

پژوهش‌نامه‌ی آموزش زبان فارسی به غیر فارسی‌زبانان

(علمی-پژوهشی)

سال سوم، شماره‌ی سوم (پیاپی ۸)، پاییز و زمستان ۱۳۹۳

تصحیح نوشته‌های غیر فارسی‌زبانان: سوگیری مصححان فارسی‌زبان با استفاده از انگاره‌ی چندوجهی راش

رجب اسفندیاری

استادیار گروه مترجمی زبان انگلیسی - دانشگاه بین‌المللی امام خمینی^(ه)

چکیده

در پژوهش حاضر با استفاده از «انگاره‌ی چندوجهی راش» به بررسی خطاهای مدرسان زبان فارسی در حین ارزیابی نوشته‌های فارسی‌آموزان غیرایرانی پرداخته شده است. به همین منظور، در زمستان ۱۳۹۳، دو مصحح/مدرس از دانشگاه بین‌المللی امام خمینی^(ه) به ارزیابی ۵۶ متن نوشته شده توسط ۲۸ دانشجوی پسر و دختر غیرفارسی‌زبان غیرایرانی در سطح پیشرفته پرداختند. به این دانشجویان که از ملیت‌های مختلف بوده و در مرکز آموزش زبان فارسی، در دوره‌های تکمیلی علوم انسانی و فنی-مهندسی مشغول به تحصیل بودند، دو موضوع نگارش ارائه شد و از آنها خواسته شد که یک متن درباره‌ی آن دو موضوع بنویسند. سپس مصححان به ارزیابی نوشته‌ها پرداخته و با یک مقیاس رتبه‌ای (از یک تا پنج) به نوشته‌های دانشجویان نمره دادند. داده‌های گردآوری شده در این پژوهش با استفاده از نرم‌افزار «فاستس» تجزیه و تحلیل شد. نتایج تحقیق نشان داد که مصححان از پنج معیار ارزیابی مورد بحث، نسبت به «انسجام معنایی» سوگیری بیشتری نشان داده‌اند و علاوه بر این، نسبت به هر دو موضوع نگارش نیز سوگیری داشته‌اند. نتایج، حاکی از تفاوت‌های معنی‌دار بین سوگیری‌های این دو مصحح می‌باشد. از نتایج این تحقیق می‌توان در آموزش مصححان در زمینه‌ی ارزیابی متن‌های نوشته شده توسط زبان‌آموزان استفاده کرد تا میزان سوگیری در ارزیابی و امتیازدهی به نوشته‌های آنان را کاهش داد.

کلیدواژه‌ها: معیار ارزیابی، انگاره‌ی چندوجهی راش، خطای مصحح، سوگیری

۱. مقدمه

در ارزیابی تولیدات زبانی یادگیرندگان زبان دوم/خارجی، خطاهای مصححان/ارزیابان (که از آنها تحت عنوان «اشتباهات مصحح» یا «سوگیری‌های مصحح» نیز یاد می‌شود)، از عوامل غیرمرتبط با توانایی زبان‌آموز محسوب می‌شوند که می‌توانند در تعیین نمره‌ی کسب شده توسط او مؤثر باشند. میزان بروز این خطاها در ارزیابی پاسخ‌های داده شده به پرسش‌های تولیدی که نیاز به صحبت کردن و نوشتن دارند، بسیار زیاد می‌باشد؛ زیرا ممکن است به جای یک پاسخ درست، چند پاسخ درست وجود داشته باشد و پاسخ زبان‌آموز، همان پاسخ مد نظر مصحح/ارزیاب نباشد. ارزیابی پاسخ‌های داده شده به این گونه پرسش‌ها، بیش از آن که مبتنی بر معیارهای عینی باشد، متأثر از نظر شخصی مصحح/ارزیاب است و نمره‌ای که به زبان‌آموز داده می‌شود، ممکن است پایین‌تر یا بالاتر از نمره‌ی واقعی خود او باشد. به همین دلیل است که در برخی موارد ممکن است، دو یا سه مصحح/ارزیاب، برگه‌ی یک زبان‌آموز را تصحیح/ارزیابی کنند تا درصد خطا کاهش یابد.

به اعتقاد اسکالن و همکاران (۲۰۰۰: ۹۵۷) خطاهای مصحح، «آن دسته‌ی وسیع از خطاها هستند که منجر به تغییرات منظم در نمره‌ی فراگیران می‌شود و بیشتر با خود مصحح ارتباط پیدا می‌کنند تا با عملکرد فراگیر». طبق این تعریف، ممکن است یک فراگیر از سطح توانایی خوبی برخوردار باشد، ولی نمره‌ی پایینی کسب کند؛ زیرا مصحح ممکن است در حین ارزیابی برگه‌ی این فراگیر خسته باشد یا حوصله نداشته باشد.

مایفورد و ولف (۲۰۰۴) و انگل‌هارد (۱۹۹۴) خطاهای مصحح را به دو دسته‌ی کلی تقسیم‌بندی کرده‌اند: «خطاهای کلاسیک»^۱ و «خطاهای غیررایج»^۲. «خطاهای کلاسیک» به آن دسته از خطاهایی گفته می‌شود که قدمت بیشتری دارند و تحقیقات بیشتری در مورد آنها انجام شده است. این دسته از خطاها عبارتند از: «سخت‌گیری/آسان‌گیری»^۳، «گرایش به حد وسط»^۴، «محدودیت دامنه»^۵، «اثر هاله‌ای»^۶، «نامنظمی»^۷ و «سوگیری»^۸. این خطاها به صورت منظم می‌توانند بدون این که زبان‌آموز خطایی مرتکب شده باشد، نمره‌ی او را کم و زیاد کنند. به اعتقاد مک‌نامرا (۱۹۹۶) برخلاف دوره‌ی آموزشی برای مصححان و تعیین معیارهای خاص برای نمره‌دهی، این خطاها کاملاً از بین نمی‌روند.

1. classic errors

2. modern errors

3. leniency/severity

4. central tendency effect

5. restriction of range

6. halo effect

7. randomness

8. bias

«خطاهای غیررایج» نیز به آن دسته از خطاهایی گفته می‌شود که جدیدتر هستند و تحقیقات کمتری در مورد آنها انجام شده است. این خطاها عبارتند از «خطای منطقی»^۱، «خطای تناقض»^۲، «اثرات ویژگی‌های پس‌زمینه‌ای ارزیاب/ارزیابی‌شونده»^۳، «خطای تقریبی»^۴، «خطای تازگی»^۵، و «تأثیرات ترتیبی»^۶ (مایفورد و مایفورد و ولف، ۲۰۰۴). این خطاها به علت دشوار بودن تشخیص و همچنین، دشواری سنجش آنها، کمتر مورد توجه محققان قرار گرفته‌اند.

یکی از خطاهای رایج مصححان در حین ارزیابی تولیدات نوشتاری یا گفتاری زبان‌آموزان، «سوگیری» می‌باشد (وینک و همکاران، ۲۰۱۳). «سوگیری» عبارت است: از تعامل بین مصحح و هر یک از ویژگی‌های محیط ارزیابی شامل «سطح سواد فراگیر»، «موضوع سؤال»، «معیارهای ارزیابی پاسخ‌ها». با توجه به مطالب بیان شده، می‌توان گفت که هدف از تحقیق حاضر، بررسی سوگیری مصححان در ارزیابی و تصحیح نوشته‌های فارسی‌آموزان و معیارهای ارزیابی پاسخ‌ها است. بنابراین، پرسش‌های پژوهش حاضر به شرح ذیل می‌باشد:

(۱) آیا بین مصححان فارسی‌زبان و موضوع نوشته‌های فارسی‌آموزان غیرایرانی سوگیری وجود دارد؟

(۲) آیا بین مصححان فارسی‌زبان و معیارهای ارزیابی پاسخ‌ها سوگیری وجود دارد؟

متناظر با پرسش‌های بالا، فرضیه‌های زیر مطرح می‌شود:

(۱) مصححان فارسی‌زبان نسبت به موضوع فارسی‌آموزان غیرایرانی سوگیری نشان می‌دهند.

(۲) مصححان فارسی‌زبان معیارهای ارزیابی پاسخ‌ها را یکسان ارزیابی نمی‌کنند.

۲. پیشینه‌ی پژوهش

در زمینه‌ی سوگیری در زبان اول تحقیقات زیادی انجام گرفته است. به‌عنوان مثال، در مطالعه‌ای که دو و همکاران (۱۹۹۶) در آمریکا در مورد نوشته‌های دانش‌آموزان دوره‌ی متوسطه انجام دادند، موارد احتمالی سوگیری را نسبت به نوع موضوع، نژاد، و جنسیت مورد بررسی قرار دادند و دریافتند که سوگیری معنی‌دار مصححان نسبت به موارد ذکر شده، شامل ۱۱٪ از کل سوگیری‌ها می‌شد. نتایج تحقیق نشان داد که مصححان نسبت به موضوع تحقیق، با توجه به جنسیت و نوع نژاد دانش‌آموزان سوگیری نشان می‌دهند؛

1. logical error

2. contrast error

3. influences of rater/ratee's background characteristics

4. proximity error

5. recency error

6. order effects

سوگیری مصححان به دانش‌آموزان سیاه‌پوست (نسبت به دانش‌آموزان سفیدپوست) بیشتر بود و همچنین، آنها به دانش‌آموزان دختر (نسبت به دانش‌آموزان پسر) سوگیری بیشتری داشتند.

درباره‌ی سوگیری در زبان دوم نیز ویگلزورث (۱۹۹۳ و ۱۹۹۴) به بررسی سوگیری در ارتباط با مصحح و معیار ارزیابی، مصحح و موضوع، و مصحح و نوع آزمون در مهارت صحبت کردن در استرالیا پرداخت. نتایج تحقیق نشان‌دهنده‌ی تفاوت‌های معنی‌داری بین سه نوع سوگیری است و این که مصححان، الگوی منحصربه‌فردی را در حین ارزیابی از خود نشان می‌دهند؛ برخی از آنها به مقوله‌ی دستور نمره‌ی کمتری می‌دهند و برخی دیگر به همین مقوله نمره‌ی بیشتری می‌دهند؛ برخی از مصححان در مقایسه با سایر محصلان نسبت به دایره واژگانی و شیوایی کلام سخت‌گیرتر هستند. نتایج تحقیق همچنین نشان داد که سوگیری مصححان نسبت به موضوع با یکدیگر تفاوت دارد.

مک‌نامرا (۱۹۹۶) نیز در استرالیا به بررسی آزمون انگلیسی شغلی برای ارزیابی مهاجران به این کشور پرداخت. آزمون انگلیسی شغلی آزمونی است که پزشکی که قصد اقامت در این کشور برای طبابت دارند باید در آن شرکت کنند تا دانش زبانی‌شان سنجیده شود. مک‌نامرا دریافت که مصححان آموزش‌دیده نسبت به مقوله‌ی دستور سوگیری نشان می‌دهند، در صورتی که خود متوجه این موضوع نیستند. در واقع، تفاوت آشکاری بین آنچه که مصححان فکر می‌کردند دارند انجام می‌دهند و آنچه که آنها واقعاً انجام می‌دهند وجود دارد.

لاملی (۲۰۰۵) مهارت نوشتاری آزمون بسندگی انگلیسی را در استرالیا مورد بررسی قرار داد. او با استفاده از دو انگاره (بلند فکر کردن^۱ و انگاره‌ی چندوجهی راش) داده‌های کیفی و کمی جمع‌آوری کرد. در ابتدا لاملی با استفاده از انگاره‌ی چندوجهی راش مصححان نامنظم را شناسایی کرد. وی نهایتاً از چهار مصحح آموزش‌دیده خواست تا هر کدام ۱۲ نمونه نوشته در دو موضوع را تصحیح کنند. نتایج تحقیق نشان داد که بین نمره‌دهی مصححان تفاوت معنی‌داری وجود دارد. در این تحقیق نیز مصححان، مانند تحقیق مک‌نامرا، نسبت به مقوله‌ی دستور سوگیری نشان دادند.

شیفر (۲۰۰۸) در پژوهشی گسترده در زمینه‌ی سوگیری، از ۴۰ مصحح انگلیسی‌زبان دعوت کرد تا هر کدام ۴۰ متن نوشته شده توسط دانشجویان ژاپنی را تصحیح کنند. بعد از آموزش مصححان از آنها خواسته شد تا با استفاده از مقیاس رتبه‌ای شش‌گزینه‌ای با استفاده از پنج معیار ارزیابی شامل: «محتوا»، «سازمان‌دهی مطالب»، «سبک و کیفیت بیان مطالب»، «کاربرد زبان و ساختار ظاهری» و «شیوایی مطالب»^۲ آنها را تصحیح/ارزیابی کنند و بعد از یک ماه تحویل بدهند. او دریافت که اگر مصححان «محتوا» و «سازمان‌دهی مطالب» را سخت‌گیرانه تصحیح کنند، کاربرد زبانی و/یا ساختار ظاهری را راحت‌تر تصحیح

1. thinking aloud

2. fluency

می‌کنند و بالعکس. او همچنین دریافت که برخی مصححان نسبت به فراگیران سطح پیشرفته سخت‌گیرتر از فراگیران سطح مبتدی هستند.

ماتسونو (۲۰۰۹) در ژاپن تحقیقی بر روی خودارزیابی، دیگرارزیابی، و معلم‌ارزیابی انجام داد. او از ۹۱ خودارزیاب و دیگرارزیاب و چهار معلم‌ارزیاب خواست تا متن‌های نوشته شده توسط فراگیران را تصحیح کنند. هدف از این تحقیق، مقایسه‌ی بین خودارزیاب‌ها و دیگرارزیاب‌ها با معلم‌ارزیاب‌ها بود تا معلوم شود که آیا این امکان وجود دارد تا بتوان از خودارزیاب‌ها و دیگرارزیاب‌ها در نظام آموزش عالی استفاده شود یا خیر. هر چهار معلم‌ارزیاب، ۹۱ متن نوشته شده توسط فراگیران را تصحیح کردند، خودارزیاب‌ها نیز متن‌های خود را تصحیح/ارزیابی کردند و دیگرارزیاب‌ها هر کدام پنج متن را که فراگیران نوشته بودند تصحیح کردند. ماتسونو به نتایج خیلی خوبی دست یافت. او دریافت که خودارزیاب‌ها متن‌های خود را خیلی سخت‌گیرانه تصحیح می‌کنند.

اسفندیاری و مایفورد (۲۰۱۳) در ایران به پژوهش درباره‌ی مصححان و سوگیری آنان نسبت به معیارهای ارزیابی و فراگیران پرداختند. در این پژوهش از سه نوع ارزیاب استفاده شد: خودارزیاب، دیگرارزیاب و معلم‌ارزیاب. ۱۸۷ خودارزیاب و دیگرارزیاب، و شش معلم‌ارزیاب در این تحقیق شرکت کردند. این ارزیاب‌ها متن‌های حاوی پنج بند (=پاراگراف) را که فراگیران انگلیسی‌آموز در ایران نوشته بودند، مورد ارزیابی قرار دادند. برای ارزیابی آنها از مقیاس رتبه‌ای شش‌گزینه‌ای و ۱۵ معیار استفاده کردند که عبارت بودند از: «محتوا»، «بسط جمله‌ی کلیدی مقاله»، «ارتباط موضوع مقاله با محتوای مقاله»، «مقدمه»، «انسجام»، «نتیجه‌گیری»، «سیر منطقی»، «تنوع کلمات»، «انتخاب واژه‌ها»، «اجزای کلام»^۱، «تنوع ساختارهای دستوری»، «دستور»، «املای واژه‌ها»، «شکل ظاهری کلمات»، و «به‌کارگیری علائم سجاوندی». بعد از آموزش دیدن یک‌ساعته‌ی ارزیابان، از آنها خواسته شد تا مقاله‌های پنج‌بندی را تصحیح/ارزیابی کنند. نتایج تجزیه و تحلیل داده‌ها نشان داد که آنها در ارتباط با معیارهای «انتخاب کلمات»، «تنوع کلمات» و «ارتباط موضوع مقاله با محتوای مقاله» نسبت به بقیه‌ی معیارهای ارزیابی سخت‌گیرانه‌تر برخورد می‌کردند؛ در صورتی که در ارتباط با «سیر منطقی» و «املای کلمات» آسان‌گیری بیشتری داشتند. علاوه بر این، نتایج تحقیق نشان داد که معلم‌ارزیاب‌ها سخت‌گیرتر، و خودارزیاب‌ها آسان‌گیرتر هستند. دیگرارزیاب‌ها بین این دو دسته ارزیاب قرار گرفتند.

ذکر این نکته ضروری است که در زمینه‌ی سوگیری مصححان فارسی‌زبان در ارزیابی تولیدات زبانی فارسی‌آموزان غیرایرانی پژوهشی مشاهده نشد.

1. parts of speech

۳. سوگیری

تعاریف متعددی از «سوگیری» ارائه شده است. انگل هارد (۱۹۹۴: ۹۸) سوگیری را «گرایش منظمی تعریف می‌کند که مصححان در حین ارزیابی فراگیران به آن مبادرت می‌ورزند [و در نتیجه] نمراتی که آنها به نوشته‌های فراگیران می‌دهند، پایین‌تر یا بالاتر از نمره‌ی واقعی خود فراگیران می‌باشد».

ویگلزورث (۱۹۹۳: ۳۰۹) نیز سوگیری را «الگوهای منظم» رفتاری می‌داند که در تعامل مصححان با محیط ارزیابی رخ می‌دهد؛ به عبارتی دیگر، ممکن است مصححی نسبت به یک موضوعی در طول ارزیابی، نمره‌ای پایین‌تر یا بالاتر از آن نمره‌ای که حق فراگیر است، بدهد و یا معیاری را راحت‌تر یا سخت‌گیرانه‌تر تصحیح کند.

فرخی و همکاران (۲۰۱۲) در تعریف سوگیری اظهار می‌کنند که سوگیری، حاصل کارکرد افتراقی مصحح می‌باشد که گرایش دارد تا نسبت به فراگیران، سؤالات ارزیابی یا معیارهای ارزیابی، تمایلات مطبوع یا غیرمطبوع نشان دهد. در تحقیق حاضر، «سوگیری» به «گرایش مصحح نسبت به موضوعات نوشته‌های فراگیران و معیارهای ارزیابی» اطلاق می‌شود.

۴. انگاره‌ی راش^۱

در دهه‌ی ۱۹۶۰ ریاضی‌دان و آمارشناس دانمارکی به نام «جورج راش» انگاره‌ی را در تعلیم و تربیت عرضه کرد که در حال حاضر از آن به عنوان «انگاره‌ی راش» نام می‌برند. انگاره‌ی راش، انگاره‌ی احتمالی و تجویزی است. این انگاره احتمالی است؛ به علت این که، توانایی فرد و درجه‌ی سختی سؤال ممکن است در شرایط مختلف تغییر کند و همیشه یک جواب ثابت وجود نداشته باشد. همچنین، انگاره‌ی تجویزی است؛ زیرا داده‌ها باید با انگاره مطابقت داشته باشند؛ اگر داده‌ها به هر نحوی متناسب با این انگاره نباشد محقق ملزم به گردآوری داده‌های جدید است؛ یعنی این انگاره برای سؤالات چندگزینه‌ای با یک پاسخ درست، طراحی شده است.

اساس این انگاره بر این فرض است که نمره‌ای که فراگیر کسب می‌کند تابعی از دو عامل «توانایی فراگیر» و «درجه‌ی سختی پرسش» است؛ این دو عامل در تعامل با هم نمره‌ی فراگیر را رقم می‌زنند. براساس این انگاره، سه حالت وجود دارد:

حالت اول: توانایی فراگیر مساوی است با درجه‌ی سختی سؤال که در این صورت احتمال جواب صحیح پنجاه-پنجاه می‌باشد؛ به این معنی که ۵۰ درصد احتمال وجود دارد که فراگیر به سؤال جواب درست بدهد و ۵۰ درصد هم احتمال وجود دارد که فراگیر به سؤال جواب غلط بدهد.

^۱. Rasch model

حالت دوم: توانایی فراگیر بیشتر از درجه‌ی سختی سؤال است. بنابراین، احتمال جواب صحیح، یک هست؛ به عبارت دیگر احتمال این که فراگیر بتواند پاسخ صحیح به سؤال بدهد ۱۰۰ درصد می‌باشد.

حالت سوم: توانایی فرد کمتر از درجه‌ی سختی سؤال است که در این صورت فراگیر نمی‌تواند به سؤال جواب صحیح بدهد؛ یعنی احتمال این که فراگیر بتواند جواب صحیح را بدهد صفر درصد می‌باشد. این انگاره بعدها توسط رایت و استون (۱۹۷۹) تشریح شد.

انگاره‌ی راش فقط برای سؤالات چندگزینه‌ای با جواب درست/نادرست، مناسب بود. در مواردی که جواب لزوماً درست یا نادرست نباشد، انگاره‌ی راش قادر به تحلیل آن نیست.

در سال ۱۹۷۸ پژوهشگری به نام اندریچ انگاره‌ی راش را توسعه داد تا بتوان سؤالات سنجیده شده براساس مقیاس‌های از نوع لیکرت را مورد بررسی قرار داد. این انگاره هم‌اکنون تحت عنوان «انگاره‌ی ارزیابی اندریچ»^۱ شناخته می‌شود. براساس این انگاره به پاسخی که فراگیر می‌دهد، بسته به کامل بودن یا ناقص بودن پاسخ، نمره‌ای بین صفر تا پنج داده می‌شود. عملکرد فراگیر و نمره‌ای که او می‌گیرد، تابعی از توانایی فراگیر و درجه‌ی سختی سؤال بر روی مقیاس رتبه‌ای است؛ به عبارت دیگر، در این انگاره علاوه بر درجه‌ی سختی سؤال و توانایی فرد، عامل سومی نیز در میان هست که از آن به‌عنوان «آستانه»^۲ یاد می‌شود. آستانه، فاصله‌ی بین پاسخ رتبه‌بندی شده از صفر تا پنج را می‌گویند. به این معنی که چه قدر احتمال دارد تا یک فراگیر با توانایی مشخص و درجه‌ی سختی سؤال مشخص نمره‌ای بین صفر تا پنج کسب کند؟

اگرچه انگاره‌ی اندریچ، انگاره‌ی راش را توسعه داد و باعث شد تا بتوان ارزیابی‌هایی را که براساس مقیاس‌های رتبه‌ای لیکرت انجام می‌شود، سنجید، این انگاره قادر به اندازه‌گیری پاسخ‌های تا حدودی درست، نیست.

در سال ۱۹۸۲ ماسترز انگاره‌ی راش را مطرح کرد تا بتواند انگاره‌ی اندریچ را توسعه دهد؛ به طوری که این انگاره قادر بود پاسخ‌های تا حدودی درست را نیز بر روی مقیاس‌های لیکرت اندازه بگیرد؛ به عبارت دیگر، مواردی هست که فراگیر به یک سؤال تشریحی پاسخ می‌دهد، اما پاسخ او درست یا غلط نیست، بلکه فقط تا حدودی درست است. به این پاسخ‌های تا حدودی درست نیز باید نصف نمره داده شود. این انگاره هم‌اکنون به «انگاره‌ی تا حدودی درست راش»^۳ معروف هست.

انگاره‌ی «اندریچ» و انگاره‌ی «تا حدودی درست راش» دو تفاوت اساسی با هم دارند. تفاوت اول این است که در انگاره‌ی «تا حدودی درست راش» برای پاسخ‌هایی که تا حدودی درست هستند، نصف نمره در نظر گرفته می‌شود، در صورتی که در انگاره‌ی اندریچ این گونه نیست. تفاوت دوم این است که در انگاره‌ی اندریچ

1. Andrich's rating scale

2. threshold

3. partial credit model

تعداد پاسخ‌ها برای هر سؤال از صفر تا پنج ثابت است؛ در صورتی که در انگاره‌ی «تا حدودی درست‌راش» ممکن است برای سؤال اول تعداد پاسخ‌ها از صفر تا پنج باشد و برای سؤال دوم از صفر تا سه باشد.

۵. انگاره‌ی چندوجهی‌راش^۱

در انگاره‌های خانواده‌ی راش، فقط دو عامل توانایی خودِ فراگیر و درجه‌ی سختی سؤال در تعیین نمره‌ی فراگیر نقش داشتند. این انگاره‌ها برای سؤالات چندگزینه‌ای و پرسش‌نامه‌هایی مناسب بودند که مصحح نقشی در آنها نداشت؛ اما در مواردی که مصحح باید پاسخ یک سؤال تشریحی را تصحیح کند و به آن نمره بدهد قضیه متفاوت است. بنابراین، در ادامه، به توضیح انگاره‌ای می‌پردازیم که به «انگاره‌ی چندوجهی راش» معروف شده است و ما نیز در این مطالعه از آن استفاده کرده‌ایم.

انگاره‌ی چندوجهی راش از خانواده‌ی انگاره‌های راش می‌باشد که در سال ۱۹۸۹ توسط لینکر مطرح شد و در سال ۱۹۹۴ بسط داده شد. این انگاره، انگاره‌های قبلی «راش» را یک قدم به جلو برد. به این معنی که نمره‌ای را که یک فراگیر می‌گیرد، تابعی از توانایی خودِ فراگیر، درجه‌ی سختی سؤال، و راحتی‌گیری و سخت‌گیری مصحح می‌داند. این انگاره در مواردی به کار می‌رود که مصحح در آنها نقش فعالی در ارزیابی دارد. این انگاره برای سؤالات تشریحی (نوشتاری و گفتاری) که در آنها فراگیر باید مطلبی را بنویسد و مصحح/ارزیاب به آنها نمره بدهد مناسب است. بعدها این انگاره از توانایی فراگیر، درجه‌ی سختی سؤال، و سخت‌گیری و آسان‌گیری مصحح نیز فراتر رفت و هر عاملی را که می‌توانست نمره‌ی فراگیر را به نوعی کم و زیاد کند مورد بررسی قرار داد و به همین خاطر است که از آن به‌عنوان «انگاره‌ی چندوجهی راش» یاد می‌شود. در انگاره‌ی چندوجهی راش نیاز به تصحیح یک برگه، توسط چند مصحح نیست؛ تا آن‌جا که بتوان بین مصحح، سؤال، و فراگیر ارتباط منطقی برقرار کرد، تک‌مصححی هم می‌تواند خیلی مفید باشد. به این معنی که یک مصحح، برگه‌ای را دو بار تصحیح کند. در انگاره‌ی «راش» تعداد نمره‌ها مهم‌تر از تعداد فراگیرها است.

لینکر (۱۹۹۷) معتقد است که نمره ایده‌آل برای این انگاره ۱،۱۵۲ نمره است، مشروط بر این که تمام مصحح‌ها تمام متن‌های نوشته شده توسط فراگیران را تصحیح کنند و همچنین، حداقل نمره برای این انگاره ۹۶ می‌باشد. این حداقل نمره‌ی مورد نیاز برای این انگاره است که کمتر از آن، هم درصد خطای محاسبه را بالا می‌برد و هم ممکن است باعث شود این انگاره نتواند بین متغیرهای مختلف ارتباط برقرار کند. بنابراین، این انگاره هم برای الگوهای کامل^۲ و هم برای الگوهای تا حدودی کامل^۳ به کار می‌رود.

1. many-facet Rasch model

2. fully crossed design

3. partially crossed design

انگاره‌ی راش دارای پیش‌فرض‌هایی است که باید تامین بشوند تا بتوان از آن استفاده کرد. اولین و مهمترین پیش‌فرض آن این است که این انگاره تک‌بعدی است (اکیز، ۲۰۰۹)؛ یعنی این که سؤالات آزمون همگی باید یک توانایی را مورد سنجش قرار بدهند. به‌طور مثال، اگر یک آزمون ۲۰ سؤالی طراحی کنیم که همه‌ی سؤالات آن، مهارت نوشتن را بسنجند، می‌توانیم ادعا کنیم که تک‌بعدی بودن لحاظ شده است. پیش‌فرض دوم این انگاره، نامتغیر بودن آن است (لینکر، ۱۹۹۴). این بدان معنی است که توانایی فراگیر و درجه‌ی سختی سؤال دو مقوله‌ی مجزا از هم هستند و برای محاسبه‌ی یکی نیاز به دیگری نداریم. به عبارت دیگر، توانایی فرد و درجه‌ی سختی سؤال نباید تحت تأثیر یکدیگر قرار بگیرند. به‌عنوان مثال، هر فراگیر با توانایی بالا باید بتواند به یک سؤال راحت جواب بدهد و جابجایی فراگیر با همان توانایی نباید تفاوتی در پاسخ دادن به سؤال، داشته باشد و سؤالی با سختی بالا را نباید فراگیری با توانایی پایین جواب بدهد و جابجایی این سؤال با همان درجه‌ی سختی نباید تفاوتی کند.

پیش‌فرض سوم این انگاره، افزودنی بودن آن است (اسمیت، ۲۰۰۴)؛ بدان معنی که توانایی فراگیر، درجه‌ی سختی سؤال، و آسان‌گیری و سخت‌گیری مصحح به صورت خطی عمل می‌کنند تا داده‌هایی را که در اصل به صورت رتبه‌ای بوده‌اند، تبدیل به داده‌های فاصله‌ای کنند. این انگاره قادر به محاسبه‌ی عوامل مختلف مؤثر در تعیین نمره‌ی فراگیر است.

پیش‌فرض آخر این انگاره، تجویزی بودن آن است (بقایی و امراهی، ۲۰۰۹). در این انگاره فرض بر آن است که داده‌های گردآوری شده با انگاره همخوانی داشته باشند و خطاهای موجود در داده‌ها پایین باشد تا بتوان از این انگاره استفاده کرد. اگر داده‌ها به هر نحوی مغایر با این انگاره باشند باید داده‌های جدید گردآوری کرد.

انگاره‌ی چندوجهی راش کاربردهای مختلفی دارد. برای اولین بار در تعلیم و تربیت، انگل هارد در سلسه مقالاتی که نوشت، توانست قابلیت این انگاره را مشخص کند. انگل هارد (۱۹۹۴) با استفاده از این انگاره توانست چهار خطای رایجی را که ۱۵ مصحح آموزش‌دیده مرتکب می‌شدند، شناسایی کند.

این انگاره در علم پزشکی نیز به کرات به کار برده شده است. اونیل و لانز (۱۹۹۷) از این انگاره استفاده کردند تا بتوانند پایداری سخت‌گیری مصححان را در چند نوبت امتحانات مورد ارزیابی قرار دهند. به این منظور، آنها از ۹ متخصص بافت‌شناسی آموزش‌دیده استفاده کردند تا ۱۵ اسلاید بافت‌شناسی را که دانشجویان بافت‌شناسی بر روی سه فعالیت طراحی کرده بودند، در ۱۷ نوبت در طی ۱۰ سال مورد ارزیابی قرار دهند. نتایج تجزیه و تحلیل داده‌ها نشان داد که مصححان از خود سخت‌گیری نشان می‌دهند و این سخت‌گیری‌ها برای برخی بیشتر و برای برخی کمتر بود و سخت‌گیری در طی ۱۰ سال تقریباً ثابت بود.

انگاره‌ی چندوجهی راش در هنرهای نمایشی نیز به‌کار گرفته شده است. مایفورد (۱۹۹۱) از سه نوع داور در هنرهای نمایشی استفاده کرد تا تک‌گویی‌هایی را که هشت دانش‌آموز دبیرستانی در رشته‌ی هنرهای نمایشی اجرا کرده بودند، مورد ارزیابی قرار دهند. داوران عبات بودند از کارگردانان حرفه‌ای هنرهای نمایشی، علاقه‌مندان به هنرهای نمایشی، و دانشجویان هنرهای نمایشی. این داوران باید فایل‌های ویدیویی دو دقیقه‌ای این هشت دانش‌آموز را دو بار نگاه می‌کردند و به آنها براساس مقیاس رتبه‌ای ۳۴ سؤال‌ی نمره می‌دادند. از این داوران خواسته شد تا پس از یک ماه مجدداً آنها را نگاه کنند و دوباره به آنها براساس همان مقیاس نمره بدهند. نتایج نشان داد که هر سه داور به مراتب سخت‌گیری خاص خود را داشتند و همچنین، سخت‌گیری علاقه‌مندان به هنرهای نمایشی، متفاوت از دو نوع داور دیگر بود و نمره‌ای که این داوران می‌دادند بیشتر نزدیک به نمرات کارگردانان بود. یافته‌ی دیگر این بود که آنها در دور دوم نمره‌دهی، به مراتب بهتر از دو دسته‌ی دیگر عمل می‌کردند.

اخیراً انگاره‌ی چندوجهی راش وارد حوزه‌ی فراگیری زبان دوم نیز شده است. به‌عنوان مثال، ناک (۲۰۱۱) در یک مطالعه‌ی بلند مدت ۱۶ ماهه با استفاده از انگاره‌ی چندوجهی راش تأثیر «بازخورد انفرادی» را بر روی ۱۹ مصحح آموزش‌دیده بررسی کرد. هدف از این مطالعه آن بود که مشخص گردد که بازخورد انفرادی چه تأثیری می‌تواند در تصحیح بعدی مصححان داشته باشد. به این منظور، مصححان با استفاده از مقیاس رتبه‌ای، بخش‌های نوشتاری و گفتاری آزمون شغلی انگلیسی را که در استرالیا برای پذیرش پزشکان مهاجر به این کشور برگزار می‌گردد، مورد ارزیابی قرار دادند. نتایج نشان داد که بازخورد انفرادی تأثیری در کاهش چهار خطای مصحح (سخت‌گیری، سوگیری، منظم بودن، و ارزیابی کلی قبل و بعد از بازخورد انفرادی) ندارد؛ اگرچه مصححان دید مثبتی نسبت به بازخورد انفرادی دارند.

۶. روش پژوهش

در این بخش از پژوهش، به بیان جزئیات کاملی در خصوص نمونه‌ی آماری، روش گردآوری داده‌ها و روش انجام پژوهش پرداخته می‌شود.

۱.۶. جامعه‌ی آماری

۲۸ غیرایرانی (دختر و پسر) از دو گروه آموزشی علوم انسانی و فنی-مهندسی که دوره‌ی فارسی‌آموزی سطح پیشرفته را در مرکز آموزش زبان فارسی دانشگاه بین‌المللی امام خمینی^(۵) در زمستان ۱۳۹۳ می‌گذراندند، در تحقیق حاضر شرکت کردند. سن این فارسی‌آموزان بین ۱۸ تا ۳۰ سال متغیر بود و میانگین سنی آنها ۲۲،۵۲ بود (که البته، هفت فارسی‌آموز (۲۵٪) سن خود را مشخص نکرده بودند). مدت یادگیری زبان

فارسی این زبان‌آموزان از ۳ ماه تا ۲۴ ماه متغیر بود و میانگین یادگیری زبان فارسی ۷,۲۶ ماه بود. اطلاعات مربوط به ملیت و زبان فارسی‌آموزان در جدول (۱) آورده شده است.

جدول ۱. ملیت و زبان مادری فارسی‌آموزان

زبان مادری فارسی‌آموزان			ملیت فارسی‌آموزان		
درصد	تعداد	زبان مادری	درصد	تعداد	ملیت
۵۰٪	۱۴	عربی	۷,۱٪	۲	عراقی
۷,۱٪	۲	کردی	۲۱,۴٪	۶	تاجیکی
۳,۶٪	۱	چینی	۳,۶٪	۱	سودانی
۳,۶٪	۱	کره‌ای	۳,۶٪	۱	عاجی
۲۱,۴٪	۶	فارسی تاجیک	۳,۶٪	۱	قزاقی
۳,۶٪	۱	فرانسوی	۱۷,۹٪	۵	سوری
۳,۶٪	۱	انگلیسی	۳,۶٪	۱	چینی
۳,۶٪	۱	قزاقی	۳,۶٪	۱	کره‌ای
۳,۶٪	۱	مشخص نشده	۱۰,۷٪	۳	لبنانی
			۱۰,۷٪	۳	یمنی
			۷,۱٪	۲	فلسطینی
			۳,۶٪	۱	نیجری
			۳,۶٪	۱	مشخص نشده
۱۰۰٪	۲۸	مجموع	۱۰۰٪	۲۸	مجموع

علاوه بر زبان‌آموزان، دو مصحح فارسی‌زبان خانم نیز برای ارزیابی نوشته‌های فارسی‌آموزان در تحقیق حاضر شرکت کردند. این دو مصحح دارای مدرک تحصیلی کارشناسی ارشد آموزش زبان فارسی به غیرفارسی‌زبانان (آزفا) بودند. هر دو مصحح تقریباً دو سال (مصحح اول یک سال، و مصحح دوم ۲ سال) سابقه‌ی تدریس در زمینه‌ی آموزش نگارش فارسی به غیرفارسی‌زبانان را داشتند.

۲.۶. گردآوری داده‌ها

داده‌های این پژوهش شامل ۵۶ متن بود که توسط فارسی‌آموزان غیرایرانی مرکز آموزش زبان فارسی دانشگاه بین‌المللی امام خمینی^(۵) در چهار کلاس درس نگارش پیشرفته در آذر ماه سال ۱۳۹۳ به رشته‌ی تحریر درآمده بودند. گردآوری داده‌ها در دو مرحله به فاصله‌ی زمانی یک هفته صورت پذیرفت. در مرحله‌ی اول، از فراگیران هر دو گروه علوم انسانی و فنی-مهندسی خواسته شد تا یک متن ۱۵ تا ۲۰ سطری را ظرف مدت ۵۰ دقیقه در مورد این موضوع بنویسند: «تأثیر تکنولوژی، از جمله اینترنت و تلویزیون، بر روابط خانوادگی و اجتماعی افراد چیست؟ به‌طور کامل توضیح دهید».

در مرحله‌ی دوم گردآوری داده‌ها (یک هفته‌ی بعد) از همان فراگیران خواسته شد تا متنی درباره‌ی موضوع زیر بنویسند: «زندگی در شهرهای بزرگ را با زندگی در شهرهای کوچک مقایسه کنید و مزایا و معایب هر یک را بنویسید». در این مرحله نیز نگارش یک متن ۱۵ تا ۲۰ سطری ظرف مدت ۵۰ دقیقه مد نظر پژوهشگر بود. سپس فارسی‌آموزانی که در مرحله‌ی اول یا دوم غایب بودند شناسایی و حذف شدند و در نهایت، ۲۸ نفر شناسایی شدند که در هر دو مرحله و درباره‌ی هر دو موضوع مطرح شده، متن نوشته بودند. هر فارسی‌آموز دو برگه تحویل داده بود که در مجموع ۵۶ برگه برای تصحیح/ارزیابی در نظر گرفته شد.

۳.۶. مقیاس ارزیابی نوشته‌ها

در این تحقیق، از مقیاس رتبه‌ای (پیوست ج) برای ارزیابی متن‌ها استفاده شد. معیارهای ارزیابی این مقیاس عبارت بودند از: «دستور»، «دایره واژگانی»، «انسجام معنایی»، «تنوع محتوا و پرورش موضوع»، و «ظاهر متن». این معیارها از یک تا پنج نمره‌گذاری شدند: ۱ (=خیلی بد)، ۲ (=بد)، ۳ (=متوسط)، ۴ (=خوب)، و ۵ (=خیلی خوب). معیارهای ارزیابی از ارزش یکسانی برخوردار بودند. «روایی» و «پایایی» مقیاس فوق را نیز با استفاده از نرم‌افزار «فاستس»^۱ به دست آوردیم که در بخش بعدی به آن پرداخته می‌شود.

متن‌های فراگیران به دو مصحح فارسی‌زبان داده شد تا آنها را بر اساس مقیاس فوق ارزیابی کنند. قبل از نمره‌گذاری برگه‌ها، توضیحاتی در مورد مقیاس مذکور و نحوه‌ی استفاده از آن به مصححان ارائه شد. برگه‌ی دستورالعمل نمره‌دهی به نوشته‌ها (پیوست د) نیز به آنها داده شد و از آنها خواسته شد تا نمره‌دهی به برگه‌ها را با توجه به توضیحات مندرج در دستورالعمل انجام دهند. همچنین، موارد مبهم برای آنها توضیح داده شد.

۷. تحلیل داده‌ها

در این پژوهش، از نرم‌افزار فاستس نسخه‌ی ۳,۷۱,۴ (طراحی شده توسط لینکر در سال ۲۰۱۴) برای تحلیل داده‌ها استفاده شده است. این نرم‌افزار برای تحلیل داده‌های تشریحی^۲ طراحی شده است که در آنها پاسخ درست یا نادرست وجود ندارد. این نرم‌افزار می‌تواند داده‌های رتبه‌ای را بر روی مقیاسی معروف به «لوجیت»^۳ به داده‌های فاصله‌ای تبدیل کند تا بتوان تأثیر متغیرهای مستقل را بررسی کرد. از جمله قابلیت‌های نرم‌افزار فاستس این است که با استفاده از آن می‌توان سوگیری و میزان آن را مشخص کرد. در

1. Facets

2. polytomous data

3. logits

این پژوهش، برای بررسی سوگیری بین مصححان و موضوعات نگارش، و مصححان و معیارهای ارزیابی، از دو مدل در این نرم افزار به شرح زیر استفاده شده است:

مدل اول، $(Model=?B,,?,?B,?,R5)$ است که در آن، علامت سؤال (؟)، نشانگر متغیر مستقل در این نرم افزار، B نشانگر سوگیری، R نشانگر مقیاس ارزیابی می‌باشد، و عدد ۵ نشانگر نمرات ۱ تا ۵ در مقیاس می‌باشد. در این مدل، عدم فاصله بین دو ویرگول، بیانگر این است که ما نمی‌خواستیم این متغیر نقشی در محاسبات نقش داشته باشد. با استفاده از جدول خروجی نرم افزار فاستس تعامل بین مصححان (متغیر اول) و موضوعات (متغیر چهارم) را که در مدل بالا ترسیم شد مشخص کردیم تا نرم افزار بتواند سوگیری و نوع آن را مشخص کند.

مدل دوم بدین شرح می‌باشد: $Model=?B,,?,?B,R5$. در این مدل نیز علامت سؤال نشانگر متغیر مستقل در این نرم افزار می‌باشد، B نشانگر سوگیری، R نشانگر مقیاس ارزیابی و ۵ نشانگر نمرات از ۱ تا ۵ در مقیاس. در این مدل، عدم فاصله بین ویرگول بیانگر این است که این متغیر نقشی در محاسبات نقش نداشته است. ما با استفاده از جدول خروجی نرم افزار فاستس، تعامل بین مصححان (متغیر اول) و معیارهای ارزیابی (متغیر پنجم) را که در مدل بالا ترسیم شد، مشخص کردیم تا نرم افزار بتواند سوگیری و نوع آن را مشخص کند.

۸. یافته‌های پژوهش

در این بخش به یافته‌های به دست آمده در این تحقیق و آزمون فرضیه‌ها می‌پردازیم؛ اما پیش از پرداختن به پرسش‌ها و فرضیه‌های پژوهش، ابتدا در مورد مقیاس ارزیابی متن نتایجی ارائه می‌شود؛ سپس به توضیح «نقشه‌ی متغیر»^۱، که به «نقشه‌ی رایت»^۲ نیز در انگاره‌ی چندوجهی راش مشهور است، پرداخته می‌شود و در نهایت به پرسش‌های این تحقیق پاسخ داده می‌شود.

۸.۱. ویژگی‌های مقیاس ارزیابی متن

در جدول (۲) اطلاعاتی در مورد ابزار ارزیابی در این پژوهش ارائه شده است.

1. variable map
2. Wright's map

جدول ۲. آماره‌های مربوط به ابزار اندازه‌گیری در این پژوهش

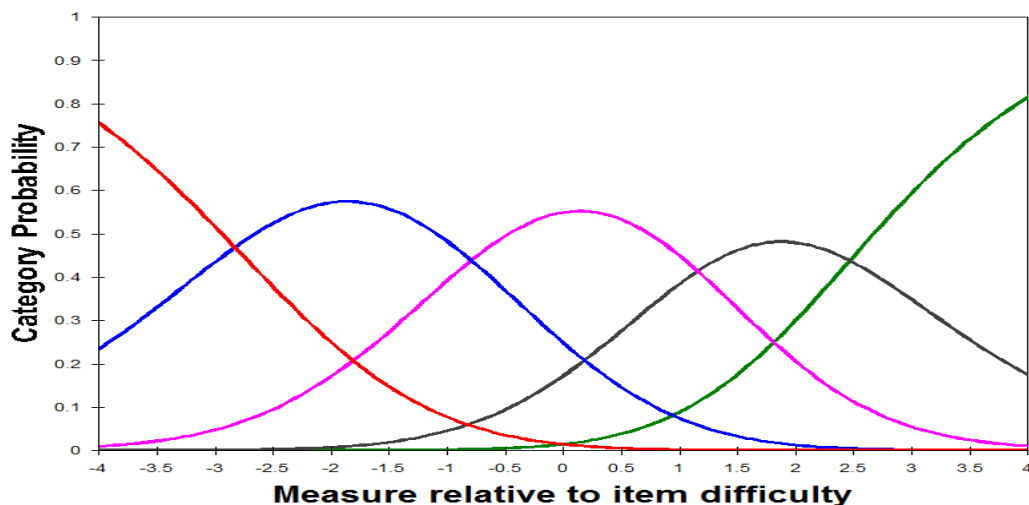
رتبه‌ها ^۱	تعداد نمره‌ها ^۲	اندازه‌گیری متوسط ^۳	اندازه‌گیری مشاهده شده و مورد انتظار ^۴	ترتیب نمره‌ها ^۵
۱	۱۵	-۰.۸۸	۱،۱	Low
۲	۱۱۴	-۰.۵۰	۱،۰	-۲،۸۳
۳	۲۲۷	.۲۹	۱،۰	-۰.۷۹
۴	۱۴۶	۱،۰۸	۱،۰	۱،۱۶
۵	۵۸	۲،۰۷	.۹	۲،۴۷

لینکر (۲۰۰۲ و ۲۰۰۴) معتقد است که مقیاسی نتایج سودمندی ارائه می‌دهد که خصوصیات زیر را داشته باشد: (۱) به هر رتبه حداقل ۱۰ نمره تعلق گرفته باشد؛ (۲) اندازه‌گیری متوسط همراه با رتبه‌ها سیر صعودی داشته باشد؛ (۳) اندازه‌گیری مشاهده شده و مورد انتظار کمتر از ۲ باشد؛ (۴) درجه‌ی دشواری آستانه‌ها حداقل تا ۱،۴ لوجیت، اما کمتر از ۵ لوجیت افزایش داشته باشد. لینکر (۲۰۰۲: ۱۲) می‌افزاید، حداقل افزایش ۱،۴ لوجیت ممکن است برای مقیاس‌های با گزینه‌های بیشتر صدق نکند و «برای مقیاس‌های پنج گزینه‌ای افزایش فاصله بین رتبه‌ها حداقل باید ۱ لوجیت باشد».

طبق این خصوصیات، مقیاس محقق‌ساخته در این پژوهش تمام موارد بالا را فراهم می‌کند. همان‌طور که در جدول (۲) مشاهده می‌شود، تعداد نمره‌ها برای هر رتبه از ۱ تا ۵ بالای ۱۰ می‌باشد؛ اندازه‌گیری متوسط سیر صعودی دارد (یعنی با حرکت از رتبه‌ی ۱ به رتبه‌ی ۲، مقدار اندازه‌گیری متوسط از -۰.۸۸ به -۰.۵۰ می‌رسد که افزایش را نشان می‌دهد). این در مورد تمام مقادیر اندازه‌گیری متوسط صادق است. اندازه‌گیری مشاهده شده و مورد انتظار، همان‌گونه که در ستون چهارم در جدول ۲ مشاهده می‌شود، زیر ۲ می‌باشد. ایده‌آل اندازه‌گیری مشاهده شده و مورد انتظار ۱ می‌باشد. همان‌طور که ملاحظه می‌شود، رتبه‌های ۲، ۳، و ۴ مقدار ۱ دارند و رتبه‌ی ۱ مقدار ۱،۱ و رتبه‌ی ۵ مقدار ۰.۹ دارند که تقریباً خیلی نزدیک به ۱ می‌باشد. ترتیب نمره‌ها نیز مناسب است. برای رتبه‌ی ۱ در ستون پنجم، نرم‌افزار فاستس «Low» را ثبت کرده است. اگر برای رتبه‌ها در ستون پنجم، «No» ثبت شده باشد، نشانگر معیوب بودن مقیاس می‌باشد. همان‌گونه که ملاحظه می‌شود، در ستون پنجم، فاصله بین رتبه‌ها به اندازه‌ی ۱،۴ لوجیت، اما کمتر از ۵ افزایش یافته‌اند، البته فاصله بین رتبه‌ی ۴ و ۵ به اندازه‌ی ۱،۳۱ لوجیت است که با توجه به مقیاس پنج‌گزینه‌ای در این تحقیق بالاتر از ۱ لوجیت می‌باشد.

1. categories
2. counts
3. average measure
4. outfit mean square
5. most probable from

نمودار (۱) نیز به صورت تصویری، کارکرد صحیح مقیاس پنج‌گزینه‌ای ما را در این تحقیق نشان می‌دهد. منحنی‌هایی که در این نمودار هستند اطلاعات سودمندی در مورد مطلوب بودن مقیاس این پژوهش ارائه می‌دهد. این منحنی‌ها باید به شکل یک سری تپه باشند که در غیر این صورت، نشانگر معیوب بودن مقیاس است. همان‌گونه که ملاحظه می‌شود تپه‌ای شکل بودن منحنی‌ها در این نمودار نمایان است.



نمودار ۱. منحنی رتبه‌های مقیاس ارزیابی متن

با توجه به اطلاعات موجود در جدول (۲) و نمودار (۱) نتیجه می‌گیریم که مقیاس محقق ساخته در این پژوهش از مطلوبیت بسیار بالایی برخوردار است و می‌توان به نتایج حاصل از آن در این پژوهش اعتماد کرد.

۲.۸. نقشه‌ی متغیر

نمودار (۲) تمام متغیرهایی را که ما در این تحقیق به کار بردیم، نشان می‌دهد و اطلاعات مهمی را بیان می‌کند که تحلیل فاستس به ما می‌دهد. با استفاده از این نقشه ما می‌توانیم اطلاعاتی را که فاستس در مورد پژوهش به ما می‌دهد به خواننده منتقل کنیم.

در نمودار (۲) از سمت چپ به راست، ستون اول مقیاس لوجیت را نشان می‌دهد. این مقیاس، مقیاس رتبه‌ای است که حکم خط‌کش را دارد که ما متغیرهای مطالعه را نسبت به آن اندازه می‌گیریم. این مقیاس از (۳) تا (۳-) متغیر است. در این مقیاس، صفر نشانگر میانگین می‌باشد. بالای صفر، درجه‌ی سختی مقیاس ارزیابی و سخت‌گیری مصحح را نشان می‌دهد. زیر صفر، درجه‌ی راحتی مقیاس ارزیابی و آسان‌گیری مصحح را نشان می‌دهد.

این مقیاس همچنین، اطلاعاتی در مورد سطح دانش فارسی‌آموزان و سخت‌گیری/آسان‌گیری در تصحیح/ارزیابی موضوع می‌دهد. ستون دوم از سمت چپ، اطلاعات مربوط به مصححانی را نشان می‌دهد که نوشته‌های فارسی‌آموزان را تصحیح کرده‌اند. همان‌گونه که ملاحظه می‌شود مصحح اول مصحح آسان‌گیر می‌باشد؛ زیرا زیر میانگین واقع شده است. این مصحح به خاطر آسان‌گیر بودنش به نوشته‌های فارسی‌آموزان نمره‌های بالایی می‌دهد. اما مصحح دوم مصحح سخت‌گیری است؛ زیرا بالای میانگین واقع شده است. این مصحح به خاطر سخت‌گیر بودنش به نوشته‌های فارسی‌آموزان نمره‌های پایینی می‌دهد.

ستون سوم از سمت چپ، تعداد فارسی‌آموزان را نشان می‌دهد که فاستس با شماره مشخص کرده است. این ستون سطح دانش زبان فارسی‌آموزان را در متن‌های نوشته شده توسط نوشته آنها را نشان می‌دهد. همان‌گونه که در این ستون مشخص شده است، فارسی‌آموزان شماره‌های ۵، ۲، ۱۲، ۴، ۲۶، ۹، ۷، ۳، ۸، ۱۸، ۱۱، و ۱۹ از سطح دانش زبانی بالایی برخوردارند؛ زیرا بالای میانگین واقع شده‌اند، اما فارسی‌آموزان شماره‌های ۲۳، ۱۷، ۱۶، ۶، ۱۴، ۲۷، ۲۴، ۱، ۲۵، ۲۲، و ۲۰ از سطح دانش زبانی پایینی برخوردارند؛ زیرا زیر میانگین واقع شده‌اند. فارسی‌آموزان ۱۰، ۱۵، ۲۱، ۲۸ بین این دو گروه فارسی‌آموز قرار می‌گیرند.

ستون چهارم از سمت چپ، موضوعات را در این پژوهش نشان می‌دهد. این ستون به ما نشان می‌دهد که مصححان در تصحیح/ارزیابی موضوع «زندگی در شهر و روستا» سخت‌گیری به خرج داده‌اند؛ زیرا این موضوع بالای میانگین واقع شده است، اما مصححان در تصحیح/ارزیابی موضوع «تأثیر تکنولوژی بر روابط» سخت‌گیری نکرده‌اند؛ زیرا این موضوع زیر میانگین واقع شده است.

ستون پنجم از سمت چپ، معیارهای ارزیابی را نشان می‌دهد. در این ستون، معیار ارزیابی بالای میانگین نشان‌دهنده‌ی این است که برای فارسی‌آموزان خیلی دشوار بوده است که نمره‌ی بالایی از این معیار بگیرند و معیار زیر میانگین نشان‌دهنده‌ی این است که رعایت کردن این معیار برای فارسی‌آموزان خیلی راحت بوده است تا نمره‌ی بالایی از این معیار بگیرند. بنابراین، معیار ۱ (=دستور) از همه‌ی معیارها سخت‌تر، و معیار ۵ (=ظاهر متن) از همه‌ی معیارها راحت‌تر بوده است. معیارهای ۲ (=دایره واژگانی)، ۳ (=انسجام معنایی)، و ۴ (=تنوع محتوا و پرورش موضوع) نیز بین دو معیار ۱ و ۵ قرار می‌گیرند.

ستون ششم از سمت چپ، نشان‌دهنده‌ی ساختار مقیاس پنج‌گزینه‌ای ما در این پژوهش است. در این پژوهش به معیارهای ارزیابی از ۱ تا ۵ نمره داده شد که این نمره‌ها در این ستون نمایش داده می‌شوند. عدد ۱ نشانگر پایین‌ترین نمره و عدد ۵ نشانگر بالاترین نمره است که در این ستون به خوبی نمایان است.

مقیاس	معیار ارزیابی-	موضوع-	فارسی آموز-	مصحح+الوجیت
(5) +	+	+	+	3 +
4	+	+	+	2 +
---	+	+	5 12 2 26 4 3 7 9 18 8 11 19	1 + 2
3	1 2 3 4 5	زندگی در شهر و روستا	* 10 15 21 28 *	* 0 *
*	*	تأثیر تکنولوژی بر روابط	17 23 16 14 6 1 24 27 13 22 25	1
---	+	+	+	-1 +
2	+	+	+	-2 +
(1)	+	+	+	-3 +
مقیاس	معیار ارزیابی-	موضوع-	فارسی آموز-	مصحح+الوجیت

نمودار ۲. نقشه‌ی متغیر برای متغیرهای پژوهش

۳.۸. سوگیری مصححان نسبت به موضوعات نوشته‌های فراگیران

برای پاسخ دادن به پرسش اول این پژوهش (آیا بین مصححان فارسی‌زبان و موضوع نوشته‌های فارسی‌آموزان غیرایرانی سوگیری وجود دارد؟)، از مدل اول که در قسمت قبل در تحلیل داده‌ها توضیح داده شد، استفاده شد. در این مدل، تعامل بین مصحح و موضوع را مشخص کردیم تا فاستس بتواند سوگیری و مقدار سوگیری را تحلیل کند. جدول (۳) آماره‌های مربوط به این تحلیل را در مورد پرسش اول نشان می‌دهد.

ستون اول از سمت راست، تعداد مصححان را با شماره نشان می‌دهد. ستون دوم، میزان سخت‌گیری و آسان‌گیری مصححان را نشان می‌دهد. همان‌طور که در بالا نیز گفته شد، مصحح اول آسان‌گیر و مصحح دوم سخت‌گیر بوده است. ستون سوم تعداد موضوعات را با شماره نشان می‌دهد. ستون چهارم سخت‌گیری/آسان‌گیری را در تصحیح/ارزیابی موضوعات نشان می‌دهد. همان‌گونه قبلاً گفته شد، مصححان، در تصحیح متن‌های مربوط به موضوع اول، آسان‌گیری، و در خصوص متن‌های مربوط به موضوع دوم سخت‌گیری به خرج داده‌اند. ستون پنجم تعداد نمراتی را نشان می‌دهد که مصححان به فارسی‌آموزان برای پنج معیار ارزیابی در ارتباط با موضوع نگارش مورد نظر این پژوهش داده‌اند. ستون ششم تعداد نمراتی را

نشان می‌دهد که فاستس پیش‌بینی کرده است که باید به فارسی‌آموزان برای پنج معیار ارزیابی در ارتباط با موضوع نگارش مورد نظر این پژوهش داده شود.

جدول ۳. آماره‌های مربوط به سوگیری مصححان نسبت به موضوعات نوشته‌های فراگیران

مصحح	پوچیت	موضوع	پوچیت	نمره‌ی مشاهده شده ^۱	نمره‌ی پیش‌بینی شده ^۲	میانگین نمره‌ی مشاهده شده و نمره‌ی پیش‌بینی شده ^۳	اندازه‌ی سوگیری ^۴	آزمون تی ^۵	خطای انحراف مدل ^۶	چندر میانگین داده‌های متعارف ^۷	چندر میانگین داده‌های غیرمتعارف ^۸
۱	-۰/۲۱	۱	-۰/۳۷	۳۹۲	۴۲۳/۴	-۰/۲۲	۰/۴۲	۳/۶۱	۰/۱۲	۰/۹	۰/۹
۱	-۰/۲۱	۲	-۰/۳۷	۴۰۰	۳۶۸/۷	-۰/۲۲	-۴/۳	-۳/۷۰	۰/۱۲	۰/۷	۰/۷
۲	۱/۱۹	۱	-۰/۳۷	۵۶۳	۵۳۱/۳	-۰/۲۲	-۰/۴۴	-۳/۶۶	۰/۱۲	۱/۱	۱/۱
۲	۱/۱۹	۲	-۰/۳۷	۴۴۳	۴۷۴/۳	-۰/۲۲	۰/۴۱	۳/۵۵	۰/۱۱	۱/۰	۱/۰

توجه^۱. خی‌دو^۹ در سطح ۰/۰۵ معنی دار است.

توجه^۲. موضوعات: ۱=تأثیر تکنولوژی بر روابط، ۲=زندگی در شهر و روستا

ستون هفتم میانگین بین نمره‌هایی که مصحح داده است و نمره‌هایی که فاستس پیش‌بینی کرده است را نشان می‌دهد. این میانگین نشانگر این است که مثلاً مصحح شماره ۱ با میانگین (-۰/۲۲) به موضوع اول نمرات پایین‌تر از حد انتظار داده است و در واقع، در تصحیح این موضوع نسبت به موضوع دیگر سخت‌گیرانه‌تر عمل کرده است. ستون هشتم، اندازه‌ی سوگیری را نشان می‌دهد. این اندازه نشانگر این است که هر مصححی هر موضوع را در حالت کلی به چه صورت ارزیابی کرده است. مصحح شماره ۱ موضوع شماره ۱ را سخت‌گیرانه‌تر از موضوع دیگر تصحیح کرده است.

ستون نهم نشانگر سوگیری می‌باشد. این آماره باید بزرگتر از (+۲) و کوچکتر از (-۲) باشد تا سوگیری وجود داشته باشد. همان‌گونه که در این ستون ملاحظه می‌شود، دو مصحح در این پژوهش نسبت به هر دو موضوع سوگیری نشان داده‌اند. مصحح اول در تصحیح/ارزیابی متن‌های مربوط به موضوع اول سخت‌گیر،

1. observed score

2. expected score

3. observed-expected average

4. bias size

5. t.score

6. model standard error

7. infit mean square

8. outfit mean square

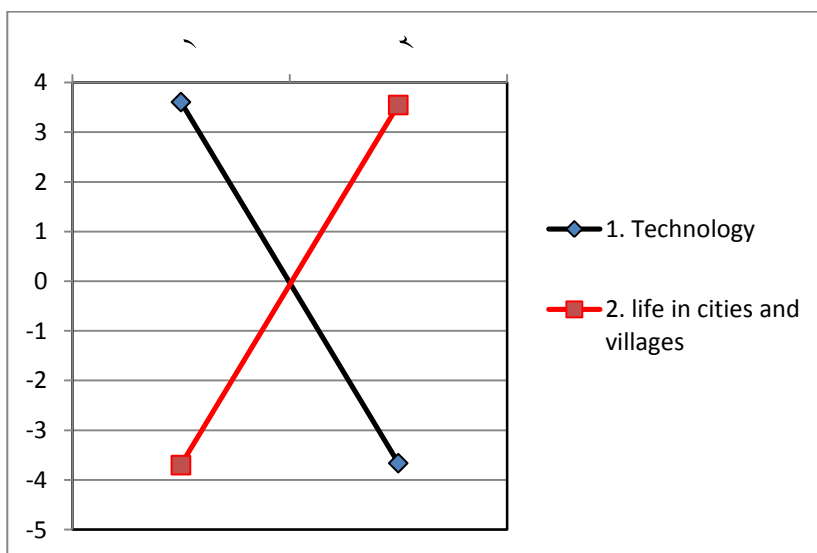
9. chi-square

ولی در تصحیح/ارزیابی متن‌های مربوط به موضوع دوم آسان‌گیرتر بوده است. در عوض، مصحح دوم در مورد موضوع اول آسان‌گیر، و در مورد موضوع دوم سخت‌گیر بوده است. این سوگیری‌ها در سطح ۰.۰۵ معنی‌دار می‌باشند ($\chi^2 = 52.7, df = 4, p = 0.00$). تفاوت معنی‌داری بین سوگیری دو مصحح در مورد موضوع اول وجود دارد (آزمون تی = ۵,۱۴؛ درجه‌ی آزادی = ۲۷۷؛ $P < 0.05$). تفاوت معنی‌داری بین سوگیری دو مصحح در مورد موضوع دوم نیز وجود دارد (آزمون تی = ۵,۱۳؛ درجه آزادی = ۲۷۷؛ سطح معنی‌داری = $P < 0.05$).

دو ستون آخر جدول (۳) به نوعی کنترل کیفیت برای سوگیری‌ها می‌باشد. مقدار ایده‌آل در این دو ستون باید ۱ باشد؛ این مقدار در شرایط خاص متغیر است تا آن جا که از ۰.۵ تا ۲ مشکل‌ساز نیست. ما در این مطالعه مقدار ۰.۵ تا ۱.۵ را در نظر گرفته‌ایم. همان‌طور که در این دو ستون ملاحظه می‌شود مقادیر این دو ستون بین حدی قرار دارند که ما مشخص کرده‌ایم و مبین این مطلب می‌باشند که داده‌های ما داده‌های نسبتاً خوبی هستند و از مقدار خطایی که نرم‌افزار پیش‌بینی کرده است تجاوز نمی‌کنند.

نمودار (۳) سوگیری مصححان را نسبت به موضوعات نوشته‌های فارسی‌آموزان در این پژوهش به صورت تصویری نشان می‌دهد. همان‌طور که پیشتر نیز گفته شد، آماره‌ی بزرگتر از (+۲) و کوچکتر از (-۲) نشانگر سوگیری می‌باشد که این مطلب در این نمودار به وضوح نشان داده شده است.

با توجه به معنی‌دار بودن آزمون خی‌دو و معنی‌دار بودن دو آزمون تی، فرضیه‌ی اول این پژوهش مبنی بر این که مصححان فارسی‌زبان نسبت به موضوعات نگارش فارسی‌آموزان غیرایرانی سوگیری نشان می‌دهند، تأیید می‌گردد.



نمودار ۳. سوگیری مصححان نسبت به موضوعات نوشته‌های فارسی‌آموزان

۴.۸. سوگیری مصححان نسبت به معیارهای ارزیابی

برای پاسخ دادن به پرسش دوم این پژوهش (آیا بین مصححان فارسی‌زبان و معیارهای ارزیابی پاسخ‌ها سوگیری وجود دارد؟) از مدل دوم که در قسمت قبل در بخش تحلیل داده‌ها توضیح داده شد، استفاده کردیم. در این مدل، تعامل بین مصحح و معیار ارزیابی را مشخص کردیم تا فاستس بتواند سوگیری و مقدار سوگیری را تحلیل کند. جدول (۴) آماره‌های مربوط به این تحلیل را در مورد پرسش دوم نشان می‌دهد.

جدول ۴. آماره‌های مربوط به سوگیری مصححان نسبت به معیارهای ارزیابی

مصحح	لوجیت	معیار ارزیابی	لوجیت	نمره‌ی مشاهده شده	نمره‌ی پیش‌بینی شده	میانگین نمره‌ی مشاهده شده و نمره‌ی پیش‌بینی شده	اندازه‌ی سوگیری	آزمون تی	خطای انحراف مدل	جدر میانگین داده‌های غیرمتعارف	جدر میانگین داده‌های متعارف
۱	-۰/۲۱	۳	-۰/۰۹	۱۷۵	۱۶۱/۰۶	۰/۲۵	۰/۴۶	-۲/۵۶	۰/۱۸	۰/۶	۰/۶
۲	۱/۱۹	۳	-۰/۰۹	۱۹۰	۲۰۳/۸۶	۰/۲۵	۰/۴۵	۲/۵۱	۰/۱۸	۱/۱	۱/۱

توجه^۱. خی‌دو در سطح ۰.۰۵ معنی‌دار است.
توجه^۲. معیار ارزیابی: ۳ = انسجام معنایی

ستون اول از سمت راست، تعداد مصححان را با شماره نشان می‌دهد. ستون دوم میزان سخت‌گیری و آسان‌گیری مصححان را نشان می‌دهد. همان‌طور که در بالا نیز گفته شد، مصحح اول آسان‌گیر و مصحح دوم سخت‌گیر بوده است. ستون سوم نیز معیار ارزیابی را با شماره نشان می‌دهد. ستون چهارم راحت و سخت بودن معیار ارزیابی را نشان می‌دهد. همان‌گونه که در این ستون مشاهده می‌شود، معیار ارزیابی ۳ با لوجیت ۰.۰۹- نسبتاً راحت می‌باشد؛ به این معنی که فارسی‌آموزان نمره‌ی بالایی از این معیار گرفته‌اند. ستون پنجم تعداد نمراتی را نشان می‌دهد که مصححان به فارسی‌آموزان برای معیار سوم در این پژوهش داده‌اند. ستون ششم تعداد نمراتی را نشان می‌دهد که فاستس پیش‌بینی کرده است که باید به فارسی‌آموزان به معیار ارزیابی سوم در این پژوهش داده شود. ستون هفتم میانگین بین نمره‌هایی که مصحح داده است و نمره‌هایی که فاستس پیش‌بینی کرده است را نشان می‌دهد. این میانگین نشانگر آن است که مثلاً مصحح شماره ۱ با میانگین ۰.۲۵، به معیار ارزیابی سوم نمرات بالاتر از حد انتظار داده است و در واقع، در ارزیابی این معیار نسبت به معیارهای دیگر، آسان‌گیرتر بوده است.

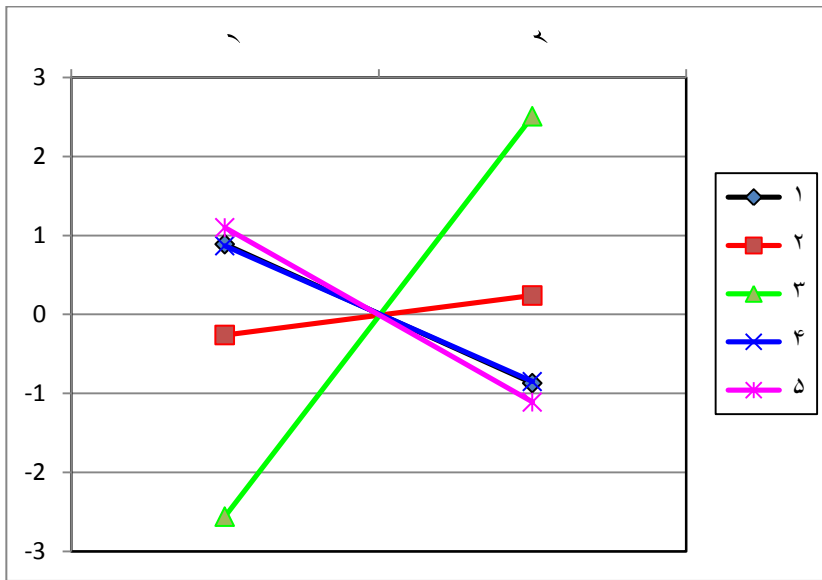
ستون هشتم اندازه‌ی سوگیری را نشان می‌دهد. این اندازه نشانگر این است که هر مصححی هر معیار ارزیابی را در حالت کلی به چه صورت ارزیابی کرده است. مصحح شماره ۱ مثلاً در تصحیح معیار ارزیابی

سوم نسبت به بقیه‌ی معیارهای ارزیابی آسان‌گیری بیشتری به خرج داده است. ستون نهم، آزمون تی را نشان می‌دهد که نشانگر سوگیری می‌باشد. این آماره باید بزرگتر از (+۲) و کوچکتر از (-۲) باشد تا سوگیری وجود داشته باشد. همان‌گونه که در این ستون ملاحظه می‌شود، دو مصحح در این پژوهش فقط نسبت به معیار ارزیابی سوم سوگیری نشان داده‌اند. مصحح اول، در تصحیح معیار سوم آسان‌گیرانه عمل کرده است؛ در عوض، مصحح دوم در تصحیح معیار سوم سخت‌گیر بوده است. این سوگیری‌ها در سطح ۰.۵. معنی‌دار می‌باشند ($\chi^2 = 18.5, df = 10, p = 0.00$). تفاوت معنی‌داری بین سوگیری دو مصحح در مورد معیار سوم وجود دارد (آزمون تی = ۳.۵۸، -، درجه آزادی = ۱۰۹، سطح معنی‌داری = $P < 0.05$).

دو ستون آخر جدول (۴) به نوعی کنترل کیفیت برای سوگیری‌ها می‌باشد. مقدار ایده‌آل در این دو ستون باید ۱ باشد؛ این مقدار در شرایط خاص متغیر است تا آن جا که از ۰.۵ تا ۲ مشکل‌ساز نیست. ما در این مطالعه مقدار ۰.۵ تا ۱.۵ را در نظر گرفته‌ایم. همان‌طور که در این دو ستون ملاحظه می‌شود، مقادیر این دو ستون بین حدی که ما مشخص کرده‌ایم قرار دارند و مبین این مطلب می‌باشند که داده‌های ما داده‌های نسبتاً خوبی هستند و از مقدار خطایی که نرم‌افزار پیش‌بینی کرده است تجاوز نمی‌کنند.

نمودار (۴) سوگیری مصححان را نسبت به معیارهای ارزیابی در این پژوهش به صورت تصویری نشان می‌دهد. همان‌طور که در پیشتر نیز گفته شد آماره‌ی بزرگتر از (+۲) و کوچکتر از (-۲) نشانگر سوگیری می‌باشد که در این نمودار به وضوح نشان داده شده است. در سمت راست نمودار معیارهای ارزیابی از ۱ تا ۵، و در سمت بالای نمودار دو مصحح نشان داده شده است. چنان که ملاحظه می‌شود معیارهای ۱، ۲، ۴، و ۵ بین (+۲) و (-۲) قرار می‌گیرند و بنابراین، سوگیری نسبت به این معیارها وجود ندارد. فقط معیار سوم است که مصحح اول در تصحیح آن آسان‌گیری و مصحح دوم، سخت‌گیری نشان داده است.

با توجه به معنی‌دار بودن آزمون خی‌دو و معنی‌دار بودن آزمون تی، فرضیه‌ی دوم این پژوهش مبنی بر این که مصححان فارسی‌زبان معیارهای ارزیابی پاسخ‌ها را یکسان ارزیابی نمی‌کنند تأیید می‌گردد.



نمودار ۴. سوگیری مصححان نسبت به معیارهای ارزیابی

۹. کاربردها و محدودیت‌های پژوهش

با توجه به نتایج حاصل از پژوهش حاضر، می‌توان به کاربردهای آموزشی زیر اشاره کرد. اولین کاربرد آموزشی این پژوهش برای تربیت مصحح/ارزیاب می‌باشد. همان‌طور که در قسمت بحث و نتیجه‌گیری اشاره خواهد شد، خطاهای مصحح از جمله سوگیری می‌توانند نمرات فراگیران را تا حد زیادی تحت تأثیر قرار دهند. بنابراین، نیاز است که به مصححان قبل از تصحیح اوراق امتحانی آموزش لازم و کافی داده شود تا آنها کمترین میزان خطا را در زمان ارزیابی و نمره‌دهی داشته باشند. این آموزش طبق نتایج این پژوهش می‌تواند به صورت مختلف انجام شود. می‌توان نمونه‌های تصحیح شده‌ی نوشته‌های فراگیران را در اختیار مصححان قرار داد تا با نحوه‌ی نمره‌دهی خوب و نمره‌دهی ضعیف آشنا شوند و دقیق‌تر نمره بدهند. همچنین، می‌توانیم با برگزاری جلسه‌ای قبل از تصحیح اوراق امتحانی به آنها چند نمونه نوشته بدهیم و از آنها بخواهیم که تصحیح‌شان کنند. برای این کار می‌توانیم معیارهای دقیق‌تری طراحی کنیم و برای آن معیارها دستورالعمل‌های مخصوص بنویسیم و از مصححان بخواهیم تا طبق همان دستورالعمل‌ها نمره بدهند. برای هرچه دقیق‌تر شدن نمرات و یک‌دست شدن آنها می‌توان مصححان را در حین ارزیابی نوشته‌ها کنترل کرد تا در صورت رعایت نکردن معیارها و دستورالعمل‌ها، به آنها تذکر داده شود تا طبق دستورالعمل و چارچوب ارائه شده عمل کنند.

دومین کاربرد مربوط به طراحی مقیاس‌های ارزیابی برای نمره‌دهی دقیق به نوشته‌ها می‌باشد. می‌توان از آماره‌هایی که در انگاره‌ی چندوجهی راش که در نرم‌افزار فاستس گنجانده شده است استفاده کرد تا بتوان

مقیاس‌های دقیق‌تر و با کارکرد بالاتری طراحی کرد تا مصححان بتوانند استفاده کنند و نمرات دقیق‌تری به نوشته‌ها بدهند. آماره‌های موجود در این نرم‌افزار به ما می‌گویند که تا چه اندازه مقیاس طراحی شده دقیق است و در صورت معیوب بودن مقیاس می‌توان با نرم‌افزار آن را ارتقا بخشید تا برای نمره‌دهی، مناسب شود. مقیاس‌های دقیق و با روایی و پایایی بالا در آزمون‌های تشریحی بسیار حائز اهمیت هستند؛ زیرا در صورت معیوب و نادقیق بودن، مقیاس‌ها می‌توانند خود باعث بالا و پایین شدن نمرات فراگیران شده و در نتیجه، به یک فراگیر نمره‌ی بیشتر از استحقاق و به فراگیری دیگر نمره‌ای کمتر از استحقاق داده شود.

با نتایج این تحقیق باید با احتیاط برخورد کرد و از تعمیم‌دهی بیش از اندازه اجتناب کرد؛ زیرا تحقیق حاضر دارای محدودیت‌هایی نیز است. یکی از این محدودیت‌ها تعداد کم مصححان در این تحقیق بود. ما فقط از دو مصحح خانم در این تحقیق استفاده کردیم و بنابراین، نتوانستیم تأثیر جنسیت مصحح را بر سوگیری بسنجیم. بنابراین، توصیه می‌شود در تحقیقات بعدی از تعداد بیشتری از مصححان خانم و آقا استفاده شود تا بتوان تأثیر جنسیت را نیز بررسی کرد. محدودیت دیگر این تحقیق این بود که ما فقط از یک نوع مقیاس استفاده کردیم. نتایج استفاده از مقیاس‌های نوع دیگر می‌تواند جالب باشد. بنابراین، توصیه می‌شود تا در تحقیقات بعدی از دو یا چند مقیاس ارزیابی استفاده شود تا تأثیر آنها را در سوگیری بررسی کرد.

۱۰. بحث و نتیجه‌گیری

هدف از پژوهش این بود که سوگیری بین مصححان فارسی‌زبان نسبت به «موضوعات نوشته‌های فارسی‌آموزان غیرایرانی» و «معیارهای ارزیابی نگارش» با استفاده از انگاره‌ی چندوجهی راش مورد بررسی قرار گیرد. بررسی سوگیری با استفاده از این انگاره مزایایی دارد؛ همان‌گونه که باسترک (۲۰۰۸: ۴۴۰) به درستی یادآوری می‌کند، با استفاده از این انگاره ما می‌توانیم وجوه مختلف داده‌ها را به صورت موردی بررسی کنیم که در روش سنتی امکان‌پذیر نیست. علاوه بر این، شیفر (۲۰۰۸: ۴۶۶) نتیجه‌گیری می‌کند که با استفاده از این انگاره، محققان می‌توانند به کنکاش و فهم منابع مختلف سوگیری مصحح بپردازند تا بتوانند راهکارهایی جهت بهبود تربیت مصحح و طراحی مقیاس‌های معتبرتر ارائه دهند.

در این پژوهش ما به یافته‌های جالبی دست پیدا کردیم. با توجه به معنی‌دار بودن آزمون خی‌دو در مورد سوگیری‌های مصححان نسبت به موضوعات نوشته‌های فارسی‌آموزان و همچنین معنی‌دار بودن دو آزمون تی در مورد سوگیری مصححان و موضوع اول و سوگیری مصححان و موضوع دوم، یافته‌ی نخست این بود که هر دو مصحح نسبت به هر دو موضوع سوگیری نشان دادند. این یافته همسو با یافته‌های مطالعات قبلی است (لاملی، ۲۰۰۵؛ ویگلزورث، ۱۹۹۴ و ۱۹۹۳؛ دو و همکاران، ۱۹۹۶). مصحح اول در تصحیح موضوع

اول، سخت‌گیر و در تصحیح موضوع دوم، آسان‌گیر بود؛ در صورتی که مصحح دوم در ارزیابی موضوع اول، آسان‌گیر و در ارزیابی موضوع دوم، سخت‌گیر بود. در حقیقت، مصحح دوم برعکس مصحح اول عمل کرده است. می‌توان این گونه بحث کرد که موضوع برای مصحح می‌تواند جذاب باشد و به خاطر همین جذابیت، ممکن است مصحح، متن را آسان‌گیرانه‌تر از حد انتظار تصحیح کند. عکس این حالت نیز صادق است؛ بدین صورت که موضوع ممکن است برای مصحح خسته‌کننده باشد و بنابراین، ممکن است آن را سخت‌گیرانه‌تر از حد انتظار تصحیح کند. این موضوع در مطالعات قبلی نیز تأیید شده است (ر.ک. بکمن، ۱۹۹۰؛ بارکویی، ۲۰۱۰؛ لاملی و مک‌نامرا، ۱۹۹۵).

دلیل دیگری که مصححان در این پژوهش به این دو موضوع سوگیری نشان دادند ماهیت خود این موضوعات بود. به لحاظ جدید بودن تکنولوژی، فارسی‌آموزان از کشورهای مختلف ممکن است آشنایی کافی با تکنولوژی نداشته باشند و در نتیجه، ممکن است اطلاعات کافی در مورد آن نداشته باشند و نتوانند موضوع را آن طور که شایسته است پرورش دهند. این باعث می‌شود مصحح به آنان نمره‌ی کمی بدهد و در تصحیح متن، سخت‌گیری شود. عکس این حالت نیز صادق است. بدین معنی ممکن است که فارسی‌آموزان زندگی در شهر و روستا را تجربه کرده باشند و برای آنها ملموس باشد و در نتیجه، بتوانند با اطلاعات بیشتر و بهتری به نگارش درباره‌ی موضوع بپردازند که خود این امر باعث می‌شود مصحح به آنها نمره بیشتری داده و متن را آسان‌گیرانه تصحیح کند.

یافته‌ی دوم در این پژوهش (با توجه به معنی‌دار بودن آزمون‌های دو در مورد سوگیری‌ها و آزمون تی در مورد سوگیری مصححان و معیار سوم) این بود که هر دو مصحح فقط به معیار سوم یعنی انسجام معنایی سوگیری نشان دادند. مصحح اول در خصوص انسجام معنایی، آسان‌گیری و مصحح دوم در خصوص همین معیار، سخت‌گیری نشان دادند. این یافته با یافته‌های مطالعات قبلی (شیفر، ۲۰۰۸؛ ماتسونو، ۲۰۰۹) همسو است. نکته‌ی جالب در این پژوهش آن بود که به اذعان خود مصححان، بعد از تصحیح نوشته‌ها انسجام معنایی برای هر دوی آنها مهم‌ترین معیار ارزیابی تلقی می‌شد. نکته‌ی جالب‌تر این جا است که خود مصححان از این که متوجه شدند فقط به این معیار سوگیری نشان داده‌اند تعجب کردند؛ زیرا اعتقاد داشتند که آنها همه‌ی معیارها را مد نظر قرار داده‌اند. این در واقع، تأیید یافته‌ی مک‌نامرا (۱۹۹۶) و لاملی (۲۰۰۵) است که دریافتند مصححان به برخی از معیارها مثلاً دستور، سوگیری نشان می‌دهند، ولی خود آنها متوجه این موضوع نیستند.

برای این مسأله که مصححان در این پژوهش به موضوعات نوشته‌ها و معیارهای ارزیابی سوگیری نشان دادند می‌توان دلایل دیگری را نیز ذکر کرد؛ دلیل اول تجربه‌ی تصحیح کردن می‌باشد. هر دو مصحح تجربه‌ی تصحیح در محیط‌های ارزیابی مهم را نداشتند و تجربه‌ی آنها محدود بود به ارزیابی نوشته‌های

کلاسی فارسی‌آموزان در مرکز زبان دانشگاه بین‌المللی امام خمینی^(۵). تحقیقات نشان داده است که مصححانی که تجربه‌ی تصحیح کردن به مدت طولانی‌مدت در محیط‌های ارزیابی مهم را داشته‌اند، سوگیری کمتری نسبت به مصححان بی‌تجربه یا کم‌تجربه نشان می‌دهند (ر.ک. بارکویی، ۲۰۱۰). دلیل دیگر مقدار آموزش به مصححان می‌تواند باشد. در این پژوهش فقط در حدود ۳۰ دقیقه به مصححان در خصوص چگونگی استفاده از معیارهای ارزیابی مقیاس پنج‌گزینه‌ای برای نمره‌دهی به نوشته‌های فارسی‌آموزان، آموزش داده شد.

تا آن جا که نگارنده اطلاع دارد پژوهشی در زمینه‌ی سنجش سوگیری مصححان فارسی‌زبان با استفاده از انگاره‌ی چندوجهی راش وجود ندارد. بنابراین، پژوهش حاضر می‌تواند شروع خوبی برای تحقیقاتی از این دست باشد تا بتوان به نتایج بهتر و گسترده‌تری در این زمینه دست یافت. نکته‌ی حائز اهمیت این است که با استفاده از انگاره‌ی چندوجهی راش می‌توان سوگیری مصححان را شناسایی نموده و در جهت رفع این نوع خطا گام‌های اساسی برداشت. سوگیری به‌عنوان خطای مصحح می‌تواند اثرات زینباری برای فراگیران در پی داشته باشد و در نتیجه، فراگیران حتی اگر از توانایی بالایی هم برخوردار باشند، قربانی خطای مصححان شوند. نکته‌ی مهم دیگر این است که اغلب، مصححان از سوگیری خودشان مطلع نیستند؛ بنابراین، می‌توان سوگیری و سایر خطاهای مصححان را با استفاده از این انگاره کشف نموده و به آنان اعلام کرد.

تقدیر و تشکر

بر خود لازم می‌دانم از جناب آقای دکتر امیررضا وکیلی‌فرد، رئیس محترم مرکز آموزش زبان فارسی دانشگاه بین‌المللی امام خمینی^(۵) به خاطر اجازه‌ی گردآوری داده‌ها در این مرکز تشکر کنم. همچنین، از زحمات جناب آقای مهدی خدادادیان که زحمت بیش از اندازه کشیدند و هماهنگی‌های لازم را برای گردآوری داده‌ها با مدرسان زبان در این مرکز انجام دادند تقدیر می‌گردد. ایشان زحمت گردآوری داده‌ها را در دو نوبت مختلف کشیدند و همچنین بنده را در معرفی کردن مصححان، بسیار یاری نمودند.

منابع:

- Andrich, D.** (1978). A Rating Formulation for Ordered Response Categories. *Psychometrika*, 43(4): 561-573.
- Bachman, L.** (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Basturk, R.** (2008). Applying The Many-Facet Rasch Model to Evaluate PowerPoint Presentation Performance in Higher Education. *Assessment & Evaluation in Higher Education*, 33(4): 431-444.
- Baghaei, P. & N. Amrahi.** (2009). Introduction to Rasch Measurement. *The Iranian EFL Journal*, 5: 139-154.

- Barkaoui, K.** (2010). Do ESL Essay Raters' Evaluation Criteria Change with Experience? A Mixed-methods, Cross-Sectional Study. *TESOL Quarterly*, 44(1): 31-57.
- Du, Y. et al.** (1996). *Differential Facet Functioning Detection in Direct Writing Assessment*. Paper presented at the Annual Meeting of the American Educational Research Association, New York.
- Eckes, T.** (2009). Many-facet Rasch measurement. In S. Takala (Ed.), *Reference Supplement to the Manual for Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, teaching, Assessment (Section H)*. Strasbourg, France: Council of Europe/Language Policy Division. Retrieved from <http://www.coe.int/t/dg4/linguistic/Source/CEF-refSupp-SectionH.pdf>
- Englehard, G.** (1994). Examining Rater Errors in the Assessment of Written Composition with A Many-Faceted Rasch model. *Journal of Educational Measurement*, 31(2): 93-112.
- Esfandiari, R. & Myford, C. M.** (2013). Severity Differences Among Self-Assessors, Peer-Assessors, and Teacher Assessors Rating EFL Essays. *Assessing Writing*, 18(2): 111-131.
- Farrokhi, F. et al.** (2012). A Many-Facet Rasch Measurement of Differential Rater Severity/Leniency in Three Types of Assessment. *JALT Journal*, 34(1): 79-102.
- Knoch, U.** (2011). Investigating The Effectiveness of Individualized Feedback to Rating Behavior- A Longitudinal Study. *Language Testing*, 28(2): 179-200.
- Linacre, J. M.** (1989/1994). *Many-Facet Rasch Measurement*. Chicago: MESA Press.
- Linacre, J. M.** (1997). *Judging Plans and Facets* (Research Note No. 3). Chicago: University of Chicago, MESA Psychometric Laboratory. Retrieved from <http://www.rasch.org/rn3.htm>.
- Linacre, J. M.** (2002). Optimizing Rating Scale Category Effectiveness. *Journal of Applied Measurement*, 3(1): 85-106.
- Linacre, J. M.** (2014). *FACETS (Version 3.71.4) [Computer Software]*. Chicago, IL: MESA Press.
- Lumley, T.** (2005). *Assessing Second Language Writing: The Rater's Perspective*. Frankfurt am Main: Peter Lang.
- Lumley, T. & McNamara, T. F.** (1995). Rater Characteristics and Rater Bias: Implications for Training. *Language Testing*, 12(1): 54-71.
- Masters, G.** (1982). A Rasch Model For Partial Credit Scoring. *Psychometrika*, 47(2):149-174.
- Matsuno, S.** (2009). Self-, Peer-, and Teacher-Assessments in Japanese University EFL Writing Classrooms. *Language Testing*, 26(1): 75-100.
- McNamara, T. F.** (1996). *Measuring Second Language Performance*. New York, NY: Longman.
- Myford, C. M.** (1991). *Judging Acting Ability: Moving from Novice to Expert*. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.
- Myford, C. M. & Wolfe, E. W.** (2004). Detecting and measuring rater effects using many facet rasch measurement: Part II. In E. V. Smith & R. M. Smith (Eds.), *Introduction to Rasch Measurement* (pp. 460-517). Maple Grove, MN: JAM Press.

- O'Neill, T. K. & Lunz, M. E.** (1997). *A Method to Compare Rater Severity Across Several Administrations*. Paper Presented at The Annual Meeting of The American Educational Research Association, Chicago, IL.
- Schaefer, E.** (2008). Rater Bias Patterns in An EFL Writing Assessment. *Language Testing*, 25(4): 465-493.
- Scullen, S. E. et al.** (2000). Understanding The Latent Structure of Job Performance Ratings. *Journal of Applied Psychology*, 85: 956-970.
- Smith, R.** (2004). Fit analysis in latent trait measurement models. In E. Smith & R. Smith (Eds.), *Introduction To Rasch Measurement* (pp. 51-83). Maple Grove: JAAM Press.
- Wigglesworth, G.** (1993). Exploring Bias Analysis As A Tool For Improving Rater Consistency in Assessing oral Interaction. *Language Testing*, 10(3): 305-335.
- Wigglesworth, G.** (1994). Patterns of Rater Behaviour in The Assessment of An Oral Interaction Test. *Australian Review of Applied Linguistics*, 17(2): 77-103.
- Winke, P. et al.** (2013). Raters' L2 Background As A Potential Source of Bias in Rating Oral Performance. *Language Testing*, 30(2): 231-252.
- Wright, B. D. & Stone, M. H.** (1979). *Best Test Design*. Chicago: MESA Press.

پیوست الف: موضوع اول

آزمون نگارش زبان آموزان دوره‌ی تکمیلی (آذرماه ۱۳۹۳)

زبان آموزان گرامی، لطفاً به‌طور دقیق، اطلاعات درخواستی زیر را وارد کنید.

نام: نام خانوادگی: ملیت:

جنس: زن مرد سن: سال

زبان مادری (زبانی که نخستین بار به آن تکلم کردید):

فارسی ترکی عربی کردی چینی کره‌ای فارسی تاجیکی فارسی دری

فرانسوی اسپانیولی انگلیسی قزاقی

سایر (لطفاً مشخص کنید:

چه مدتی است که زبان فارسی را یاد می‌گیرید؟ ماه

لطفاً در مورد موضوع زیر، یک متن ۱۵ تا ۲۰ سطری بنویسید:

تأثیر تکنولوژی از جمله اینترنت و تلویزیون بر روابط خانوادگی و اجتماعی افراد چیست؟ به‌طور کامل توضیح دهید.

پیوست ب: موضوع دوم

آزمون نگارش زبان آموزان دوره‌ی تکمیلی (دی ماه ۱۳۹۳)

زبان آموزان گرامی، لطفاً به‌طور دقیق، اطلاعات درخواستی زیر را وارد کنید.

نام: نام خانوادگی: ملیت:

جنس: زن مرد سن: سال

زبان مادری (زبانی که نخستین بار به آن تکلم کردید):

فارسی ترکی عربی کردی چینی کره‌ای فارسی تاجیکی فارسی دری

فرانسوی اسپانیولی انگلیسی قزاقی

سایر (لطفاً مشخص کنید:

چه مدتی است که زبان فارسی را یاد می‌گیرید؟ ماه

لطفاً در مورد موضوع زیر، یک متن ۱۵ تا ۲۰ سطری بنویسید:

زندگی در شهرهای بزرگ را با زندگی در شهرهای کوچک مقایسه کنید و مزایا و معایب هر یک را بنویسید.

پیوست ج: مقیاس ارزیابی متن

نمره‌گذاری از ۱ تا ۵					
خیلی خوب	خوب	متوسط	بد	خیلی بد	معیارهای ارزیابی متن
۵□	۴□	۳□	۲□	۱□	دستور
۵□	۴□	۳□	۲□	۱□	دایره واژگانی
۵□	۴□	۳□	۲□	۱□	انسجام معنایی
۵□	۴□	۳□	۲□	۱□	تنوع محتوا و پرورش موضوع
۵□	۴□	۳□	۲□	۱□	ظاهر متن

پیوست ۵: دستورالعمل نمره‌دهی به نوشته‌ها

(۱) دستور

۱. تطابق نادرست فعل و فاعل
۲. کاربرد نادرست زمان افعال
۳. استفاده‌ی نادرست از ترکیبات وصفی و اضافی
۴. استفاده‌ی نادرست از حروف اضافه‌ی مناسب
۵. آرایش جمله؛ جایگاه نادرست فعل و فاعل

(۲) دایره واژگانی

۱. به‌کارگیری واژه‌های نادرست در جمله
۲. به‌کارگیری نادرست واژه‌ها؛ مثل استفاده از صفت به جای اسم در جمله
۳. عدم تنوع واژگانی
۴. استفاده از واژه‌ها و عبارات تکراری
۵. به‌کار نبردن ترکیبات و اصطلاحات رایج زبان فارسی

(۳) انسجام معنایی

۱. استفاده از عوامل انسجام معنایی؛ مانند «اما»، «بنابراین» و...
۲. ارتباط معنایی بین جملات
۳. استفاده از جملات و ایده‌ای مرتبط با موضوع
۴. برقراری وحدت بین متن و موضوع

(۴) تنوع محتوا و پرورش موضوع

۱. استفاده از ایده‌های اصلی در پاراگراف‌ها
۲. استفاده از تنوع ایده‌های فرعی؛ مانند «مثال»، «نقل قول»، «ذکر آمار و ارقام» و...
۳. تازگی مطالب و تکراری نبودن آنها

(۵) ظاهر متن

۱. مرتب بودن متن و جدا کردن پاراگراف‌ها
۲. به‌کاربردن علائم سجاوندی؛ مانند «نقطه»، «علامت سؤال» و...
۳. طول متن: متن باید بین ۱۵ تا ۲۰ سطر باشد.
۴. غلط‌های املائی