



رواسازی آزمون مبتنی بر چهارچوب مرجع اتحادیه اروپا و برابری ارزشیابی

سید محمد علوی *

پروفسور آموزش زبان انگلیسی دانشگاه تهران، تهران، ایران
Email: smalavi@ut.ac.ir



حسین کرمی **

استادیار آموزش زبان انگلیسی دانشگاه تهران، تهران، ایران
Email: hkarami@ut.ac.ir



داود سپاهی ***

(نویسنده مسئول)
دانشجوی دکتری تخصصی آموزش زبان انگلیسی دانشگاه تهران پردیس بین‌المللی کیش، تهران، ایران
Email: dsepahi@ut.ac.ir



چکیده

برابری در ارزشیابی از جمله اهداف مهم در زمینه آموزش و سنجش زبان می‌تواند باشد. متخصصان ارزشیابی به رابطه بین روایی و برابری اذعان می‌کنند. بعضی معتقدند روایی خود بخشی از برابری است. از سوی دیگر، بعضی بر این باورند که دستیابی به برابری در واقع همان کسب روایی آزمون است. چهارچوب مرجع اتحادیه اروپا (چما) برای دستیابی به برابری آموزشی و ارزشیابی و یادگیری سامانه‌ای از توصیفگرها تهیه و تدوین کرده است. هدف از این تحقیق رواسازی آزمونی شامل ۲۵ آیتم است که بر اساس توصیفگرهای چما ساخته شده است. برای دستیابی به اطلاعات لازم به منظور تعیین میزان روایی آزمون ۱۵۰ نفر شرکت‌کننده زن و مرد به سؤالات چندگزینه‌ای پاسخ دادند. سپس، داده‌های به‌دست‌آمده برای تعیین میزان افتراق عملکرد هریک از آیتم‌ها با روش نظریه سؤال پاسخ (نسب) تحلیل شد. علاوه بر این، از پرسش‌نامه ۷۵ آیتم (دورنیه) برای تفکیک گروه‌های شرکت‌کنندگان بر اساس علائق و اهداف استفاده شد. نتایج به‌دست‌آمده نشان داد علاوه بر معیارهای معمول برای دسته‌بندی شرکت‌کنندگان اعم از جنس و پیشینه زبانی و فرهنگی، علائق و اهداف شخصی افراد می‌تواند در نظر گرفته شود. همچنین روشن شد که این معیارها برای میزان روایی آزمون می‌تواند به کار برود.

اطلاعات مقاله

تاریخ ارسال: ۱۳۹۷/۱۰/۰۱
تاریخ پذیرش: ۱۳۹۹/۰۴/۲۴
تاریخ انتشار: بهار ۱۴۰۱
نوع مقاله: علمی پژوهشی

کلیدواژگان:

«برابری»، «چما»،
«ارزشیابی»، «کارکرد افتراقی»
«سؤال»، «نظریه سؤال پاسخ»

شناسه دیجیتال DOI: 10.22059/JFLR.2020.272015.588

علوی، سید محمد، کرمی، حسین، سپاهی، داود. (۱۴۰۱). رواسازی آزمون مبتنی بر چارچوب مرجع اتحادیه اروپا و برابری ارزشیابی. پژوهش‌های زبان‌شناختی در زبانهای خارجی، ۱۲ (۱)، ۱۱۱-۱۳۱.

Alavi, D., Karami, H., Sepahi, D. (2022). CEFR based test Validation and Fairness Assessment. Foreign Language Research Journal, 12 (1), 111-131.

* پروفسور آموزش زبان انگلیسی دانشگاه تهران، تهران، ایران

** استادیار آموزش زبان انگلیسی دانشگاه تهران، تهران، ایران

*** دانشجوی دکتری تخصصی آموزش زبان انگلیسی دانشگاه تهران پردیس بین‌المللی کیش، تهران، ایران



Validation of a CEFR Based Test and Fairness in Assessment



Sayyed Mohamad Alavi *
Professor of TEFL University of Tehran
Email: smalavi@ut.ac.ir



Hossein Karami **
Assistant Professor of TEFL University of Tehran
Email: hkarami@ut.ac.ir



Davood Sepahi ***
(corresponding author)
Ph.D. Candidate University of Tehran Kish International Campus
Email: dsepahi@ut.ac.ir

ABSTRACT

Fairness in assessment can be one of the important goals in the field of language teaching and assessment. Assessment experts acknowledge the relationship between validity and fairness. Some believe that validity itself is part of fairness. On the other hand, some believe that achieving fairness is in fact the achieving a test validity. The Common European Framework of Reference (CEFR) has developed a system of descriptors to achieve educational equality and assessment of learning. The purpose of this research is to validate a test consisting of 25 items based on CEFR descriptors. In order to obtain the necessary information to determine the validity of the test, 150 male and female participants answered multiple-choice questions. Then, the data were analyzed to determine the differential functioning of each item (DIF) by item response theory (IRT). In addition, a 75-item questionnaire was used to separate groups of participants based on their interests and goals. The results showed that in addition to the usual criteria for categorizing participants, including gender and linguistic and cultural background, personal interests and goals can be considered. It also became clear that these criteria could be used for determining test validity.

ARTICLE INFO

Article history:
Received: December 22, 2018

Accepted: July 14, 2020

Available online:
Spring2022

Keywords:

"Fairness", "CEFR",
"Assessment",
"Differential Functioning", "Item Response Theory", "Item"

DOI: 10.22059/JFLR.2020.272015.588

Alavi, D., Karami, H., Sepahi, D. (2022). CEFR based test Validation and Fairness Assessment. Foreign Language Research Journal, 12 (1), 111-131.

* Sayyed Mohamad Alavi Professor University of Tehran

** Hossein Karami Assistant Professor University of Tehran

*** Davood Sepahi Ph.D. Candidate University of Tehran

بندی سطوح توانایی زبان وفق معیارهای چما قابل توجه است. به این ترتیب حیطة کاربرد چما به طور مضاعف با این آزمون‌های شناخته شده افزایش یافته، به طوری که زبان-آموزان می‌توانند سطح آموخته‌های خود را نسبت به این آزمون‌ها بسنجند. همچنین، ناشران معروف بین‌المللی نیز کتاب‌های آموزشی خود را نسبت به چما تعریف و هم-راستا کرده‌اند که باز به گستره کاربرد این چهارچوب می-افزاید.

زمینه به کار بستن چما تنها محدود به حیطة آموزش زبان نیست بلکه در خصوص تهیه آزمون برای دانشجویان داوطلب تحصیل به زبان انگلیسی در رشته‌های غیر زبان هم به کار رفته است. شا و امام (۲۰۱۳) در تحقیقی نشان دادند چما برای آموذن سطح مهارت شناختی زبان آکادمیک داوطلبان می‌تواند به کار برود. از نظر آنها حداقل مهارت زبانی مورد نیاز برای داوطلبان سطح ب ۲ است. همچنین سطح مهارت س ا هم می‌تواند زمینه خوبی برای پیشرفت در یادگیری موضوعاتی چون تاریخ و جغرافی باشد. به پیشنهاد ایشان چما می‌تواند زمینه‌ای برای تهیه بسته زبانی مناسب برای محتوای آموزشی غیرزبانی باشد. یعنی، آموزگاران که محتوای غیرزبانی -مثلاً تاریخ یا جغرافی- درس می‌دهند به کمک این بسته زبانی می‌توانند محتوای آموزشی را به دانشجویان غیرانگلیسی‌زبان بهتر بیاموزند.

از جنبه دیگر کاربرد چما تاکنون بر سنجش زبان تأثیر شگرفی گذاشته است (لیتل، ۲۰۱۴). از نظر لیتل زبان-آموزی، و ارزشیابی در پیوندی تنگاتنگ قرار دارند. در این دیدگاه زبان‌آموز خود عهده دار یادگیری است. نظارت شخصی بر یادگیری زبان‌آموز با استفاده از جدول‌های «من می‌توانم» (can do) انجام‌پذیر است. در این زمینه تحقیقاتی انجام شده که مؤید نظر لیتل است.

ویژگی‌های چما

در اسناد مربوط به چما آمده است که توصیف موقعیت‌های زبانی بر اساس نیازهای زبان‌آموزان فارغ از زبان اول و دوم

برابری در ارزشیابی (fairness in assessment) یکی از اهداف مهم در زمینه آموزش و سنجش (testing) زبان است، به طوری که در دهه‌های اخیر توجه دانشمندان را به خود جلب کرده است (مک نامارا ۲۰۰۵، کونان ۲۰۰۰ و ۲۰۰۴، خلی ۲۰۱۰). هر چند برخی این مباحث را چندان مهم نمی‌دانند (دیویس ۲۰۱۳). با این حال پژوهش برای تعریف و تعیین حدود برابری و نیز ارتباطش با روایی آزمون (test validity) اهمیت بسزا دارد. مک نامارا و رایان (۲۰۱۱) برابری را چنین تعریف می‌کنند: «تضمین میزان کیفیت آزمون، به ویژه کیفیت روان‌شناختی آن و برابری در رویه‌ها برای فرد و زیرگروه شرکت‌کنندگان در آزمون و کفایت بازنمایی ساختار در مواد و روش‌های آزمایش» (ص ۱۶۳). این تعریف سوای تعریف از عدالت (Justice) در آزمون است. شناسایی نقطه آغاز، نوع مواد آموزشی، کیفیت ارزشیابی و نیز اهداف آموزشی از جمله اطلاعاتی است که برای ایجاد برابری در ارزشیابی اهمیت دارد. تحقیقات درباره برابری در دو جهت نظری و عملی انجام پذیرفته است. هریک نیز به نوبه خود باعث تحول در دیدگاه افراد ذی‌نفع (stakeholders) می‌شود.

چهارچوب مرجع اتحادیه اروپا (چما) (Common European Frame of Reference) که با پشتوانه تحقیقاتی گسترده تدوین شده است، می‌تواند زمینه‌ای برای تحقق برابری ارزشیابی به دست دهد. این چهارچوب در سال ۲۰۰۱ معرفی (شورای اروپا، ۲۰۰۱) و توصیه شده تا به تدریج در نظام آموزشی کشورهای عضو اتحادیه به کار رود. علاوه بر این کشورها، پژوهش برای به کار بستن چما در سطوح مختلف در استرالیا (مک نامارا و رایان، ۲۰۱۱) چین (هوانگ و جیا، ۲۰۱۲) ژاپن (ناگایی و اودویر، ۲۰۱۱) تایوان (وو و وو، ۲۰۰۷)، ترکیه (پیچی، ۲۰۱۲)، کانادا (الاتیا، ۲۰۱۱) صورت پذیرفته است. همچنین هم‌راستاسازی (aligning) آزمون‌های معروف بین‌المللی مانند تافل (TOEFL) و آیلتس (IELTS) با طبقه-

تعریف شده است. هریک از این تعاریف در جهت یادگیری، آموزش و ارزیابی قابل کاربرد است. گروه‌های درگیر امر آموزش زبان عبارت‌اند از: سیاست‌گذاران (policy makers) تدوین‌کنندگان متون آموزشی (content developers) آموزگاران (teachers) و سایر ذی‌نفعان هریک می‌توانند از توصیف‌های به دست داده شده در چما استفاده مطلوب و منسجم ببرند. برای مثال تهیه مواد آموزشی به تعریف دقیق موقعیت زبانی (scenario) محتمل نیاز دارد. نمونه‌هایی از این موقعیت‌ها در اسناد منتشره چما وجود دارد. زبان‌آموزان می‌بایست به روشنی در مورد اهداف یادگیری (learning purposes) خود یاری شوند و بتوانند در هر مرحله توفیق یا عدم آن را در آموزش خودآزمایی (self-assessment) کنند. آزمون‌سازان (test makers) وقتی از اهداف و تعاریف موقعیت‌ها و سطوح متعدد یادگیری آگاهی داشته باشند می‌توانند آزمون‌های متناسب با اهداف مدنظر هریک از ذی‌نفعان تهیه کنند. به این ترتیب روایی تفسیر نمرات (validity of test score interpretation) بیشتر تضمین می‌شود (شورای اروپا، ۲۰۰۱ ص ۵).

توصیف‌گرهای چما:

این توصیف‌گرها برای شناخت عملکرد زبانی افراد در موقعیت‌های واقعی ارتباطی طراحی شده است. طبقه‌بندی سه‌گانه اولیه این توصیف‌گرها به ترتیب از بالا به پایین عبارت است از: کاربر مسلط (C1) (proficient user) و کاربر مستقل (C2) (independent user) (B1-B2) و کاربر مبتدی (A1-A2) (basic user). این سطوح مهارت زبانی با جمله‌های واضح در مهارت‌های درک مطلب خواندن (understanding reading) شنیداری (understanding listening comprehension) ، سخن گفتن (speaking) و ارتباط کلامی (spoken interaction) و همچنین نوشتن (writing) توصیف می‌شوند. نمونه‌ایی

از دو سطح توصیف شده A1 و B2 در جدول ۱ دیده می‌شود. این توصیف‌گرها به «جملات می‌توانم ...» (can do statements) نیز شناخته می‌شوند.

جدول ۱. توصیف‌گرهای A1 و B2 در مهارت‌های زبانی CEFR (شورای اروپا، ۲۰۰۱)*.

* ترجمه مولف

	درک		سخن گفتن	نوشتار
	شنیداری	خواندن		
A1	بتوانم کلمات آشنا و عبارتهای بسیار ابتدایی را درباره خودم، خانواده- ام و اطرافم را وقتی دیگران به زبان ساده و به آرامی صحبت می‌کنند بفهمم.	بتوانم اسامی آشنا، کلمات و جمله‌های خیلی ساده را مثل اعلامیه‌ها و پوسترها یا در کاتالوگ‌ها درج می‌شود بفهمم.	بتوانم به آسانی ارتباط برقرار کنم به شرطی که شخص دیگر حاضر باشد جمله یا بخش‌های آن را تکرار کند یا به عبارتی دیگر درآورد و به آهستگی بگوید. و در بیان آن چه را که می‌خواهم بگویم کمک کند. بتوانم درباره چیزهای ضروری و موضوعات آشنا سؤال کنم و جواب بدهم.	بتوانم مثلاً پشت یک کارت پستال را برای تبریک کوتاه و ساده بنویسم. بتوانم فرم- های ساده را با مشخصات فردی پر کنم، مثلاً نوشتن نام، ملیت، و نشانی در کارت ثبت نام هتل.
B2	بتوانم سخنان پیوسته و سخنرانی را دنبال کنم حتی اگر از نظر استدلالی پیچیده باشد در صورتی که موضوع به طور معقول آشنا باشد. بتوانم بیشتر برنامه‌های خبری و رویدادهای اخیر را بفهمم. بتوانم اکثر فیلم- هایی را بفهمم که به لهجه استاندارد است.	بتوانم مقاله‌ها و گزارش‌های مربوط به رویدادهای جاری را که نویسنده رویکرد و روش خاصی را در آن داشته بفهمم. بتوانم نوشته‌های نثر معاصر را بفهمم.	بتوانم ارتباط برقرار کنم تا حدی از شیوایی و هم- زمانی که ارتباط با گویشور این زبان امکان‌پذیر باشد. بتوانم به طور فعال در بحث- ها در بافت آشنا شرکت کنم و طی آن به استدلال برای اثبات دیدگاهم بپردازم.	بتوانم توصیف- های واضح و با جزئیات درباره گستره وسیعی مرتبط با موضوعات مورد علاقه‌ام بنویسم. بتوانم مقاله یا گزارشی برای ارائه اطلاعات یا دلایل در حمایت از یا در ضدیت با دیدگاهی خاص بنویسم. بتوانم نامه‌هایی بنویسم برداشتهای مهم شخصی را از رویدادها و تجربه‌ها مشخص کنم.

پژوهش‌های زبان‌شناختی در زبان‌های خارجی، دوره ۱۲، شماره ۱، بهار ۱۴۰۱، از صفحه ۱۱۱ تا ۱۳۱

این چهارچوب در ابتدا برای بزرگسالان فراهم شده است. به این سبب برای تهیه مطالب آموزشی و نیز آزمون‌های خاص کودکان و نوجوانان، توصیف‌گرها نیازمند بازنگری

محدودیت‌های چما:

است. به طوری که مطالعات نشان می‌دهد سطوح تعریف شده در چما فراتر از درک شناختی کودکان است. دیگر اینکه گذر از هر سطح به سطح دیگر مطابق روند رشد شناختی کودکان نیست. به منظور رفع این کاستی‌ها هدفهای زیر باید در نظر گرفته شود: تاحدامکان شناسه‌هایی را تعریف کند که نشان دهنده دنیا و زبان کودکان باشد، سطحی بالاتر از ب ۱ را تعریف نکند، و میان سطوحی (مثلاً ۱+، یا ۲+) را تعریف کند که کودکان پیشرفت واقعی را تجربه کنند. از جمله ایراداتی که بر چما وارد می‌دانند (آلدسون ۲۰۰۷) کمبود شواهد تجربی در روایی به کاربردن آن است.

۲. پیشینه تحقیق:

. هم‌راستاسازی چما:

هم‌راستاسازی درونی

هم‌راستاسازی چما با سامانه‌های آموزشی و تدریسی موجود یکی از روش‌های مؤثر در کاربردن و پژوهش درباره چما است. این پژوهش‌ها در سامانه‌های آموزش و ارزشیابی زبان دوم در هر کشور با پیشینه تاریخی و فرهنگی و آموزشی متفاوت انجام یافته است. این فرآیند خود نشان دهنده تلاش برای نشان دادن روایی کاربرد چما در محیط‌های گوناگون است. زیرا، ضرورت پوشش مسائل ناشی از پدیده چندزبانی و چندفرهنگی در زمینه‌های آموزش، یادگیری و ارزشیابی از ابتدای پیدایش و طراحی چما مطرح بوده است. نگیشی و همکاران (۲۰۱۲) در تحقیقی به روش سؤال پاسخ به تحلیل نتایج به دست آمده از آزمون واژگان در گروه کثیری از شرکت‌کنندگان ژاپنی پرداختند. یافته‌ها نشان داد که ترتیب ارائه افعال ترکیبی لزوماً مطابق پیشنهادهای مندرج در توصیفگرهای چما نیست. باین حال اذعان کردند شاید ماهیت زبان ژاپنی باعث این وضعیت است و پیشنهاد کردند تا همین تحقیق در زبان‌های دیگر نیز صورت پذیرد. در پژوهشی دیگر در مورد استفاده از چما در طراحی برنامه درسی مبتنی بر نیازهای ارتباطی زبان‌آموزان

در سطح دانشگاهی ارسال و اوزنیچی (۲۰۱۷) پس از ذکر فایده‌ها، پیشنهاد می‌کنند در طراحی برنامه‌های آموزش زبان این چهارچوب به کار برود. از جمله این پژوهش‌ها تدریس زبان انگلیسی عمومی به دانشجویان غیر رشته زبان در ویتنام است. گزارش شده است که هم‌راستاسازی چما در سامانه آموزشی باعث تغییر کیفی در خودارزیابی دانشجویان، آزمون‌های پایان نیمسال تحصیلی و نیز سنجش متقابل دانشجویان شده است (لی اچ.ت.ت.، ۲۰۱۸).

. هم‌راستاسازی بیرونی

از چما برای تهیه آزمون‌های پایه زبان‌های عمده و غیراروپایی نظیر عربی استفاده شده است. در این حیطه از توصیفگرهای مهارت‌های عملی زبان برای تعریف سطح آشنایی و تسلط زبان استفاده می‌شود و بر اساس این تعاریف آزمون سنجش مهارت در آن زبان به دست می‌آید. برای مثال، علی (۲۰۱۸) در گزارشی درباره بخشی از پژوهش هم‌راستاسازی آزمون مهارت عملی زبان عربی (TALP: Test of Arabic Language Proficiency) بیان می‌دارد: «برای تطبیق با اقتضات زبان عربی بسیاری از نکات ویراسته شده‌اند بنابراین چهارچوب توصیفی این تحقیق حاوی نکات زیر است:

- دشواری لغات در هر سطح و چگونگی گزینش و مرتب‌سازی
 - عناصر دستوری، طبقه‌بندی‌ها و بخش‌های ساختاری و نیز فرایندها و روابط
 - عناصر واژگانی و فرایندها
 - روابط معناشناختی
 - مهارت‌های آواشناختی و توزیع هر یک در سطوح مختلف
 - گروه‌های اجتماعی در جامعه مقصد که فراگیران می‌بایست آشنا شوند
- علاوه بر نکات فوق توصیف درباره مسائل ذیل نیز ارائه شده است:

- اصول انتخاب نقش‌های خرد و کلان و چینه‌بندی آنها
 - شناسایی حیطه‌ها و چرایی اهمیت آنها
 - تعیین فعالیت‌های زبانی که باید در آزمون و برنامه درسی گنجانده شود؟
 - تعیین تعداد ساعات تدریس متناسب با سطوح زبانی به میزان نیاز
- پژوهش‌های بینا فرهنگی چما**

پرسش مهم درباره کارایی و به‌خصوص روایی برابر چما در عرصه فرهنگ‌های متفاوت اهمیت بسزایی دارد. آیا چما که مرجعی کلان‌سطح در برنامه‌ریزی زبانی فراهنگی معرفی شده است، می‌تواند با روایی مناسب در سطح خرد برنامه‌ریزی روایی مطلوب به دست دهد؟ در پژوهشی بینا فرهنگی، محتوای کتاب‌های زبان انگلیسی دوره سوم و چهارم مدارس ابتدایی ترکیه و پرتغال با توجه به ارجاعات توصیفی چما در سطح 1A مقایسه شد. نتایج مقایسه نشان دهنده تفاوت و تشابه در اجرای چما در هر یک از این دو کشور بود. پیشنهادهای مبتنی بر نتایج تحقیق شامل آموزش آموزگاران در آشنایی با ویژگی‌های چما، تهیه مطالب چند فرهنگی متناسب با سن زبان‌آموزان در سطوح ابتدایی و نهایتاً توصیه بر پیوستگی و تداوم برنامه‌ریزی درسی در چهارچوب فرهنگی زبانی هر دو کشور ارائه شده است (گوئرا، ۲۰۱۸).

کاربرد چما در سطح‌بندی زبان اول

نمونه جالب در کاربرد چما برای طراحی آزمون زبان اول - و نه زبان دوم - است. در تحقیقی، نیازمندی‌های توصیفی کلی یادگیری سطح ابتدایی و ساختار و نیز ارزیابی و سنجش زبان تاتار در سطح 2A از سامانه توصیف‌گرهای چما استفاده شده است. از نظر پژوهشگران تعریف زبان تاتار در این سامانه انتقال روش‌های نوین آموزش زبان دوم را تسهیل می‌کند. در واقع چما پلی برای تبادل روش‌های آموزش، سنجش و ارزشیابی می‌شود. این پژوهشگران خواسته‌اند بدین وسیله از مرزهای داخلی روسیه فراتر روند و ارتباطی بین‌المللی برقرار کنند و زبان تاتاری را برای

دیگر فرهنگ‌ها دانستنی کنند. همچنین از این طریق نیازسنجی‌های زبان تاتار با زمینه‌ای پیشرفته هماهنگ بشود تا در نتیجه این ارتباط، سبب روزآمدی پیوسته روش‌های آموزشی در حیطه زبان تاتار بشود (شاکرووا، ۲۰۱۸).

ارزیابی زبانی شهروندی (Citizenship Language Assessment)

در کشورهای مهاجرپذیر و چندزبانی - سوئیس - و چند فرهنگی - ایالات متحده - دانش و توانش زبانی مهاجران و تطبیق با هنجارهای زبانی جامعه مقصد همواره مسئله‌ای بغرنج بوده است. در سوئیس، برای نمونه، فرایند پذیرش مهاجر باید به نحوی صورت پذیرد که درک فرهنگ و زبان در جامعه به اندازه انجام بشود. از این روی به منظور استانداردسازی همه فرایندها توصیف‌گرهای چما ترجمه شد تا با الزامات حداقلی و حداکثری قوانین استان-ها (contons) تطبیق داده شود (آریغی، ۲۰۱۸).

پیکره زبانی (language corpus) بر مبنای چما

انباشت متون حاصل از نگارش متن توسط زبان‌آموزان (زبان دوم) یکی از راهکارهای مناسب برای ساختن پیکره زبانی است. از این پیکره زبانی در تهیه مطالب آموزشی و آزمون‌های لازم می‌توان بهره وافر برد. به این منظور در سوئد که کشوری مهاجرپذیر است متن‌های نوشته شده توسط فراگیرندگان زبان سوئدی در سامانه‌های مختلف زبان‌آموزی براساس توصیف‌گرهای چما تهیه شد. هدف از تشکیل این پیکره زبانی بازتعریف انتقادی و ارزیابی روش‌های آموزش و سنجش سامانه فعلی و تجدید نظر برای تهیه مطالب و تدریس و ارزشیابی با موقعیت برابر برای زبان‌آموزان است (مگیسی و همکاران، ۲۰۱۸).

تهیه متون درسی (materials preparation)

در حیطه تدوین متون درسی نکات متعددی از جمله دانش زبان‌شناختی، جامعه‌شناختی و نیز کاربردشناسی زبان اهمیت بسزا دارد. برای مثال آگهی‌های تبلیغاتی، تنها یکی از منابع تهیه متون برای آشناسازی زبان‌آموزان با پیکره‌ای از

اطلاعات زبانی در زبان مقصد است. از چما برای سطح - بندی متن‌ها می‌شود استفاده کرد (پرزداکایه، ۲۰۱۸).

تعیین سطح گفتاری و شنیداری

به‌منظور تعیین سطح، چما در هر چهار مهارت زبانی کاربردهای مؤثری دارد. از جمله برای تعیین میزان پیچیدگی (complexity) صحت (accuracy) و روانی (fluency) گفتار از سطح‌بندی‌های چما استفاده می‌شود. در تحقیق کرمی و همکاران (۲۰۱۹) متون منتخب در گستره A1 تا C1 در نظر گرفته شد تا پوشش کاملی از توانش گفتاری شرکت‌کنندگان تعیین شود (کرمی و همکاران، ۲۰۱۹). در تحقیقی دیگر (دومنا و زفیری، ۲۰۱۸) برای پژوهش در میزان یادگیری زبان‌آموزان مدرسه خصوصی آزمون‌هایی در سطح A1 به‌عنوان پیش‌آزمون به شرکت‌کنندگان دادند. هدف از این تحقیق تعیین تفاوت تدریس متفاوت و متعارف در مدارس خصوصی بود. بدین منظور آزمون‌های مربوطه متناسب با سطوح مرتبط با چما به کار رفت زیرا بنا به گفته محققان برنامه آموزشی این مدارس بر اساس چما تهیه و ارائه شده است.

ارزشیابی پیشرفت

برای تشخیص میزان پیشرفت زبان‌آموزان در سطوح بالاتر، از آزمون‌های متعارف کمتر می‌توان استفاده کرد. زیرا این آزمون‌ها نشانگر دقایق زبانی (مانند، همایی (collocation) زبانی) در سطوح بالا (B2, C1) نیستند. برای تعیین این سطح از مهارت در یادگیری زبان فرانسه تعدادی همایهای اسم-فعل تهیه شد. در نتیجه این تحقیق معلوم شد کاربرد دقیق همایندها نشان دهنده سطح بالای دانش‌زایای زبانی است (لوندل و دیگران، ۲۰۱۸).

تعریف نظری روایی

در زمینه آموزش (و ارزشیابی آموزشی) موضوعات به هستی‌شناسی (ontology)، معرفت‌شناسی (epistemology) و اخلاق مربوط می‌شود. اینکه چه

چیز را حقیقت بدانیم، چگونه به حقیقت دست بیابیم، و عواقب (consequences) سنجش چیست (فولچر به نقل از هوندریچ، ۱۹۹۵، ص ۶۶۶)، تقریباً همه در حیطه سنجش و ارزشیابی آموزشی بر سر اصول اخلاق (حرفه‌ای، نیز برابری و عواقب سنجش در زندگی افراد) اتفاق نظر دارند. لیکن تفاوت آرا در زمینه چگونگی دستیابی به تعریف واحد و نیز چگونگی رسیدن به تعاریف عملیاتی حاصل از تعریف مفاهیم بیشتر مربوط به روش‌های هستی-شناختی و معرفت‌شناختی است. مسائل مطرح در کار شناخت بیشتر به دو دیدگاه ختم می‌شود. در یک دیدگاه «حقیقت» مستقل از ما وجود دارد. در دیدگاه دیگر حقیقت آن چیزی است که ما از حقیقت برداشت می‌کنیم. برای هر یک از این دیدگاهها استدلال‌های متعددی در طول تاریخ علم ارائه شده است که در اینجا ما به اینها نمی‌پردازیم. بلکه تأثیر هر یک از این دو دیدگاه را در حیطه سنجش و ارزشیابی آموزشی بیان می‌کنیم. از آنجا که بحث درباره ماهیت و چیستی و نیز چگونگی تأثیرگذاری نتایج آزمون و ارزشیابی به خارج از حیطه محاسبات آماری گسترش پیدا می‌کند، و از طرف دیگر برقراری ارتباط بین یافته‌های آماری و تفسیر با تفاسیر منتج از نمره آزمون در تعیین روایی آن تأثیر دارد، به روش استدلالی نیاز است که تبیین روشنی از داده‌های آماری و غیر آن حاصل از اجرای آزمون به دست دهد. در ادامه ضمن مرور کلی تعریف روایی به معرفی شیوه استدلالی درباره ارتباط نتایج آزمون و عواقب آن و نیز تفسیر آن پرداخته می‌شود.

ساده‌ترین تعریف رایج از روایی: «آزمون چیزی را بیازماید که برای آزمودنش به وجود آمده» است. این تعریف در خوانش اول بسیار ساده و قابل فهم به نظر می‌رسد. اما مسائل مربوطه در آموزش، سنجش و ارزشیابی چنان پیچیده هستند که این تعریف نمی‌تواند روشن‌نگر باشد. به این ترتیب دانشمندان سنجش و ارزشیابی در پی یافتن تعریفی هستند تا علاوه بر توضیح و تبیین روایی و تفسیر و کاربرد (test interpretation and use) نتایج آن بتوانند تأثیر نمره

را در زندگی فرد و عواقب پس از آن پژوهش کنند. مسیک (Messick) با ارائه نظریه «روایی یکپارچه» (validity as concept a unitary) تقسیم‌بندی‌های روایی را در یک الگوی واحد قرار داد. پیش از آن روایی متشکل از چند بخش تعریف می‌شد که هریک بخشی از واقعیت منتج از آزمون را پوشش می‌داد. در تعریفی دیگر که روایی را از اجزای اساسی آزمون‌ها می‌شناسد چنین آمده است: میزانی که تفسیر و کاربرد نتیجه آزمون بنا به مقصود کاربران با نظریه و شواهد تأیید می‌شود. (انجمن روان‌شناسی آمریکا، ۱۹۹۹؛ ص ۹ مذکور در کامینز، ۲۰۱۲).

روایی‌سازی بر اساس استدلال

این روش از رواسازی آزمون در پی ابداع روش استدلالی تولمین (۲۰۰۳) توسط کین (۱۹۹۲، ۲۰۰۶) برای رواسازی استدلالی آزمون‌های زبان بسط داده شد. این ایده از دو بخش تشکیل شده است اول بیان استدلال تفسیری (interpretive argument) و سپس تدوین و ارزیابی شواهد استدلال و نیز در نظر گرفتن موارد بالقوه نقض وابطال (potential counterevidence) استدلال. در این راستا بکمن (۲۰۰۵) بر لزوم و تحقیق و کاربرد شیوه استدلالی برای رواسازی آزمون تأکید کرد. سپس صورت‌بندی روابط مفهومی بین تفاسیر مترتب بر آزمون، اتخاذ تصمیم بر اساس نتایج و در نظر گرفتن عواقب آن و همچنین توجه به انصاف در ارزشیابی را بکمن و پالمر (۲۰۱۰) در قالب رویکرد استدلال کاربردی ارزشیابی (assessment use argument) قرار دادند. کونان (۲۰۰۴) در مفهوم‌سازی برابری دو چهارچوب تکمیل‌کننده یکی چهارچوب برابری آزمون و دیگری چهارچوب بافت اجتماعی ارائه کرده است. چهارچوب آزمون شامل روایی، نبود سویه‌گیری، دسترسی، اجرای آزمون و عواقب اجتماعی است و هریک از این کیفیات در تدوین و کاربرد آزمون اهمیت بسزا دارد. چهارچوب دیگر ناظر بر عوامل محیطی تأثیر گذار بر برابری آزمون است (ص ۲۷). از آنجا که تعریف و محاسبه تأثیر یکایک عوامل مذکور به‌سادگی

انجام‌پذیر نیست حداقل این عوامل عمده مؤثر در - تقریباً - همه محیط‌ها را در هنگام تفسیر نتایج آزمون باید در نظر داشت. این عوامل عمده عبارت‌اند از: سیاسی، اقتصادی، آموزشی، اجتماعی، فرهنگی و قانونی و اخلاقی. (کونان، ۲۰۰۸، ص ۲۴۰).

برابری در ارزشیابی

برابری در آزمون به عواقب نتایج آزمون برای افراد و گروه‌ها یا جامعه مربوط می‌شود. این مفهوم هم به روایی و هم به نوعی شاخص سنجش توانایی زبانی و در نتیجه به تساوی اجتماعی مربوط می‌شود (دی ویس و دیگران، ۱۹۹۹). بر همین منوال می‌توان گفت تحقیق درباره برابری به از طرفی به آزمون و نیز به کاربرد نتیجه آن در بافت اجتماعی ارتباط مستقیم دارد. از طرف دیگر می‌توان عدم وجود سویه در آزمون، در رویکرد روان‌سنجی، مفهوم یا کیفیت برابری را تأمین می‌کند. پرسش از وجود این کیفیت در آزمون در واقع وقتی مطرح می‌شود که گروه‌های مختلفی از آزمون‌شوندگان در فرایند سنجش قرار می‌گیرند. آیا آزمون تنها مهارت زبانی را می‌سنجد و برای هیچ‌یک از شرکت‌کنندگان برتری ایجاد نمی‌کند؟ برای پاسخ به این سؤال روش‌های متعدد آماری به کار رفته و نیز گروه‌بندی‌های مختلف از جمله زبانی، جنسیتی، و فرهنگی بیشتر مورد پژوهش قرار گرفته است. این پژوهش از ابتدای فرایند تهیه آیت‌ها تا اجرا و شیوه نمره‌گذاری آغاز می‌شود و حتی در نوع تفسیر نتایج و عواقب اجتماعی ادامه می‌یابد. تهیه‌کنندگان و کاربران آزمون با استفاده از روش‌های آماری متناسب در پی کاستن از اثرات نابرابر بر آزمون هستند تا آزمون را تا سرحد امکان برای حصول مقصود بهینه کنند (کونان، ۲۰۰۴؛ مک نامارا و روور، ۲۰۰۶؛ بکمن و پالمر، ۲۰۱۰). روند بررسی اثر برابری در آزمون در آثار محققانی مانند مسیک (۱۹۸۹)، کین (۲۰۰۰)، کونان (۲۰۰۴) و شی (۲۰۱۰) دیده می‌شود.

این خط روند پژوهشی نشان می‌دهد که برابری در آزمون به انصاف، روایی و عدم سویه‌گیری مرتبط است.

در همین راستا از نظر خلی (۲۰۱۰) برابری همان روایی است که برای گروهها یکسان تعریف شده و در تهیه و اجرا و تفسیر نتیجه آزمون اعمال می‌شود. در راستای پژوهش درباره رابطه بین برابری و روایی، از نظر کونان (۲۰۰۰) وایی از طریق تعریف روایی سازه متساوی برای تفسیر نتیجه آزمون برای همه گروههای آزمون‌شوندگان است. این مهم از نظر او با کنترل سوبه‌گیری محتوا، شکل ارائه و کنش آیتم (item functioning) و نیز زبان ارائه آزمون‌ها امکان‌پذیر است.

روایی و برابری در منظر کین (۲۰۱۰) بسیار به یکدیگر نزدیک‌اند و میزان فراپوشش و فروپوشش هریک به تعریف آنها در موقعیت‌های مختلف متفاوت است. باین‌حال به نظر او در حالت کلی هیچ‌یک از این دو آن دیگری را در بر نمی‌گیرد و هریک هویت مستقلی دارد.

مک نامارا و روور (۲۰۰۶) بدون ارائه تعریف نظری مشخص و ارائه راهکار معین ادعا می‌کنند مسائل مربوط به برابری و نقش اجتماعی آزمون به‌قدر کفایت مدون نشده است. ایشان به اشاره بر ضرورت برابری اجتماعی و جلوگیری از سوبه‌گیری آزمون‌ها به نفع یا ضرر گروههای اجتماعی تأکید می‌کنند. باین‌حال پیشنهاد می‌کنند پژوهش در تفاوت‌های فردی آزمون‌شوندگان و تأثیر آن بر عملکرد ایشان در آزمون و در نظر داشتن معیارهای اخلاق حرفه‌ایی از طریق کنش افتراقی آیتم (کآ) (differential item functioning) امکان‌پذیر می‌باشد.

ضرورت انجام تحقیق:

چما به تدریج نقش اساسی در مسائل مرتبط به آموزش و سنجش زبان در میان کشورهای اروپایی پیدا می‌کند (از جمله: نگیشی و دیگران ۲۰۱۲). تحقیقات متعددی در شاخه‌های مختلف اعم از آموزش، سنجش، تهیه و تدوین مطالب آموزشی، تربیت آموزگار و غیره بر اساس چما انجام می‌پذیرد. این تحقیقات تنها محدود به کشورهای اروپایی نیست و در دیگر نقاط جهان هم در جریان است.

این خود نشان دهنده اهمیت نقش آفرینی چما در آینده آموزش زبان - و امور مرتبط - است. از این روی شناخت این چهارچوب و تحقیق در چگونگی به کار بستن و یا حتی انتقاد از دست آوردهای آن امری اجتناب‌ناپذیر می‌نماید.

ایجاد ارتباط بین آزمون مبتنی بر چما از یک‌سو نیازمند طراحی بنا به مقصود است و سپس اجرا و محاسبه داده‌ها در چهارچوب آماری معین. از طرف دیگر مرتبط کردن نتایج حاصله از اجرای آزمون با عوامل محیطی و مقاصد شرکت‌کنندگان در آزمون و در نهایت تفسیر نتایج زمینه را برای توضیح و توجیه روایی آزمون مهیا می‌کند. در انتها تخمین (extrapolation) نتایج آزمون و تفسیر یا تفاسیر مرتبط با آن انجام می‌شود.

تطبیق نمره حاصل از آزمون با استانداردهای رایج (از جمله چما و ACTFL) معمولاً با واژگان خاصی بیان می‌شود. مفهوم منتسب به این جملات هریک مستقیماً به نوع تعریف سطوح زبان‌آموزی و برداشت افراد ذی‌نفع ممکن است از محیطی به محیط دیگر متفاوت باشد. علاوه - براین افراد با مقاصد مختلف نیاز دارند درک روشنی از مفهوم نتیجه آزمون و ارتباط آن با مقصود شخصی خود داشته باشند. فراهم آوردن شواهد برای ایجاد ارتباط بین نتایج آزمون و تفاسیر مرتبط با آن بخش مهمی از استدلال روایی نتایج است (بکمن، ۲۰۰۳). بنابراین، تعیین چهارچوب فرایندی اساسی در استدلال روایی نتیجه آزمون است که بنا به مقصود مشخصی با توصیف‌گرهای زبانی به وجود آمده است. در این خصوص چند مسئله باید مد نظر قرار گیرد. چه کسانی در فرایند روایی آزمون ذی‌نفع هستند و معنی نتایج برای هر گروه از ذی‌نفعان باید با زبان روشن و با توجه به اهداف و مقاصد هر گروه بیان شود. دیگر اینکه فرایند روایی آزمون در یک مرحله به پایان نمی‌رسد، بلکه نتایج با توجه به تغییر شرایط و اشخاص ذی‌نفع و اهداف بیان شده تفسیر متفاوتی را پدید می‌آورد.

نتایج حاصل از آزمون اطلاعاتی به تدوین‌کنندگان و کاربران آزمون می‌دهد که به‌عنوان شاهد برای استدلال روایی استفاده کنند که به‌نوبه خود در ایجاد ارتباط بین نتایج و تفسیر و تصمیم‌های منطبق بر آن تأثیر می‌گذارد (کین، ۱۹۹۲؛ خی، ۲۰۰۸). انتخاب نوع تحلیل آماری و تفسیر نتایج بستگی به این دارد که آیا آزمون nrt است یا crt. در این تحقیق آزمون از نوع crt فرض شده است زیرا بر اساس چهارچوب‌های چما تهیه شده است. یکی از ویژگی‌های تحقیق حاضر این است که ارتباطی بین روش استدلال روایی و کاربرد آزمون آن با چهارچوب‌های ارائه شده در چما در تعیین سطح آزمون‌شوندگان بر اساس علایق فرهنگی آنها برقرار می‌کند. به سخن دیگر، شرایط ذهنی آزمون‌شوندگان از طریق پرسش‌نامه تعیین شده در تفسیر نتایج دخیل است. به‌این‌ترتیب اطلاعات حاصل از تجزیه و تحلیلی آماری به تصمیم‌گیرندگان و دیگر افراد ذی‌نفع اطلاعاتی ارائه می‌دهد که تا سرحد امکان بر میزان روایی تفسیر و عواقب آن در زندگی آزمون‌شوندگان تأثیر بهینه دارد.

۳. روش تحقیق:

تعریف عملیاتی: در تعریف عملیاتی روش‌های گوناگون آماری و محاسباتی به کاررفته است. یکی از این روش‌ها کنش افتراقی آیتم است. برای محاسبه کلاً شرکت‌کننده‌ها را معمولاً براساس معیار یا معیارهایی به دو گروه هدف و مرجع تقسیم می‌کنند. این تقسیم‌بندی بر اساس مسائل مبتلابه جامعه و اولویت‌های پژوهشی متفاوت است. در جوامعی مانند ایالات متحده مسئله پیشینه نژادی افراد (اسپانیایی زبان و دیگر زبانهای اقلیت) در پژوهش مدنظر قرار می‌گیرد. همچنین مسائلی دیگر از جمله جنسیت در بیشتر جوامع مسئله‌ایی بغرنج و شایسته پژوهش می‌باشد. در اینجا باید توجه داشت این تقسیم‌بندی‌ها نه بر اساس روند تحقیقات انجام شده قبلی بلکه باید بر اساس نظرگاه شناختی و چپستی آنها طبق تعریف و نیاز آنی (immediate need) انجام شود. مثلاً می‌توان گروه‌های

افتراقی را بر اساس انگیزه و هدف یا موقعیت و دسترسی به عوامل آموزشی معین کرد. روش‌های آماری به کار برده شده هریک به نوعی برای موقعیت و تعاریف متفاوت متناسب هستند.

در تحقیق حاضر پاسخ شرکت‌کنندگان به سؤالات آزمون عملکرد مهارتی زبان انگلیسی که بر مبنای توصیفگرهای چما ساخته شده است تصحیح و دسته‌بندی شد. آیتم‌های سازنده آزمون از نظر سازه مهارت زبانی شناسایی و دسته‌بندی شد. سپس با روش نظریه سؤال پاسخ نِسپ (item response theory) نتایج به‌دست‌آمده تحلیل شد.

جامعه آماری

جامعه آماری این تحقیق ۱۵۰ دانشجوی زن و مرد - در رشته‌های مهندسی نفت و پتروشیمی، مکانیک سیالات و جامدات، کامپیوتر و اطلاع رسانی، برق و شیمی است. آنها با شرکت در آزمون ورودی به این رشته‌ها پذیرفته شده‌اند. طیف سنی آنها بین ۱۸ تا ۲۳ سال متغیر است. اکثر ایشان در درس زبان انگلیسی عمومی در یک نیمسال تحصیلی در این تحقیق شرکت داشته‌اند.

ابزار تحقیق

در این پژوهش از یک آزمون مهارت زبانی انگلیسی با ۲۵ آیتم برای تعیین سطح مهارت زبان انگلیسی شرکت‌کنندگان استفاده شد. انتخاب آیتم‌ها بر اساس توصیفگرهای زبانی و مهارتی چما است. سؤالات از آسان به سخت چیده شده‌اند. جواب منفی برای هریک از پاسخ‌ها در نظر گرفته نشد. در یک طبقه‌بندی سازه از پنج صاحب‌نظر خواسته شد که سازه‌های هریک از آیتم‌ها را معین کنند. در جدول ۴،۱ خلاصه نظر افراد صاحب‌نظر آمده است:

جدول ۲ خلاصه نظر صاحب‌نظران (۱) درباره سازه‌های آیتم‌ها (۲). همبندی و واژگان (۳). دستور (۴). کاربردشناسی

۱	صاحب نظر ۱	صاحب نظر ۲	صاحب نظر ۳	صاحب نظر ۴
۲	۶۰۹ و ۱۰۱ و ۱۱ و ۱۲ و ۱۴ و ۱۵ و ۱۷ و ۱۸ و ۱۹ و ۲۰ و ۲۱ و ۲۲ و ۲۳ و ۲۴ و ۲۵	۶۰۹ و ۱۰۱ و ۱۱ و ۱۲ و ۱۴ و ۱۵ و ۱۷ و ۱۸ و ۱۹ و ۲۰ و ۲۱ و ۲۲ و ۲۳ و ۲۴ و ۲۵	۶۰۹ و ۱۰۱ و ۱۱ و ۱۲ و ۱۴ و ۱۵ و ۱۷ و ۱۸ و ۱۹ و ۲۰ و ۲۱ و ۲۲ و ۲۳ و ۲۴ و ۲۵	۳ و ۶ و ۹ و ۱۰ و ۱۱ و ۱۲ و ۱۴ و ۱۵ و ۱۷ و ۱۸ و ۱۹ و ۲۰ و ۲۱ و ۲۲ و ۲۳ و ۲۴ و ۲۵
۳	۲ و ۳ و ۴ و ۷ و ۱۳ و ۱۶ و ۲۰ و ۲۱ و ۲۷	۷ و ۱۳ و ۱۴ و ۱۵ و ۱۶ و ۲۰ و ۲۱ و ۲۷	۲ و ۳ و ۷ و ۱۳ و ۱۶ و ۲۰ و ۲۱ و ۲۷	۷ و ۱۳ و ۱۴ و ۱۵ و ۱۶ و ۲۰ و ۲۷
۴	۴ و ۵	۲ و ۳ و ۴ و ۵	۴ و ۵	۲ و ۳ و ۴ و ۵

نتایج تحلیل آماری نسپ

برای آزمودن ارتباط بین سازه‌های زیرساختی (underlying constructs) آیت‌های آزمون و آیت‌های گروه‌های پرسش‌نامه همبستگی پیرسون به کار برده شد. آیت‌ها چنان‌که در جدول ۴،۲ مشاهده می‌شود رفتار متفاوتی نشان دادند.

جدول ۳ رابطه بین بخش‌های آزمون چما به‌عنوان نشان دهنده استدلال کاربرد ارزشیابی (تعداد=۱۵۰)

علاوه بر ابزار فوق، از نسخه فارسی پرسش‌نامه ۷۵ آیت (دورنیه و تاگوچی، ۲۰۰۹) استفاده شد تا نظرات شرکت‌کنندگان در آزمون درباره «موقعیت تحصیلی درآینده، موقعیت شغلی در آینده، زبان‌آموزی از نظر دیگران، نظر اعضای خانواده درباره زبان‌آموزی و دیدگاه درباره فرهنگ جامعه زبانی مقصد پرسیده شود. سپس این چند دسته‌بندی برای انجام محاسبات سازه‌ای و دیگر تجزیه تحلیل‌های لازم به کار بسته شد.

توانش منظورشناختی	دستور زبان	واژگان و هماینها	آر	موقعیت شغلی آینده
-۰/۰۱۲	۰/۱۱۱	۰/۰۵۳	آر	موقعیت شغلی آینده
۰/۸۸۷	۰/۱۷۸	۰/۵۱۹	سیگ	موقعیت شغلی آینده
-۰/۰۶۲	۰/۰۱۸	-/۰۰۶	آر	موقعیت تحصیلی آینده
۰/۴۵۰	۰/۸۲۷	۰/۹۴۵	سیگ	موقعیت تحصیلی آینده
-۰/۰۲۰	۰/۰۲۵	-/۰۰۶۵	آر	نگرش فرهنگی
۰/۸۰۸	۰/۷۶۲	۰/۴۲۶	سیگ	نگرش فرهنگی

مطالعه انجام شده توسط نعمت زاده (۲۰۱۸) نشان می‌دهد که مشخصه‌های فرهنگی و پیش‌زمینه‌های خاص مانند سن، جنسیت، و زبان مادری تفاوتی در آزمون نشان نداد. در تحقیقی با استفاده از کارکرد افتراقی آیتم (کآ، DIF) درباره جنسیت در آزمون توانایی زبانی سطح بالا (NUEEFL)، بردبار (۲۰۲۰) نتیجه می‌گیرد که نمرات حاصل از آزمون بدون واریانس بی‌ارتباط با سازه نیستند و برابری کلی آزمون تأیید نمی‌شود.»

تأثیر نمرات آزمون چما بر نشانه‌های استدلال کاربرد ارزشیابی

همه آیتم‌ها در سه گروه کلی واژگان، دستور زبان و منظورشناختی دسته‌بندی شدند. سپس، برای آزمودن ارتباط بین این سه گروه از آیتم‌های مبتنی بر چما و استدلال

سنجش کاربرد رگرسیون چندگانه خطی انجام شد. نتایج به‌دست‌آمده در جدول نشان می‌دهد که آیتم‌های مربوط به دستور زبان نسبت به دو دسته دیگر آیتم‌ها قدرت پیش‌بینی بیشتری دارند. یعنی، آیتم‌های ۱، ۳، ۷، ۱۳، ۱۶، ۲۰ و ۲۱ اطلاعات بیشتری را درخصوص ارزیابی توانایی زبانی شرکت‌کنندگان در آزمون فراهم می‌آورند. نکته جالب اینکه همین آیتم‌ها در تحلیل کنش افتراقی آیتم‌ها هم از اهمیت ویژه‌ای برخوردارند که بعداً ارائه خواهد شد. نکته قابل توجه دیگر اینکه آیتم‌های ۱ و ۴ توانش منظورشناختی در مقایسه با دو بخش دیگر دستور زبان و واژگان و هماینها قدرت پیشگویی بیشتری برای دسته آیتم‌های نگرش فرهنگی پرسش‌نامه از خود نشان دادند. این اطلاعات در جداول زیر خلاصه شده است.

جدول ۴						
مدل	ب	خطای استاندارد	بتا	تی	سیگ	
۱	۷۲/۲۰۰	۸/۲۲۵		۸/۷۷۹	۰/۰۰۰	(ثابت)
	۰/۰۰۳	۰/۱۲۱	۰/۰۰۲	۰/۰۲۲	۰/۹۸۳	واژگان هماینها
	۰/۲۰۰	۰/۱۶۶	۰/۱۲۰	۱/۲۰۱	۰/۲۳۲	دستور
	-۰/۰۷۴	۰/۱۵۴	-۰/۰۴۴	-۰/۴۸۳	۰/۶۳۰	توانش منظورشناختی
متغیر وابسته: موقعیت شغلی آینده						

موقعیت شغلی آینده = $۷۲/۲۰ + ۰/۰۰۳$ (واژگان و هماینها) - $۰/۰۷۴$ توانش منظورشناختی

$۰/۲۰۰ +$ دستور زبان

جدول ۵						
مدل		ب	خطای استاندارد	بتا	تی	سیگ
۱	(ثابت)	۶۹/۹۶۴	۷/۱۱۰		۹/۸۴۱	۰/۰۰۰
	واژگان هماینها	۰/۰۰۵	۰/۱۰۵	۰/۰۰۶	۰/۰۵۲	۰/۹۵۹
	دستور	۰/۰۴۸	۰/۱۴۴	۰/۰۳۴	۰/۳۳۴	۰/۷۳۹
	توانش منظورشناختی	-۰/۱۰۷	۰/۱۳۳	-۰/۰۷۳	-۰/۸۰۴	۰/۴۲۲
متغیر وابسته: موقعیت تحصیلی آینده						

موقعیت تحصیلی آینده = $۶۹/۹۶ + ۰/۰۰۵$ (واژگان و $- ۰/۱۰۷$ (توانش هماینها)
 منظورشناختی)

$+ ۰/۰۴۸$ (دستور زبان)

جدول ۶						
مدل		ب	خطای استاندارد	بتا	تی	سیگ
۱	(ثابت)	۴۷/۸۳۴	۷/۸۱۹		۶/۱۱۸	۰/۰۰۰
	واژگان هماینها	-۰/۱۳۱	۰/۱۱۵	-۰/۱۲۱	-۱/۱۳۷	۰/۲۵۷
	دستور	۰/۱۴۶	۰/۱۵۸	۰/۰۹۳	۰/۹۲۳	۰/۳۵۷
	توانش منظورشناختی	۰/۰۱۱	۰/۱۴۶	۰/۰۰۷	۰/۰۷۴	۰/۹۴۱
متغیر وابسته: نگرش فرهنگی						

نگرش فرهنگی = $۴۷/۸۳ - ۰/۱۳۱$ (واژگان و هماینها)

$۰/۱۴۶$ (دستور زبان)

$۰/۰۱۱$ توانش منظورشناختی

تحلیل کاآ آزمون چما بر اساس نشانه‌های استدلال کاربرد ارزشیابی

آمار توصیفی زیر گروه‌ها (تعداد-۱۵۰)		
انحراف معیار استاندارد	میانگین	
15.12	79.58	فرصت شغلی آینده
13.01	68.33	فرصت تحصیلی آینده
14.35	48.45	نگرش فرهنگی

برای تخمین میزان سویه‌گیری آزمون مبتنی بر چما به نسبت نشانه‌های استدلال کاربرد ارزشیابی از تحلیل کاآ استفاده شد. بیشتر مطالعات کاآ معطوف به یافتن آیتم‌هایی بوده است که سویه‌گیری جنسیتی، نژادی، و غیره داشته‌اند. ایده طبقه‌بندی در تحقیق حاضر بر مبنای ویژگی‌های طبیعی شرکت‌کنندگان نیست. زیرا ایشان در انتخاب این ویژگی‌های دخالتی ندارند و صرفاً اکتساب شده است. حال آنکه ویژگی‌هایی در طول زندگی برای ایشان به وجود می‌آید، هر چند تحت تأثیر ویژگی‌های طبیعی است ولی خود افراد سهمی در انتخاب و تحول و تکوین این ویژگی‌های ثانویه دارند. به این ترتیب می‌توان گفت این ویژگی‌های به دست آمده بر اساس انتخاب و تصمیم در روش زندگی آنها و تصورات و مقاصد ایشان تأثیرگذار است. علاوه بر این می‌توان گفت برای بررسی اثر آزمون بر عملکرد شرکت‌کنندگان و عواقب نتایج تفسیر حاصل از نمرات آزمون ارتباط مستقیمی با روایی آزمون دارد. زیرا زندگی شغلی و تحصیلی ایشان و نیز نگرش فرهنگی آنان تحت تأثیر ویژگی‌های ثانویه حاصل انتخاب آنها است. همچنین، گروه‌بندی‌های شغلی در عصر حاضر می‌تواند بر اساس علائق و انتخاب‌های افراد صورت پذیرد. این مسئله جنسیت و زبان و دیگر ویژگی‌های این گونه را در جامعه کشوری ما کم‌رنگ می‌کند. بر این مبنای می‌توان گفت آن دسته از ویژگی‌های طبیعی که در محیط چندزبانی و چندفرهنگی مثل ایالات آمریکا مسئله است برای ما اهمیت چندانی ندارد. بلکه، بخاطر یکسان بودن زبان مادری و نژاد افراد این گونه مسائل را نباید در مطالعات برابری یا عدم سویه‌گیری دخالت داد. جدول ۴،۵ اطلاعات مربوط به عملکرد شرکت‌کنندگان و پرسش‌نامه را نشان می‌دهد.

جدول ۷

از آنجایی که انحراف خطای استاندارد همه آیتم‌ها در هر دو گروه بالای ۰/۰۰ است، مقداری کاآ در داده‌ها مشاهده می‌شود. سنجه کاآ میزان سختی یک آیتم برای یک گروه است، وقتی همه شرایط تأثیرگذار دیگر یکسان باشد. کنتراست کاآ «میزان تأثیر» بر مبنای لاجبیت است و تفاوت بین دو سنجه کاآ یعنی اندازه کاآ در دو گروه طبقه‌بندی شده. کنتراست مثبت کاآ نشانگر سختی بیشتر آیتم برای گروه در حال مقایسه است. بنابر تعریف کاآ یک گروه از شرکت‌کنندگان نمره بیشتری از دیگر گروه دریافت می‌کند. دلیل این تفاوت بنابر ملاحظه لیناکر (۲۰۱۹) می‌تواند به خاطر موارد زیر باشد:

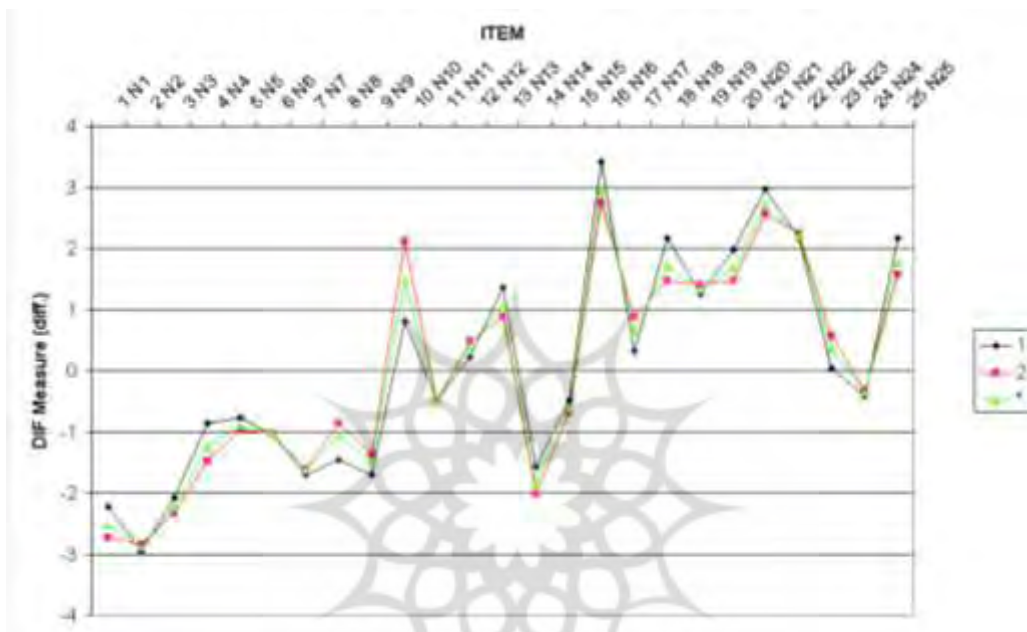
۱. گروه در حد طبیعی توانایی خود بر آیتم عمل می‌کند، گروه دیگر بهتر از حالت عادی خود عمل کرده است.
۲. گروه در حد طبیعی توانایی خود بر آیتم عمل می‌کند، گروه دیگر بدتر از حالت عادی خود عمل کرده است.
۳. آیتم سختی عادی خود را برای گروهی دارد، اما برای گروه دیگر بیشتر از حد طبیعی سخت است.
۴. آیتم سختی معمولی خود را برای یک گروه دارد، اما برای گروه دیگر ساده‌تر از حالت عادی است.

توجه به این نکته مهم است که اطلاعات آماری بین این توجیحات تفاوتی ایجاد نمی‌کند، بلکه ارتباط نتایج تحلیل کاآ با تصمیم‌های لازم بر مبنای مقاصد معین است که

تفاوت‌ها را معنادار می‌کند. پس در صورتی که آیت‌ها به نفع یک گروه و هم‌زمان به ضرر گروه دیگر باشد می‌گوییم آزمون متشکل از این آیت‌ها نابرابر (unfair) است. از ابتدای تعیین هدف برای تهیه آزمون و حین نگارش آیت‌ها و سپس در اجرای اولیه و اجرای اصلی و نیز در هنگام نمره‌گذاری و تفسیر نمرات می‌بایست داده‌های آماری را با اطلاعات کیفی کنار هم و در راستای دستیابی به برابری در شرایط آزمون بررسی کرد. باین‌حال موارد زیر از نظر لینا کر (۲۰۱۹) از لحاظ آماری بر نتایج کا تأثیرگذار است:

۱. اندازه تأثیر کا
۲. اندازه طبقه‌بندی گروه

نتایج تحلیل کا در جدول‌ها و نمودارها نشان دهنده اطلاعات در مورد تک‌تک آیت‌هایی است که مقدار قابل توجهی سویه‌گیری گروهی داشته‌اند.



شکل ۱ توضیح سنجه کا فرصت شغلی آینده

زیادی در هر دو گروه ۱ و ۲ دارد. آیت ۱۵ سه میزان بالاتر از خط مرجع (۰) قرار گرفته و برای هر دو گروه خاصیت یکسانی از خود نشان می‌دهد. نکته جالب توجه دیگر اینکه آیت ۲۴ و آیت ۱۰ در یک سطح قرار دارند. با این همه، روند فراز و فرود آیت‌ها در مجموع نشان می‌دهد شرکت - کنندگان روالی صعودی در اندرکنش با آیت‌ها احساس می‌کنند.

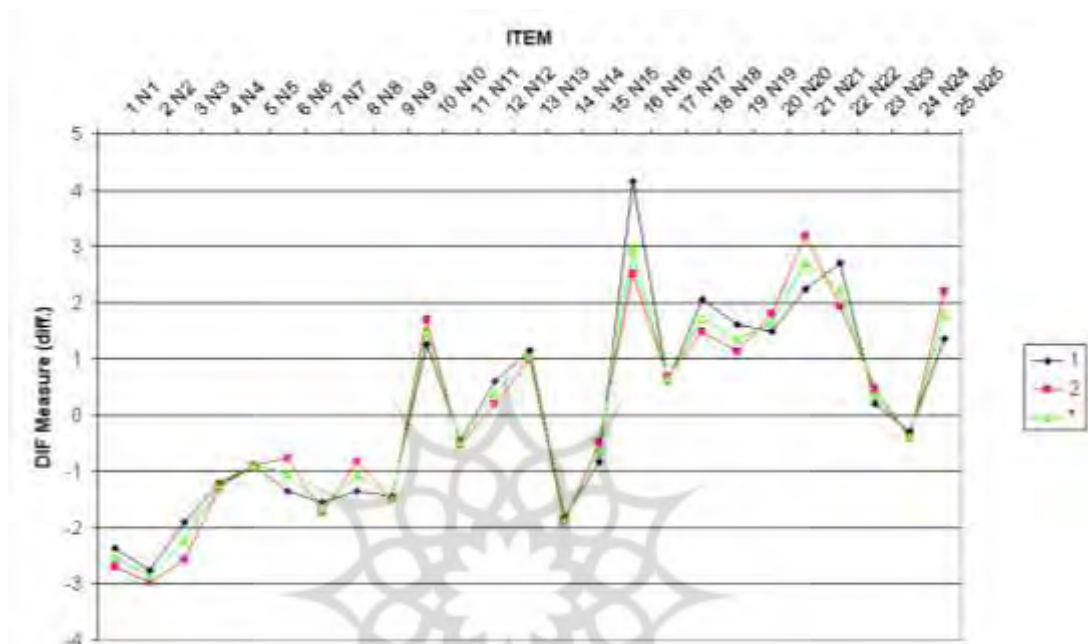
چنان‌که در شکل ۱ مشاهده می‌شود آیت ۱ تا ۸ تقریباً برای هر دو گروه پاسخ دهنده کا یکسانی دارند. اما، آیت‌های ۹ و ۱۰ تا حدودی نسبت به آیت‌های قبل کا بیشتری دارند و اندک تفاوتی در گروه ۱ و ۲ دیده می‌شود. باین‌حال، آیت - های ۱۱ و ۱۲ از نظر تفکیک مانند آیت‌های قبل نیستند و در سطح پایین‌تری در دو گروه ۱ و ۲ عملکردی تقریباً مشابه دارند. آیت ۱۴، اما توجه برانگیز است قدرت تفکیکی آن نسبت به آیت‌های قبل از خود فروکاسته شده و تشابه

جدول ۸ کا برای گروه فرصت شغلی آینده

Item Number	Name	Person Class	DIF Measure	Person Class	DIF Measure	DIF Contrast	Rasch-Welch		
							t	df	prob.
4	N4	1	-0.86	2	-1.48	.62	1.57	119	.1183
8	N8	1	-1.46	2	-.86	-.61	-1.50	106	.1356

10	N10	1	.80	2	2.12	-1.32	-2.74	144	.0070
16	N16	1	3.42	2	2.75	.67	.78	95	.4372
18	N18	1	2.17	2	1.48	.68	1.26	101	.2106

آیتم‌های ۴، ۶ و ۱۸ به نفع گروه دو و آیتم‌های ۸، ۱۰ به نفع گروه یک ساده است



۱ تفکیک بیشتری نشان می‌دهد. در نظرگاه کلی به نمودار و اندازه تفکیک کاآ و نیز نظر شرکت‌کنندگان همسانی زیادی دیده نمی‌شود. شاید این ناهمسانی مربوط به ویژگی‌هایی دیگر باشد که در تحلیل دیگری باید بررسی شود.

شکل ۲ سنجه کاا فرصت شغلی آینده

شکل ۲ نشان دهنده کاآ برحسب تفکیک پاسخ‌ها با توجه به نظر شرکت‌کنندگان درباره فرصت شغلی آینده است. دو

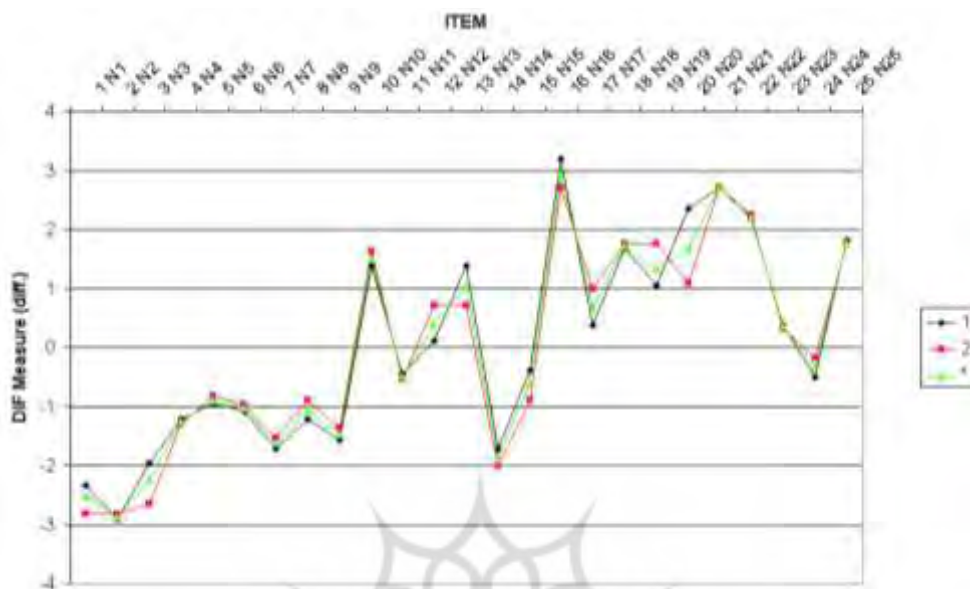
آیتم ۱۰ و ۲۰ هر یک تفکیک بیشتری برای شرکت‌کنندگان گروه ۲ از خود نشان می‌دهند. همچنین، آیتم ۱۹ و ۲۵ در درجه دوم تفکیک قرار می‌گیرند. آیتم ۱۸ برای گروه ۱ تفکیک بالاتری نشان می‌دهد. پس از آن آیتم ۳ برای گروه

جدول ۹ کاا برای گروه فرصت تحصیلی آینده

Item Number	Name	Person Class	DIF Measure	Person Class	DIF Measure	DIF Contrast	Rasch-Welch		
							t	df	prob.
3	N3	1	-1.91	2	-2.57	.67	1.43	147	.1538
16	N16	1	4.17	2	2.52	1.65	1.50	88	.1367
21	N21	1	2.24	2	3.19	-.95	-1.37	146	.1729
22	N22	1	2.70	2	1.93	.77	1.21	118	.2279
25	N25	1	1.37	2	2.20	-.83	-1.63	147	.1050

آیتم‌های ۳، ۱۶ و ۲۲ به نفع گروه ۲ و آیتم‌های ۲۱ و ۲۵ به نفع گروه یک است.

توضیح کاا در نگرش فرهنگی



یک هستند. در تفسیر نمره افراد شرکت‌کننده در آزمون‌های هدف آن تعیین تأثیر نگرش فرهنگی شرکت‌کنندگان درباره یادگیری زبان مقصد است مشخص کردن هر یک از این ویژگیها در میزان روایی آزمون تأثیر دارد.

شکل ۳ سنجه کاا برای نگرش فرهنگی

چنان‌که در شکل ۳ دیده می‌شود اکثر آیتم‌ها برای هر دو گروه تقریباً یکسان عمل کرده‌اند. با این وجود با توجه دقیق تر به اطلاعات به دست آمده دیده می‌شود که آیتم‌های ۳، ۱۳ و ۲۰ به نفع گروه دو و آیتم‌های ۱۲، ۱۷ و ۱۹ به نفع گروه

جدول ۱۰ کاا برای گروه نگرش فرهنگی

Item Number	Name	Person Class	DIF Measure	Person Class	DIF Measure	DIF Contrast	Rasch-Welch		
							t	df	Prob.
3	N3	1	-1.96	2	-2.66	.69	1.42	129	.1579
12	N12	1	.12	2	.72	-.60	-1.56	140	.1206
13	N13	1	1.39	2	.72	.67	1.59	147	.1148
17	N17	1	.38	2	.98	-.60	-1.51	138	.1340
19	N19	1	1.04	2	1.76	-.72	-1.54	131	.1269
20	N20	1	2.35	2	1.08	1.27	2.51	144	.0133

آیتم‌های ۳، ۱۳ و ۲۰ به نفع گروه ۲ و آیتم‌های ۱۲، ۱۷ و ۱۹ به نفع گروه یک است.

International Journal of Languages' Education and Teaching. Volume 5, Issue 3, p. 12-36

Bachman, L. F. (2003). Constructing an assessment use argument and supporting claims about test taker-assessment task interactions in evidence-centered assessment design. *MEASUREMENT-LAWRENCE ERLBAUM ASSOCIATES-*, 1, 63-65.

Bachman, L. F. (2005). Building and supporting a case for test use. *Language Assessment Quarterly*, 2, 1-34.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford, England: Oxford University Press.

Bordbar, S. (2020). Differential Item Functioning of University Entrance Exam: Using Rasch Analysis. *Foreign Language Research Journal*, 10(1), 37-55. doi: 10.22059/jflr.2019.278170.611

Council of Europe. (2001). *Common European Framework of Reference for Languages*. Cambridge, UK: Cambridge University Press.

Cummins, A. (2012). *Validation of language assessments*. The encyclopedia of applied linguistics.

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing*. Cambridge, England: UCLES/Cambridge University Press.

Davis, A. (2013). Fifty Years of Language Assessments. In *The Companion to Language Assessment*, A. J. Kunnan (Ed). doi:10.1002/9781118411360.wbcla127

Dörnyei, Z., & Taguchi, T. (2009). *Questionnaires in second language research: Construction, administration, and processing*. Routledge.

Domna, A., & Zafiri, M. (2018). *A Case Study of Two Groups of A1 Level Students in English*. Education and Linguistic Research.

Elatia, S. (2011). *Choosing language competence descriptors for language assessment: validity and fairness issues*. Synergies Europe, 6, 165-175.

بهترین آزمون‌ها تنها در شرایطی که در آن تعریف و تهیه می‌شوند روایی دارند. بنابراین برای هر کاربری در موقعیت دیگر نیاز به بازتعریف روایی دارند. تعریف روایی پدیده‌ی سیال است. به عبارت دیگر هر آزمون برای اهداف تعریف شده‌اش در زمان و مکان معین روایی دارد. چما زمینه مناسب برای تحقیق و تحقق روایی آزمون زبانی را در اختیار قرار می‌دهد. زیرا برای همه زبان‌ها تعریف شده است و به یک زبان خاص وابسته نیست.

استدلال کاربرد ارزشیابی ساختاری منطقی است که در آن داده‌ها در چهارچوبی معین و قابل مقایسه باهم شواهد لازم را برای مشخص ساختن میزان روایی هر آزمون به دست دهد. این چهارچوب می‌تواند انواع محاسبات آماری را که به‌صورت کمی تهیه شده است در خود جای داده و ارزیابی کیفی لازم را بنا بر شرایط مختلف به وجود آورد.

همچنین داده‌های تحقیق حاضر مشخص می‌کند که کارکرد افتراقی آیت‌های یک آزمون را نمی‌بایست ثابت و قطعی دانست بلکه لازم است تفسیر کاربردی نمرات را با توجه به شرایط خاص شرکت‌کنندگان در آزمون و اهداف مد نظر آنان و نیز سازمان‌ها و افراد ذی‌نفع تعیین کرد.

۸. منابع

Alderson, J.C., (2007). *The CEFR and the Need for More Research The Modern Language Journal*, 91(4), pp. 659-663.

Ali, Z. M., Ali, F., Radzuan, N. M., Alwi, N. N. M., Abu, N. L., & Kassim, Z. (2018). Contextualising the CEFR: the Universiti Malaysia Pahang English language proficiency writing test. In 11th Annual International Conference of Education, Research and Innovation (ICERI 2018) (pp. 4892-4902).

Arrighi, J. T., & Piccoli, L. (2018). SWISSCIT Index on Citizenship Law in Swiss Cantons: Conceptualisation, Measurement, Aggregation. Université de Neuchâtel.

Arsalan, A. & Özeinci, S. (2017) *A CEFR-based Curriculum Design for Tertiary Education Level*

- Linacre, J. M. (2019). *Winsteps*[Computer program]. Chicago, IL: Winsteps.com.
- Little, D. (2014). Learning, teaching, assessment: An exploration of their interdependence in the CEFR. In *5th international conference on teaching English as a foreign language assessment in ELT: Opportunities and challenges, FCSH, Lisbon new university, Portugal* (pp. 21-22).
- Lundell, F., Lindqvist, C., & Edmonds, A. (2018). Productive Collocation Knowledge at Advanced CEFR Levels: Evidence from the Development of a Test for Advanced L2 French. *Canadian Modern Language Review*, 74(4), 627-649.
- Martyniuk, W. (Ed.). (2010). *Aligning Tests with the CEFR: Reflections on Using the Council of Europe's Draft Manual* (Vol. 33). Cambridge University Press.
- Megyesi, B., Granstedt, L., Johansson, S., Prentice, J., Rosén, D., Schenström, C. J., ... & Volodina, E. (2018, November). Learner Corpus Anonymization in the Age of GDPR: Insights from the Creation of a Learner Corpus of Swedish. In *Proceedings of the 7th Workshop on NLP for Computer Assisted Language Learning (NLP4CALL 2018) at SLTC, Stockholm, 7th November 2018* (No. 152, pp. 47-56). Linköping University Electronic Press.
- McNamara, T. (2005). 21st century shibboleth: Language tests, identity and intergroup conflict. *Language Policy*, 4(4), 351-370. <https://doi.org/10.1007/s10993-005-2886-0>
- McNamara, T., & Ryan, K. (2011). *Fairness versus justice in language testing: The place of English literacy in the Australian Citizenship Test*. *Language Assessment Quarterly*, 8(2), 161-178.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York: American Council on Education and Macmillan.
- Nagai, N., & O'Dwyer, F. (2011). *The actual and potential impacts of the CEFR on language education in Japan*. *Synergies Europe*, 6, 141-152.
- Negishi, M., Tono, Y., & Fujita, Y. (2012). *A validation study of the CEFR levels of phrasal*
- Fulcher, G., & Davidson, F. (Eds.). (2013). *The Routledge handbook of language testing*. Routledge.
- Guerra, L., Gonçalves, O., Fiske, F. N., & Gungor, M. N. (2018). A CEFR-based comparison of ELT curriculum and course books used in Turkish and Portuguese primary schools.
- HUANG, T., & JIA, G. (2012). *The feasibility study of linking language test to the CEFR—Taking College English Test as an example* [J]. *Foreign Language Testing and Teaching*, 1, 007.
- Kane, M. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527–35.
- Karami, M., Jafarigohar, M., Tajeddin, Z., & Rouhi, A. (2018). Input-induced Variation in EFL Learners' Oral Production in Terms of Complexity, Accuracy, and Fluency. *Iranian Journal of English for Academic Purposes*, 6(2), 70-85.
- Karimi, F., Chalak, A., & Biria, R. (2019). *Pedagogical Utility of Pre-Listening Activities for Improving Iranian Elementary EFL Learners' Listening Comprehension*. *International Journal of Instruction*, 12(1), 1127-1140.
- Kunnan, A. (2000). *Fairness and justice for all*. In A. J. Kunnan (Ed.), *Fairness and validation in language assessment*, Vol. 9, *Studies in language testing* (pp. 1–14). Cambridge, UK: UCLES/CUP.
- Kunnan, A. (2004). Test fairness. In M. Milanovic & C. Weir (Eds.), *European language testing in a global context* (pp. 27–48). Cambridge, England: UCLES/Cambridge University Press.
- Kunnan, A. (2008). Towards a model of test evaluation: Using the test fairness and test context frameworks. In L. Taylor & C. Weir (Eds.), *Multilingualism and assessment* (pp. 229–51). Cambridge, England: UCLES/Cambridge University Press.
- Le, H. T. T. (2018). Impacts of the CEFR-Aligned learning outcomes implementation on assessment practice. *Hue University Journal of Science: Social Sciences and Humanities*, 127(6B), 87-99.

Language testing and assessment (2nd ed., pp. 177-96). New York, NY: Springer.

Xi, X. (2010). *How do we go about investigating test fairness?* *Language Testing*, 27, 147-170

verbs in the English Vocabulary Profile. *English Profile Journal*, 3, e3
doi:10.1017/S2041536212000037

Nematzadeh, A. (2018). Construct Irrelevant Factors and Test Validity: Investigating the Relationship among Gender, Age, Mother Tongue, Field of Study and TOEFL IBT® Results. *Foreign Language Research Journal*, 8(1), 139-166. doi: 10.22059/jflr.2018.242996.405

Peachy, W. S. (2012). Şen, Y. (2012). *The Common European Framework of Reference for Languages (CEFR) and English Language Teaching in Higher Education in Turkey*. In International Higher Education Symposium.

Piccardo, E. (2012). *Multidimensionality of assessment in the Common European Framework of References for Languages (CEFR)*. *Les Cahiers de l'ILOB/OLBI Working Papers*, 4, 37-54.

Pérez de la Calle, S. (2018). The use of advertisement as didactic resource in the foreign language classroom according to sociocultural, linguistic, sociolinguistic and pragmatic aspects.

Razavipour, K. (2019). Philosophy of Research in Language Testing: Investigating Papers Published in Iranian, Peer-reviewed, Domestic Journals from 2008 to 2018. *Foreign Language Research Journal*, 9(3), 831-860. doi: 10.22059/jflr.2019.262709.535

Shakirova, D. S., Zamaletdinov, R. R., & Ashrapova, A. K. (2018). The Impact of the Cefr in Testing Tatar as a Foreign Language (A2 Level). *The Journal of Social Sciences Research*, 4, 36-39.

Shaw, S., & Imam, H. (2013). *Assessment of international students through the medium of English: Ensuring validity and fairness in content-based examinations*. *Language Assessment Quarterly*, 10(4), 452-475.

Toulmin, S. E. (2003). *The uses of argument*. Cambridge university press.

Wu, J. R., & Wu, R. Y. (2007). Using the CEFR in Taiwan: The perspective of a local examination board. *The Language Training and Testing Center Annual Report*, 56, 1-20.

Xi, X. (2008). Methods of test validation. In E. Shohamy & N. H. Hornberger (Eds.), *Encyclopedia of language and education*. Vol. 7: