

Critical Studies in Texts and Programs of Human Sciences,
Institute for Humanities and Cultural Studies (IHCS)
Monthly Journal, Vol. 21, No. 10, Winter 2022, 1-25
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

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



Copyright © 2018, 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 به نظر می‌رسد. کتاب دوجلدی *تحلیل داده‌های کمی در سنجش زبان: تکنیک‌های بنیادین و روش‌های پیشرفته* به‌منظور پرکردن این خلأ از سوی وحید آریادوست (Vahid Aryadoust) و میشل راکوئل (Michelle Raquel) گردآوری شده است و انتشارات راتلج (Routledge) آن را منتشر کرده است. 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 به عدم شفافیت در انتخاب روش‌ها اشاره کرد. کتاب حاضر می‌تواند منشأ تولید

* دانشیار آموزش زبان انگلیسی، عضو هیئت علمی دانشگاه بین‌المللی امام خمینی (ره)، قزوین، ایران،
esfandiari@hum.ikiu.ac.ir

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



Copyright © 2018, This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International, which permits others to download this work, share it with others and Adapt the material for any purpose.

پژوهش‌های سودمند در این حوزه باشد، به‌گسترش آن کمک شایانی کند، و به تولید علم در سنجش زبان منجر بشود. لذا، مطالعه آن به جامعه آزمون‌سازی و سنجش زبان مؤکداً توصیه می‌شود.

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

۱. مقدمه

داده‌ها نقش کلیدی را در علوم امروزی بازی می‌کنند، منجر به نظریه‌پردازی، باعث تأیید یا رد ادعای صاحب‌نظران می‌شوند، و به گسترش 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 بوده است، تاجایی‌که حتی امروزه داده‌های کمی مهم‌ترین نوع داده در این زیرشاخه محسوب می‌شود. علت این اهمیت نیز ماهیت علمی این زیرشاخه است، به‌سبب 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 تحت‌تأثیر دو رشته تخصصی زبان‌شناسی کاربردی و اندازه‌گیری بوده است، لذا آمار و ارقام به‌عنوان بهترین نوع داده کمی در سنجش زبان مطرح می‌شود (Ockey and Green 2020).

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

باوجود اهمیت تحلیل داده‌های کمی در سنجش زبان، شوربختانه توجه خاصی به آن‌ها نشده است. روش‌های تحلیل داده‌ها در سنجش زبان یا همان روش‌های رایج در زبان‌شناسی کاربردی‌اند یا به‌صورت 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 *راهنما نامه* سنجش زبان دوم توسط تاساگاری و بنرجی (Tsagari and Banerjee 2017) و در ویرایش سوم کتاب شش صد صفحه‌ای *آزمون‌سازی و سنجش زبان* که توسط شوهایمی و اور (Shohamy and Or 2017) منتشر شد یک فصل هم در مورد تحلیل داده‌های کمی سنجش زبان یافت نمی‌شود. در تنها دایرة‌المعارف تخصصی چهارجلدی *The Companion to Language Assessment*، که کونان (Kunnan 2014) به سنجش زبان اختصاص داده است، فقط ۱۴۰ فصل از 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 مؤسسه 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 هیئت تحریریه مجلاتی چون *ارزش‌یابی آموزشی (Educational Assessment)* است. دارای بیش از پنجاه مقاله چاپ شده در مجلات معتبر بین‌المللی نظیر این مجلات است: *فصل‌نامه سنجش زبان (Language Assessment Quarterly)*، *آزمون‌سازی (Language Testing)*، *فصل‌نامه آموزش زبان انگلیسی به غیرانگلیسی‌زبانان (TESOL Quarterly)*، و *یادگیری زبان با استفاده از رایانه (Computer Assisted Language Learning)*، و روان‌شناسی تربیتی.

این پژوهش‌گر ایرانی تبار صاحب کتب 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 انگاره‌اش در سنجش زبان، طراحی آزمون‌های شنیداری، و اعتبارسنجی سازه‌آزمون‌های زبانی نوشته است که در کتب زبان‌شناسی کاربردی در انتشارات معتبر بین‌المللی هم‌چون انتشارات دانشگاه کمبریج، انتشارات بلک وایلی، و انتشارات راتلج به چاپ رسیده است. در سال ۲۰۱۹ به پاس خدمات ارزنده علمی - تحقیقاتی در جامعه بین‌الملل به کسب جایزه بین‌المللی Intercontinental Academia Fellowship موفق شد.

میشل راکوئل (Michelle Raquel) دانشیار مرکز مطالعات انگلیسی کاربردی در دانشگاه هنگ‌کنگ است. او بیش از پانزده سال سابقه تدریس دروس زبان‌شناسی کاربردی در مقاطع مختلف تحصیلات تکمیلی را در دانشگاه‌های مختلف هنگ‌کنگ دارد. خانم راکوئل از سال ۲۰۱۸ سرپرست پروژه دروس فشرده ارتباطات است و سابقه طولانی در مدیریت طراحی آزمون‌های مؤسسات زبان دارد. مقالات مختلفی در مجلات معتبر بین‌المللی نظیر فصل‌نامه سنجش زبان به چاپ رسانده است و نویسنده چندین فصل کتاب در کتب زبان‌شناسی کاربردی بوده است که در انتشارات معتبر بین‌المللی مانند انتشارات فرهیختگان کمبریج (Cambridge Scholars Publishing) چاپ شده است.

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 را 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 دوجلدی مورد تحلیل محتوا قرار می‌گیرد تا تصویر درست‌تر و منسجم‌تری از داده‌ها ترسیم شود.

۴. معرفی اجمالی 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 یازده 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 ده فصل نسبتاً مفصل تشکیل شده است و در 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 سی 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 گفته وی معمولاً اتفاق نمی‌افتد. به اعتقاد گرین، ضرورت این کار به این دلیل است که این سؤالات در واقع مبین خصوصیات یک آزمون خوب است که عبارت‌اند 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 واسیدای 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 اندازه خصوصیات روان‌سنجی آن اهمیت دارد. لذا، با استفاده از این نظریه، به بررسی پایایی آزمون و تعمیم‌پذیری 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، به سبب این که نظریه مذکور می‌تواند هم‌زمان چند عامل اثرگذار 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. کایربی 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 فناوری نان‌یانگ سنگاپورند که تجربه‌های ارزنده‌ای در کاربردی‌کردن 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 رفع می‌کند، نظریه بی‌عیب و نقصی نیست. مشکل اساسی نظریه، که نویسندگان نیز در نتیجه‌گیری به آن اشاره می‌کنند، این است که تحلیل کلی از متغیرها را فراهم می‌کند 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: یک‌بعدی‌بودن (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 یک‌بعدی‌بودن و استقلال موضعی که، به گفته نویسندگان، «در پیشینه تحقیقات قبلی سنجش زبان معمولاً نادیده گرفته شده است» (Aryadoust and M. Raquel 2019: 99) رعایت شود. در ادامه فصل، نویسندگان با استفاده از انگاره امتیازدهی پاره‌ای راش (partial credit Rasch model) و نرم‌افزار Winsteps به بررسی یک‌بعدی‌بودن و استقلال موضعی آزمون شنیداری می‌پردازند که بخشی از آزمون بسندگی انگلیسی بود و در دانشگاه فودان چین ساخته و برگزار شد. توصیه نویسندگان این است که برای به‌حداقل رساندن عوامل غیرمرتبط بر روایی سازه (construct validity) هر آزمونی و برای کمک به خوانندگان برای فهم بهتر نتایج بهتر است که 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 انگاره راش را در تحلیل کارکرد افتراقی سؤال (differential item functioning) در تحقیقات مربوط به سنجش زبان مورد بحث قرار می‌دهد. این فصل دراصل ادامه‌دهنده بحث مربوط به ویژگی‌های انگاره راش است که در فصل قبلی معرفی شد. نویسنده کارکرد افتراقی سؤالات قسمت شنیداری آزمون دلتا (DELTA) را با استفاده از انگاره راش بررسی می‌کند. نتایج بررسی مشخص می‌کند که اگرچه تعدادی از سؤالات این آزمون تشخیصی کارکرد افتراقی داشتند، نتایج آزمون تحت‌تأثیر این سؤالات معیوب قرار نگرفتند.

در دو فصل قبلی، انگاره ساده راش، ویژگی‌های این انگاره، و دو الزام مهم آن مورد بررسی قرار گرفت. انگاره ساده یکی از انگاره‌های خانواده راش است. در این فصل، ایکیو چوی (Ikkyu Choi) کاربرد دو انگاره دیگر از خانواده راش را به نام‌های انگاره مقیاس نمره‌دهی راش (Rasch rating scale model) و انگاره امتیازدهی پاره‌ای راش (partial credit Rasch model) را در مطالعات سنجش زبان معرفی می‌کند و آن‌ها را 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 انگاره امتیازدهی پاره‌ای راش برای بررسی ساختار پرسش‌نامه مناسب‌تر است. درخاتمه، نویسنده به یک جمع‌بندی خیلی جالب می‌رسد و جمله زیبایی دارد؛ این‌که با توجه به قدمت این دو انگاره، «می‌توان با اطمینان گفت که آن‌ها بخش جدایی‌ناپذیر روش‌های تحلیل داده‌های چندحالتی در مطالعات سنجش زبان شده‌اند» (ibid.: 149).

فصل آخر در این بخش به انگاره چندوجهی راش و کاربردهای آن در سنجش زبان تعلق دارد. انگاره‌ای که از آن به‌عنوان پرکاربردترین انگاره خانواده راش (Aryadoust et al. 2020) نام برده می‌شود و مک‌نامرا (McNamara 2011: 436) پیدایش آن را به «جهش کوانتومی» (quantum leap) تعبیر می‌کند. تامس اکیز (Thomas Eckes)، با تجربه چندین ساله عملی و تحقیقاتی خود در این زمینه، به معرفی این انگاره می‌پردازد. نویسنده، بعد از معرفی و توضیح آن و مرور مطالعات مرتبط با سنجش زبان، تلاش می‌کند تا با استفاده از داده‌های آماده یک آزمون نوشتاری کاربرد این انگاره را نشان بدهد. برای این کار از نرم‌افزار FACETS استفاده کرد که به همین منظور طراحی شده است. 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 خانواده‌ی راش امکان‌پذیر نیست. در صفحات پایانی فصل، اکیز به پیشرفت‌هایی اشاره می‌کند که در این زمینه صورت گرفته است و کاربردهای عملی‌ای که این انگاره دارد. او همچنین به موارد بحث‌برانگیز استفاده از این انگاره گریزی می‌زند و این فصل را با معرفی نرم‌افزارهای دیگری هم‌چون ConQuest برای 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)، نویسنده 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 و این‌که اگر پیش‌فرض‌های آزمون تی و تحلیل واریانس نقض شوند و مطالعاتی را که از این دو آزمون در سنجش زبان استفاده کرده‌اند مرور می‌کند. توضیحات مبسوط نویسنده باعث می‌شود شبیهاتی که خواننده در مورد این دو آزمون دارد برطرف شود. نویسنده در قسمت دوم مقاله با استفاده از این دو آزمون در دو مجموعه داده متفاوت مرحله‌به‌مرحله خواننده را با کاربرد آن‌ها آشنا می‌کند. با استفاده از یک آزمون تلفظ، تأثیر آهنگ‌های انگلیسی را در 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 به خواننده توضیح بدهد. اگرچه اکثر خوانندگان ممکن است با این دو آزمونی که در این فصل مطرح شد آشنایی نسبی داشته باشند، نحوه ارائه آن‌ها، به‌کارگیری 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 نهم‌اند و تلاش می‌کنند کاربرد تحلیل کوواریانس (analysis of covariance) و تحلیل کوواریانس چندمتغیره (multivariate analysis of covariance) را در سنجش زبان توضیح دهند. 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 (برنامه 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. براساس پیشینه 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 به‌جای استفاده از تحلیل واریانس از این دو آزمون استفاده شود، زیرا در این نوع طرح‌ها معمولاً متغیرهای کمکی وجود دارند و روی متغیر وابسته اثر می‌گذارند و نتایج را ممکن است تغییر بدهند. علی‌رغم اهمیت و مفید بودن این دو آزمون، به‌اعتقاد 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 تاهربهای (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 پیش‌بینی‌کننده (predictor) یک یا چند متغیر در متغیر وابسته مورد بررسی قرار بگیرد. به همین 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)، و توزیع نرمال داده‌ها (normality) توضیح داده می‌شود. نمونه داده برای بررسی رگرسیون خطی که در این فصل استفاده شد، بخشی از داده‌های آزمون بسندگی (proficiency) بود که بر روی پانصد زبان‌آموز اجرا شده بود. بعد از تحلیل داده‌ها، ارائه، و بحث درباره نتایج، نویسندگان به نکات 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 این است که اگر صرفاً از این آزمون‌ها برای پیش‌بینی استفاده شود، واقعاً نیازی نیست پژوهش‌گران نگران نقض برخی از پیش‌فرض‌های رگرسیون خطی باشند، زیرا این آزمون آماری به اندازه کافی استوار (robust) است که نقض برخی از پیش‌فرض‌ها نتواند ضرایب رگرسیون (regression 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 به عنوان روش آماری چندمتغیری و یکی از روش‌های داده‌گاهی (data reduction) و دو نوع مهم آن - تحلیل عاملی اکتشافی و تحلیل عاملی تأییدی (confirmatory) - ارائه می‌شود. سپس، تحلیل عاملی اکتشافی و پنج مرحله 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 یا آزمون استفاده می‌شود؛ در واقع، روایی سازه هر ابزاری که برای سنجش آزمودنی‌ها (testees) مورداستفاده قرار بگیرد. به همین منظور، مؤلفان با استفاده از پرسش‌نامه راه‌بردهای خواندن (reading test strategy use questionnaire) به شناسایی 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 به استفاده درست تحلیل عاملی اکتشافی اختصاص دارد که به پژوهش‌گران توصیه می‌شود که 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

آغازکننده بخش اول از جلد دوم فصلی درباره استفاده از انگاره ترکیبی راش (mixed Rasch model) برای سنجش خواندن و درک مطلب است. 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 به پایان 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 انگاره 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 انگاره یک‌بعدی راش می‌دانند، به سبب 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 مناسب‌ترند 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 چه‌طور از آن می‌شود برای محاسبه دقیق‌تر پایایی آزمون شنیداری و سایر آزمون‌ها استفاده کرد. این فصل با معرفی کوتاه سایر انگاره‌های چندبعدی راش و استفاده آن‌ها در سنجش زبان به پایان می‌رسد.

در فصل سوم، یکی دیگر از انگاره‌های چندبعدی مطرح می‌شود. انگاره شناختی تشخیصی لگاریتم خطی (log-linear cognitive diagnosis modeling) به نوعی از انگاره چندبعدی غیرراش اطلاق می‌شود که ویژگی‌های دو علم روان‌سنجی و روان‌شناسی را با

هم 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 درباره ویژگی‌های ریاضی و مبانی نظری انگاره، نویسندگان مطالعاتی را که با استفاده از آن در سنجش زبان انجام شده است مرور می‌کنند. در ادامه، برای فهم بیشتر، با استفاده از یک آزمون شنیداری کاربردها و خصوصیات آن را نشان می‌دهند و براساس تحلیل داده‌ها، ارائه، و بحث درباره 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 انگاره سلسله‌مراتبی تشخیصی - دسته‌بندی (hierarchical diagnostic classification model) را در سنجش خواندن و درک مطلب توضیح می‌دهد. مانند انگاره شناختی تشخیصی لگاریتم خطی در فصل قبلی، انگاره 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، به سبب این‌که می‌تواند روابط بین متغیرهایی که عملکرد آزمودنی را توضیح می‌دهد شناسایی کند و هم این‌که وابستگی متغیرها به هم‌دیگر را کشف کند. مقاله با ارائه راه‌کارهای کاربردی برای انتخاب مدل برای 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. به عنوان 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 را که آن‌ها در آزمون IELTS می‌گیرند پیش‌بینی کند. علاوه بر کاربرد پیش‌بینی‌کنندگی، نویسندگان معتقدند که مدل‌سازی معادلات ساختاری را می‌توان برای اعتبارسنجی روایی سازه 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) یا SGP نام دارد و برای محاسبه 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 آزمون متفاوت یکسان باشد، نمرات واقعی SGP آن‌ها با هم فرق دارد (نویسندگان در ص 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 به هم‌ارزسازی آزمون (test equating) نیاز 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 انجام شده بود استفاده کردند و نشان دادند که چگونه می‌توان گرایش آزمودنی‌ها برای کسب نمره GSP را که لازمه گرفتن نمره برش است اندازه گرفت. نویسندگان بر این باورند که این روش روش خوبی برای بررسی تغییراتی است که در چهار مهارت شنیداری، گفتاری، خواندن، و نوشتن در طی بازه طولانی مدت 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 ورود پیدا کنند. سپس، این روش را در سنجش زبان دوم تشریح می‌کنند و گله‌مندند که 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 ملاحظات در استفاده از این روش نظیر حجم داده‌ها، داده‌های گم‌شده (missing data)، و نوع نرم‌افزار اشاره می‌کنند. درخاتمه، رهنمودهایی برای استفاده از این روش برای مطالعات آتی در حوزه سنجش زبان دوم ارائه می‌دهند.

فصل پایانی به‌نوعی ادامه‌دهنده فصل قبلی در این بخش است، به سبب 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 بلندمدت گردآوری می‌شود، با این روش بررسی می‌کنند. به همین سبب، 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 نرم‌افزار HLM6 داده‌ها تحلیل می‌شوند، نتایج مطالعه با جزئیات ارائه می‌شوند، و در انتهای فصل به محدودیت‌های مطالعه اشاره می‌کنند. این فصل درحالی‌به‌پایان 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 درحالی‌به‌پایان می‌رسد که در این بخش پایانی دو روش پیشرفته داده‌کاوی معرفی می‌شوند که در سنجش زبان بسیار تازگی دارند و به‌ندرت می‌توان در مورد آن‌ها در مجلات سنجش زبان 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 به‌کار می‌رود (aryadoust and Goh 2020). کاربردهای مهم داده‌کاوی عبارت‌اند از خوشه‌بندی (clustering)، طبقه‌بندی (classification)، پیش‌بینی، و خلاصه‌سازی (Winne and Baker 2013). در این بخش، دو روش داده‌کاوی به خواننده 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 به جامعه سنجش زبان معرفی می‌کنند. به همین منظور، ابتدا مقدمه کوتاهی بیان می‌کنند و

مفاهیمی چون داده‌کاوی، طبقه‌بندی، و رگرسیون را توضیح می‌دهند. بعد از این مقدمه کوتاه، با استفاده از یک مثال فرضی، به تبیین این روش می‌پردازند و درباره مفاهیم بنیادی آن از جمله اعتبارسنجی متقابل (cross-validation)، برازش، و شاخص اهمیت متغیر (variable importance index) بحث می‌کنند. در ادامه، تنها دو مطالعه‌ای را که در این زمینه در سنجش زبان انجام شده است خلاصه می‌کنند. سپس، با استفاده از این روش، سطح دشواری ۳۲۱ سؤال آزمون شنیداری را بررسی می‌کنند. با استفاده از تحلیل داده‌ها به مجموعه‌ای از حکم‌های اگر-آن‌گاه (if-then rules) دست می‌یابند که سطح دشواری را به ویژگی‌های زبانی مرتبط می‌سازد. به‌علت جدید بودن آن در سنجش زبان، نویسندگان پیش‌نهاد می‌کنند تا از آن در محیط‌های دیگر استفاده شود تا بتوان به نتایج تعمیم‌پذیر 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) برای تعیین رابطه بین خواندن و دانش دستوری-واژگانی. آریادوست، نویسنده این فصل، مبحثی را مطرح می‌کند که فقط خود او توانسته است در حوزه سنجش زبان در مجلات غیرزبان‌شناختی مقاله چاپ کند. از این رو، این فصل می‌تواند شروع خوبی برای تحقیقاتی از این دست در سنجش زبان باشد. به‌دلیل جدید بودن آن، آریادوست تمام تلاش خود را به‌کار می‌گیرد تا مفاهیم و اصطلاحات فنی این روش را با زبان ساده‌ای برای خواننده توضیح دهد و در این مسیر رنج بسیار زیادی هم متحمل می‌شود. نویسنده با استفاده از تحلیل داده‌های یک آزمون بسندگی با این روش و مقایسه آن با یک روش رگرسیون خطی نشان می‌دهد که این روش مناسب‌تری برای پیش‌بینی توانایی خواندن آزمودنی‌هاست و نتایج دقیق‌تری را ارائه می‌کند. درخاتمه، نویسنده هشدار می‌دهد که نباید مجذوب معادلات پیچیده 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 به‌دست آوریم.

۵. امتیازات کتاب

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

۱.۵ جامع‌بودن 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 با سایر زیرشاخه‌های زبان‌شناسی کاربردی است. علاوه بر این، متغیرهایی که در سال‌های اخیر در این زیرشاخه مورد تحلیل قرار گرفته‌اند به مراتب پیچیده‌تر از قبل‌اند. 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 حوزه 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 را نظیر تحلیل واریانس و هم روش‌های پیشرفته آماری را مانند مدل‌سازی معادلات ساختاری با وسواس خاصی برای خوانندگان توضیح می‌دهند. شاید مهم‌ترین بخش کتاب دو فصل آخر جلد دوم باشد که نویسندگان تمام همت خود را به‌کار بسته‌اند تا پیچیده‌ترین و پیشرفته‌ترین روش‌های تحلیل داده‌ها را نظیر روش‌های بهینه‌سازی 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 به‌لحاظ ساختاری و محتوایی

گردآورندگان این کتاب دوجلدی هر جلد را به سه بخش کلی تقسیم‌بندی کرده‌اند که عبارت‌اند از نظریه‌های اندازه‌گیری، روش‌های آماری، و داده‌کاوی. این سه بخش درعین کلی‌بودن کامل نیز هستند، به‌سبب این‌که گردآورندگان را قادر ساخته است تا بتوانند از طرفی انواع روش‌های تحلیل داده را زیر این سه بخش مدیریت کنند و از طرفی دیگر این امکان را فراهم ساخته است تا به‌نوعی پوشش‌دهنده روش‌های تحلیل داده در سنجش زبان باشند. هر فصل تصویر کاملی از روش تحلیل داده ارائه می‌کند. هر فصل با مقدمه 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 یا داده‌هایی را که صرفاً برای آن روش گردآوری شده است تحلیل می‌کند. این تحلیل تمام بخش‌های یک مقاله کامل را نظیر روش‌شناسی، نتایج، بحث، و نتیجه‌گیری دارد.

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

۳.۵ انسجام و پیوستگی مطالب

تقسیم‌بندی کتاب به سه بخش نظریه‌های اندازه‌گیری، روش‌های آماری، و داده‌کاوی به‌لحاظ موضوعی باعث شده است 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 می‌شوند. هر فصل، ضمن مستقل بودن، انسجام خود را به لحاظ ساختاری با فصل‌های قبل و بعد از خود با ارجاعات متعدد حفظ می‌کند.

روش‌ها در فصل‌ها طوری انتخاب شده‌اند که به لحاظ دشواری پیش‌نیاز روش‌های بعدی در فصل‌های بعدی‌اند. به عنوان نمونه، در بخش دوم از جلد اول، چهار فصل گنجانده شده در آن به ترتیب عبارت‌اند 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، به سبب این که 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 مرتب‌کردن آن‌ها باعث انسجام، پیوستگی، و هدف‌مندی مطالب شده است.

۴.۵ تنوع روش‌های کمی و نرم‌افزارهای تحلیل داده‌ها

در مجموع، ۲۷ انگاره، نظریه 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 از روش‌های مرسوم در سنجش زبان داشته باشند و بتوانند تصمیم‌های سنجیده‌تر و آگاهانه‌تری برای انتخاب آن‌ها برای انجام طرح‌های پژوهشی داشته باشند. تنوع در روش باعث شده است 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 AMOS، Eureka، Facets، 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. گاه مشاهده می‌شود در بعضی از فصل‌ها از دو یا چند نرم‌افزار به لحاظ پیچیدگی روش و داده استفاده شده است. تناسب نرم‌افزارها با روش‌های مطرح‌شده در کتاب برای تحلیل داده‌ها به خواننده 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 گرفته است. از نویسندگانی استفاده شده است که تخصص و تجربه فراوان در کارهای تحقیقاتی دارند و در زمینه‌ای که کار می‌کنند صاحب تألیفات زیادی‌اند. 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 به‌زبان ساده‌ای بیان کنند تا خوانندگانی که دانش کافی در مورد این روش‌ها ندارند به‌راحتی 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 افرادی‌اند 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 چاپ‌شده در مجلات، منابع مهم اینترنتی، کتاب‌ها و دایرة‌المعارف‌های بین‌المللی، و مقالات مهم کنفرانسی‌اند. تا حد امکان 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، به‌علت 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

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 پیچیده 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

همان‌طور که پیش‌تر نیز به آن اشاره شد، این اولین کتابی است که در زمینه تحلیل داده‌های کمی در سنجش زبان نوشته شده است. علی‌رغم مزایایی که در قسمت قبل درمورد آن‌ها توضیح داده شد، کتاب خالی از ایراد نیست. در این قسمت به این معایب پرداخته می‌شود.

۱.۶ شفافیت‌نداشتن در انتخاب روش‌های تحلیل داده‌های کمی

اولین ایرد مهم این کتاب این است که مشخص نیست معیار انتخاب این روش‌ها چه بوده است و متأسفانه گردآورندگان آن نیز توضیحی درمورد معیارهای انتخاب نمی‌دهند. تنها توضیحی که گردآورندگان در مقدمه پایانی دو جلد به آن اشاره می‌کنند این است که باید نقش نظریه و ماهیت سؤالات تحقیق را در استفاده از این روش‌ها مدنظر 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؟ یا معیار انتخاب دیدگاه‌های پیش‌کسوتان این حوزه بوده است؟ یا این که سلیقه خود گردآورندگان بوده است؟ دانستن معیارهای انتخاب می‌تواند برای خواننده مهم باشد، به سبب این که 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 شفاف می‌کردند تا تکلیف خواننده مشخص می‌شد.

۲.۶ اختصاص ندادن فصلی به تحلیل داده‌های کیفی و ترکیبی

دومین عیب اساسی کتاب این است که اصلاً مطلبی درمورد تحلیل داده‌های کیفی و ترکیبی وجود ندارد. درست است که کتاب فقط به تحلیل 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 گردآورندگان می‌توانستند عنوان کتاب را طوری انتخاب کنند که بتواند 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 ننگ‌جاندن منابع مرتبط برای مطالعه بیشتر

در کتبی که جنبه 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 از موضوع کسب کند. این ویژگی در این کتاب رعایت نشده است و نقص بزرگی برای آن تلقی می‌شود، به سبب این که 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 سنتی‌تر

شاید عیب دیگر این کتاب این باشد که از معرفی روش‌های سنتی‌تر غافل شده است. در سنجش زبان، ما غالباً با موقعیت‌هایی مواجه می‌شویم که نیاز است تا رابطه (نه هم‌بستگی) بین دو یا چند متغیر را بررسی کنیم یا تأثیر یک متغیر مستقل را در چند متغیر وابسته اندازه بگیریم. در چنین شرایطی، از روش‌هایی مثل χ^2 دو (Chi Square) و تحلیل واریانس چندمتغیره (multivariate analysis of variance) استفاده می‌شود که از اضافه کردن آن‌ها صرف‌نظر شده است. مشخص نیست چرا گردآورندگان 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 شصت 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 با قابلیت‌های بیش‌تر، 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، نحوه ارائه مطالب، نوع نگاه متفاوت نویسندگان به روش‌های استفاده‌شده، و معرفی روش‌های جدیدتر و پیچیده‌تر در کتاب می‌تواند آغازگر تحقیقات جدید و ارزش‌مندی در حوزه سنجش زبان قرار بگیرد، باعث گسترش این زیرشاخه از زبان‌شناسی کاربردی شود، و در نتیجه به تولید علم منجر شود. بنابراین، مطالعه آن را به تمامی دانشجویان تحصیلات تکمیلی زبان‌شناسی کاربردی، اساتید دانشگاه در این حوزه، و پژوهش‌گران برای انجام پژوهش در سنجش زبان و روان‌سنجی توصیه می‌کنم.

کتاب‌نامه

- Ary, D. et al. (2019), *Introduction to Research in Education*, Boston: Cengage.
- Aryadoust, V. and M. Raquel (eds.) (2019), *Quantitative Data Analysis for Language Assessment Volume I: Fundamental Techniques*, New York: Routledge.
- Aryadoust, V. and M. Raquel (eds.) (2020), *Quantitative Data Analysis for Language Assessment Volume II: Advanced Methods*, New York: Routledge.
- Aryadoust, V. and C. Goh (2020), "Classification and Regression Trees in Predicting Listening Item Difficulty", in: *Quantitative Data Analysis for Language Assessment Volume II: Advanced Methods*, V. Aryadoust and M. Raquel (eds.), New York: Routledge.
- Aryadoust, V., L. Y. Ng, and H. Sayama (2020), "A Comprehensive Review of Rasch Measurement in Language Assessment: Recommendations and Guidelines for Research", *Language Testing*: Available at: <<https://doi.org/10.1177%2F0265532220927487>>.
- Green, R. (2019), "Item Analysis in Language Assessment", in: *Quantitative Data Analysis for Language Assessment Volume I: Fundamental Techniques*, V. Aryadoust and M. Raquel (eds.), New York: Routledge.

- Hyland, K. (2019), *Second Language Writing*, Cambridge: Cambridge University Press.
- Koizumi, R. and Y. In'nami (2013), "Vocabulary Knowledge and Speaking Proficiency among Second Language Learners from Novice to Intermediate Levels", *Journal of Language Teaching and Research*, vol. 4, no. 5.
- 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: *The Routledge Handbook of Research Methods in Applied Linguistics*, J. McKinley and H. Rose (eds.), London: Routledge.
- McNamara, T. (2011), "Applied Linguistics and Measurement: A Dialogue", *Language Testing*, vol. 28, no. 4.
- Ockey, G. J. and B. A. Green (eds.) (2020), *Another Generation of Fundamental Considerations in Language Assessment*, Singapour: 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*, vol. 14.
- Shohamy, E. and L. Or (eds.) (2017), "Language Testing and Assessment", in: *Encyclopedia of Language and Education*, M. Stephen (ed.), vol. 7, Cham: Springer.
- Tsagari, D. and J. Banerjee (eds.) (2017), *Handbook of Second Language Assessment*, Berlin: de Gruyter.
- Winne, P. H. and R. S. J. D. Baker (2013), "The Potentials of Educational Data Mining for Researching Metacognition, Motivation, and Self-regulated Learning", *Journal of Educational Data Mining*, vol. 5, no. 1.

