای تحلیل داده‌های ساده از جمله تفاوت میانگین نمرات دو یا چند گروه و داده‌هایی مانند خوشه‌بندی و روابط غیرخطی در ۲۱ فصل این کتاب استفاده شده است. گاهی مشاهده می‌شود در بعضی از فصل‌ها از دو یا چند نرم‌افزار به‌لحاظ پیچیدگی روش و داده استفاده شده است. تناسب نرم‌افزارها با روش‌های مطرح شده در کتاب جهت تحلیل داده‌ها به‌خواننده این امکان را می‌دهد تا بتواند در مواجهه با داده‌های مشابه نرم‌افزار مناسب را انتخاب کند.

#### ۵.۵ تخصص نویسندگان

انتخاب نویسندگان برای نوشتن فصل‌های کتاب بسیار آگاهانه صورت گرفته‌است. از نویسندگانی استفاده شده‌است که تخصص و تجربه بالایی در کارهای تحقیقاتی دارند و در

زمینه‌ای که کار می‌کنند صاحب تألیفات زیادی هستند. تخصص‌گرایی در نوشتن فصل‌ها باعث شده است تا طیف وسیعی از نویسندگان از سرتاسر دنیا از ایران گرفته تا آمریکا همکاری کنند و لذا پراکنندگی جغرافیایی نیز لحاظ شده است. تخصص نویسندگان باعث شده است تا نویسندگان بتوانند مفاهیم پیچیده و آماری را به زبان ساده‌ای بیان کنند تا خوانندگانی که دانش کافی در مورد این روش‌ها ندارند به راحتی آنها را متوجه شوند. نکته‌ای که حائز اهمیت است و باید به آن اشاره کرد این است که نویسندگان افراد صاحب‌نام و پیش‌کسوتی چون بکمن (Bachman)، آلدرسن (Alderson)، مکنامرا (McNamara)، و فولچر (Fulcher) نیستند، بلکه افرادی هستند که بیشتر در حوزه‌های روانسنجی سنجش زبان کار کرده‌اند و تسلط، تجربه، و تخصص بالایی در زمینه تحلیل داده‌های کمی دارند.

## ۶.۵ منابع معتبر در هر فصل

یکی از ویژگی‌های بارز این کتاب این است که از منابع معتبری برای نوشتن فصل‌ها استفاده شده است. این منابع شامل مقالات چاپ شده در مجلات، منابع مهم اینترنتی، کتاب‌ها و دایرةالمعارف‌های بین‌المللی، و مقالات مهم کنفرانسی هستند. حتی المقدور سعی شده است هم از منابع کلاسیک و هم از منابع جدید استفاده شود تا خواننده بتواند روش‌های موجود را در بستر تاریخی خود بررسی کند. در مواردی هم به علت فقدان منابع موجود در سنجش زبان، نویسندگان از رشته‌های علمی مشابه دیگر استفاده کرده‌اند تا بتوانند مفاهیم را برای خوانندگان توضیح بدهند که باعث بین‌رشته‌ای بودن سنجش زبان نیز می‌شود.

## ۷.۵ زبان علمی، ساده، و روان

یکی از ویژگی‌های متون علمی نظیر مقالات که آنها را از سایر متون مانند روزنامه‌ها متمایز می‌کند زبان علمی است. زبان علمی زبانی است صریح، مستقیم، و شفاف که از واژه‌های مبهم و متکلف، ساختارهای پیچیده‌ی غیرضروری، و ترکیبات نامأنوس اجتناب می‌کند (Hyland, 2019). چنین زبانی در تمام فصل‌های این کتاب مشاهده می‌شود. مطالب، مفاهیم نظری و آماری، و اصطلاحات با زبانی ساده و روان برای خواننده توضیح

داده می‌شود تا خواننده حداقل درگیر زبان مبهم نشود. از کژتابی‌ها، حشو، و خلاقیت زبانی اجتناب می‌شود تا متنی روان در اختیار خواننده قرار بگیرد.

## ۶. کاستی‌های کتاب

همان‌طوری که پیش‌تر نیز به‌آن اشاره شد، این اولین کتاب است که در زمینه تحلیل داده‌های کمی در سنجش زبان نوشته شده است. علیرغم مزایایی که در قسمت قبل در مورد آن‌ها توضیح داده شد، کتاب خالی از ایراد نیست. در این قسمت به‌این معایب پرداخته می‌شود.

### ۱.۶ عدم شفافیت در انتخاب روش‌های تحلیل داده‌های کمی

اولین ایراد مهم این کتاب این است که مشخص نیست معیار انتخاب این روش‌ها چه بوده است و متأسفانه گردآورندگان آن نیز توضیحی در مورد معیارهای انتخاب نمی‌دهند. تنها توضیحی که گردآورندگان در مقدمه‌ی پایانی دو جلد به‌آن اشاره می‌کنند این است که باید نقش نظریه و ماهیت سؤال تحقیق را در استفاده از این روش‌ها مد نظر قرار داد، اما هیچ اشاره‌ای به معیارهای انتخاب این روش‌ها نشده است. آیا معیار انتخاب بسامد بالای آن‌ها در مجلات، کتب، و رساله‌ها و پایانامه‌ها بوده است؟ آیا معیار انتخاب اهمیت و کاربرد آن‌ها در سنجش زبان است؟ و یا معیار انتخاب دیدگاه‌های پیش‌کسوتان این حوزه بوده است؟ یا اینکه سلیقه خود گردآورندگان بوده است؟ دانستن معیارهای انتخاب می‌تواند برای خواننده مهم باشد بخاطر اینکه اگر صرفاً انتخاب آن‌ها سلیقه‌ای بوده باشد، واقعاً چه لزومی دارد که خواننده وقت خود را صرف مطالعه روشی مثل رگرسیون نمادین مبتنی بر الگوریتم تکاملی کند که ممکن است اصلاً از آن استفاده‌ای نشود؟ بهتر بود گردآورندگان این موضوع را شفاف‌کرده بودند تا تکلیف خواننده مشخص می‌شد.

### ۲.۶ عدم اختصاص فصلی به تحلیل داده‌های کیفی و ترکیبی

دومین عیب اساسی کتاب این است که اصلاً مطلبی در مورد تحلیل داده‌های کیفی و ترکیبی وجود ندارد. درست است که کتاب فقط به‌تحلیل داده‌های کمی اختصاص

یافته است و بنابراین بحث در مورد تحلیل داده‌های دیگر محلی از اعراب پیدا نمی‌کند، اما داده‌های کیفی و ترکیبی نیز بخشی از داده‌های این زیرشاخه از زبان‌شناسی کاربردی را تشکیل می‌دهند و گردآوردگان می‌توانستند عنوان کتاب را طوری انتخاب کنند که بتوانند روش‌های تحلیل داده‌های کمی، کیفی، و ترکیبی را پوشش دهند تا هم تصویر کامل‌تری از انواع تحلیل ارائه دهند و هم به‌عنوان اولین کتاب در این زمینه مبنای تألیف و گردآوری کتب تخصصی‌تر باشند.

### ۳.۶ نگنجاندن منابع مرتبط برای مطالعه پیش‌تر

در کتبی که جنبه‌ی گردآوری مقالات تألیفی دارند رسم بر این است تا در انتهای هر فصل قبل از فهرست منابع، تحت عنوان مطالعه اضافی (Further Reading)، چند منبع کاملاً مرتبط با موضوع معرفی شود تا در صورت نیاز و مطالعه بیشتر خواننده به آنها مراجعه کند تا دانش عمیق‌تری نسبت به موضوع کسب کند. این ویژگی در این کتاب رعایت نشده است و نقص بزرگی برای آن تلقی می‌شود بخاطر اینکه خواننده ممکن است نتواند منابع مرتبط را در فهرست منابع شناسایی کند و در نتیجه ممکن است از کسب اطلاعات بیشتر محروم بماند.

### ۴.۶ جای خالی برخی از روش‌های سستی‌تر

شاید عیب دیگر این کتاب این باشد که از معرفی روش‌های سستی‌تر غافل شده است. در سنجش زبان، ما غالباً با موقعیت‌هایی مواجهه می‌شویم که نیاز است تا رابطه‌ی (نه همبستگی) بین دو یا چند متغیر را بررسی کنیم و یا تأثیر یک متغیر مستقل را بر روی چند متغیر وابسته اندازه بگیریم. در چنین شرایطی از روش‌هایی مثل  $\chi^2$  (Chi Square) و تحلیل واریانس چندمتغیره (Multivariate Analysis of Variance) استفاده می‌شود که از اضافه کردن آنها صرف‌نظر شده است. مشخص نیست چرا گردآوردگان کتاب روش ساده‌ای مانند آزمون تی را مطرح کرده‌اند، ولی از طرح چنین روش‌های پرکاربردی غافل شده‌اند.



## ۷. نتیجه‌گیری

در مقاله حاضر، نگارنده به معرفی، نقد و بررسی اولین کتاب تحلیل داده‌های کمی در سنجش زبان پرداخت. چاپ چنین کتابی در سنجش زبان از جهات مختلف حائز اهمیت است. اول اینکه، بعد از ۶۰ سال که سنجش زبان فعالیت رسمی خود را شروع کرده است، این اولین کتابی است که روش‌های تحلیل داده‌های کمی را به صورت جامع معرفی می‌کند. می‌توان این کتاب را "کتاب طلایی" تحقیق تحلیل کمی در سنجش زبان بنامیم. دوم، تنوع روش‌های معرفی شده سنتی و جدید، تحلیل متنوع با استفاده از نرم‌افزارهای رایج و پرکاربرد، از یک طرف، و نرم‌افزارهای جدیدتر و با قابلیت‌های پیش‌تر، از طرف دیگر، و بحث و نتیجه‌گیری شفاف با استفاده از داده‌های واقعی می‌تواند از طرفی منبع مناسبی برای دروس روش تحقیق کمی و آزمون‌سازی در مقاطع ارشد و دکتری در رشته زبان‌شناسی کاربردی باشد و از طرفی دیگر مرجع ارزشمندی برای اساتید، زبان‌شناسان کاربردی، دانشجویان مقاطع تحصیلات تکمیلی در سنجش زبان، و پژوهش‌گران قرار بگیرد. سوم، نحوه‌ی ارائه‌ی مطالب، نوع نگاه متفاوت نویسندگان به روش‌های استفاده شده، و معرفی روش‌های جدیدتر و پیچیده‌تر در کتاب می‌تواند آغازگر تحقیقات جدید و ارزشمندی در حوزه‌ی سنجش زبان قرار بگیرد، باعث گسترش این زیرشاخه از زبان‌شناسی کاربردی شود، و در نتیجه منجر به تولید علم شود. بنابراین، مطالعه آن را به تمامی دانشجویان تحصیلات تکمیلی زبان‌شناسی کاربردی، اساتید دانشگاه در این حوزه، و پژوهشگران جهت انجام پژوهش در سنجش زبان و روانسجی توصیه می‌کنم.

## کتاب‌نامه

- Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2019). *Introduction to research in education*. Boston: Cengage.
- Aryadoust, V., & Raquel, M. (Eds.). (2019). *Quantitative data Analysis for language assessment Volume I: Fundamental techniques*. New York, NY: Routledge.
- Aryadoust, V., & Raquel, M. (Eds.). (2020). *Quantitative data analysis for language assessment Volume II: Advanced methods*. New York, NY: Routledge.
- Aryadoust, V., Goh, C. (2020). Classification and regression trees in predicting listening item difficulty. In V. Aryadoust & M. Raquel (Eds.). *Quantitative data Analysis for language assessment Volume II: Advanced Methods* (pp. 15-30). New York, NY: Routledge.

- Aryadoust, V., Ng, L. Y., & Sayama, H. (2020). A comprehensive review of Rasch measurement in language assessment: Recommendations and guidelines for research. *Language Testing*, 1-35. 10.1177/0265532220927487
- Green, R. (2019). Item analysis in language assessment. In V. Aryadoust & M. Raquel (Eds.), *Quantitative data Analysis for language assessment Volume I: Fundamental techniques* (pp. 15-30). New York, NY: Routledge.
- Hyland, K. (2019). *Second language Writing* (2<sup>nd</sup> ed.). Cambridge: Cambridge University Press.
- Koizumi, R., & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research*, 4(5), 90-91.
- Kunnan, A. (Ed.). (2014). *The Companion to language assessment*. Singapore: Black Wiley.
- McKinley, J. (2020). Introduction: Theorizing research methods in the golden age of applied linguistics research. In J. McKinley & H. Rose (Eds.), *The Routledge handbook of research methods in applied linguistics* (pp. 1-12). London: Routledge.
- McNamara, T. (2011). Applied linguistics and measurement: A dialogue. *Language Testing*, 28(4), 435-440.
- Ockey, G. J., & Green, B. A. (Eds.). (2020). *Another generation of fundamental considerations in language assessment*. Singapore: Springer.
- Rasch, G. (1960 /1980). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research, 1960. (Expanded edition, Chicago: The University of Chicago Press, 1980).
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14, 271-282.
- Shohamy, E., & Or, L. (Eds.). (2017). Language testing and assessment. In M. Stephen (Ed.), *Encyclopedia of language and education* (Volume 7). Cham: Springer.
- Tsagari, D., & Banerjee, J. (Eds.). (2017). *Handbook of second language assessment*. Berlin: de Gruyter.
- Winne, P. H., & Baker, R. S. J. D. (2013). The potentials of educational data mining for researching metacognition, motivation, and self-regulated learning. *Journal of Educational Data Mining*, 5(1), 1-8.