

چالش های ارزشیابی در آموزش زبان خارجی

سمیرا خدادادی^۱، روح الله رحمتیان^۲

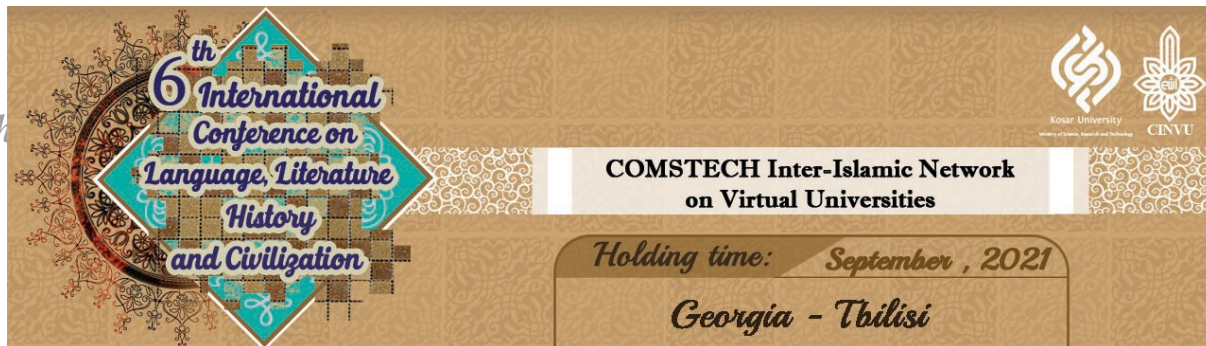
۱- دانشجوی دکتری آموزش زبان فرانسه دانشگاه تربیت مدرس تهران

۲- دانشیار گروه آموزش زبان فرانسه دانشگاه تربیت مدرس تهران

چکیده

ارزشیابی همواره یکی از چالش برانگیزترین مسائل در آموزش و خصوصاً آموزش زبان بوده است، چراکه علاوه بر وجود نظریات مختلف و گاهاً متناقض، در آموزش زبان خارجی با امر ارزشیابی مهارت های زبانی نیز سروکار داریم که ارزشیابی را برای ارزیابی کنندگان و مدرسان دشوارتر می کند. در تحقیق حاضر سعی شده تا با شناسایی و دسته بندی مسائل چالش برانگیز ارزشیابی در آموزش زبان های خارجی، این مساله را بیشتر مورد بررسی قرار دهیم تا در نهایت بتوان با شناسایی بهتر این چالش ها، راه های بهتری برای رویارویی با آنها پیدا کرد. بر اساس مطالعات صورت گرفته از تحقیقات و کتاب های موجود در این حوزه، چالش های آموزش زبان خارجی را به سه دسته چالش های مربوط به ارزیابی کننده، چالش های مربوط به آزمون و فرایند ارزشیابی و چالش های مربوط به محتوای مورد ارزیابی تقسیم نموده و مسائل مربوط به هر کدام مورد بررسی قرار دادیم. در نهایت، نتایج تحقیق حاضر برای متخصصان حوزه ارزشیابی و آموزش زبان خارجی که سعی بر درک ابعاد مختلف چالش های ارزشیابی در آموزش زبان دارند مناسب است.

واژگان کلیدی: ارزشیابی، آزمون، آموزش زبان های خارجی، چالش های ارزشیابی آموزشی



مقدمه

ارزشیابی که بعنوان فرایند تعیین و گردآوری اطلاعات لازم درباره مطلوبیت اهداف، برنامه های عملیاتی، طرح های اجرایی و نتایج حاصل از آنها به منظور هدایت و تصمیم گیری جهت بهبود و پیشرفت فعالیت ها برای کسب بازدهی مورد نظر (بازرگان، ۱۳۷۴) تعریف می شود، همواره یکی از پرچالش ترین موضوعات در علوم آموزشی بوده است، زیرا بعنوان یک عنصر سازمان دهنده در سیستم آموزش محسوب می شود که هم بر اقدامات آموزشی و هم بر فعالیت فراگیران تاثیر گذار است (وان دو کرخف^۱، ۲۰۱۷). از زمانی که بیچر^۲ (۱۹۸۲) وجود نظریه های متفاوت و گاه متناقض در آموزش را دلیلی بر هرچه چالش برانگیزتر شدن ارزشیابی آموزشی برشمرد تا امروز که صاحب نظران با ایجاد چارچوب اروپایی مشترک مرجع برای زبان ها^۳ سعی در یکپارچه سازی آموزش زبان ها در همه حوزه ها از جمله ارزشیابی دارند، هنوز ارزشیابی خصوصاً در امر آموزش زبان های خارجی پر از چالش است.

برای نشان دادن پرچالش بودن ارزشیابی در آموزش ذکر این نکته کافی است که حتی اصطلاح ارزشیابی، خود یک مفهوم چالش برانگیز است، چراکه در ۲۰ سال اخیر بیش از صد تعریف برای ارزشیابی ارائه شده است (پاتون^۴، ۱۹۸۱) و طبیعتاً باید در ابتدا در مورد مفهوم ارزش به اتفاق نظر برسیم تا بتوانیم تعریف مناسبی از ارزشیابی ارائه دهیم و در نهایت و بر اساس آن تعریف، ارزشیابی را در عمل پیاده سازی کنیم. در تعاریف و نظریات مختلف ارزش به گونه های مختلفی تعریف شده و همین مساله تعریف ارزشیابی را دشوار کرده است.

در اینجا به منظور پرداختن بهتر به چالش های ارزشیابی آنها را در سه گروه طبقه بندی می کنیم و در ادامه به هریک از آنها به صورت جداگانه خواهیم پرداخت. این چالش ها عبارت اند از:

- ۱- چالش های مربوط به ارزیابی کننده
- ۲- چالش های مربوط به محتوای مورد ارزشیابی
- ۳- چالش های مربوط به آزمون و فرایند ارزشیابی

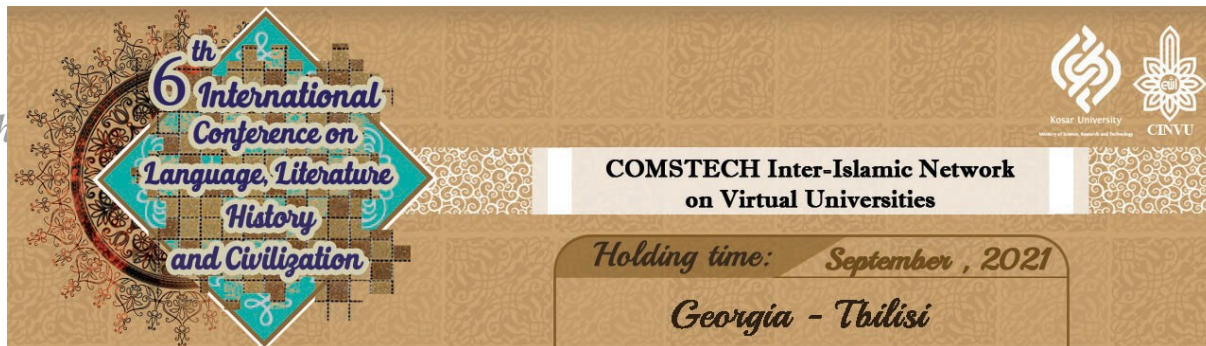
^۱ واژه چالش که از حدود ۳۰ سال پیش در متون فارسی رواج بیشتری یافت، در فرهنگ های دهخدا و معین معانی مختلفی نظیر تلاش، زد و خورد و مباشرت دارد اما معمولاً معنای کاربردی آن با معنای لغت نامه ای آن متفاوت است و در کاربردی به معنای مبارزه طلبی و موقعیت دشوار پیش رو است که بعنوان معادل کلمه انگلیسی Challenge و کلمه فرانسوی l'enjeu است. فرهنگ لغات BBC واژه challenge را بصورت "وضعیت و پدیده ای جدید و دشوار که مواجهه با آن تلاشی سخت و تعیین کننده را ایجاد کند" تعریف می کند و منظور ما از چالش در مطالعه حاضر نیز همین است.

^۲ Van de Kerkhove

^۳ Baecher

^۴ CECRL (Cadre Européen Commun de Référence pour les Langues) / CEFRL (Common European Framework of Reference for Languages)

^۵ Patton



بعد از ارائه توضیحات لازم در خصوص هر چالش سعی می‌کنیم روش‌های علمی استفاده و پیشنهاد شده در منابع و مقالات علمی مختلف را بعنوان روش رویارویی با هر چالش به معلمان و بطور کلی ارزیابی کنندگان آموزشی ارائه دهیم.

۱- چالش‌های مربوط به ارزیابی کننده

از آنجایی که بسیاری از معلمان، آموزش قابل توجهی در رابطه با ارزشیابی ندیده‌اند، بسیاری از آنها با چالش‌های درونی در رابطه با عدالت و ... مواجه هستند و طبیعی است که در بسیاری موارد بصورت سلیقه‌ای و طبق برداشت شخصی خودشان از عدالت آموزشی عمل می‌کنند. همانطوری که اشاره شد، یکی از مهم‌ترین دلایل روی آوردن معلمان به رویکردهای سلیقه‌ای و شخصی، عدم وجود آموزش کافی و عملی در رابطه با ارزشیابی آموزشی برای معلمان است و حتی با نگاهی به منابع موجود، به این نکته پی می‌بریم که نوعی عدم توازن بین تعداد منابعی که معلمان برای آماده‌سازی محتوای تدریس در اختیار دارند و منابعی که برای یادگیری و بدست آوردن ابزار ارزشیابی در اختیار دارند، وجود دارد (مورو^۶، ۲۰۱۸). از طرف دیگر، بر اساس تعریفی که پرتی^۷ ارائه می‌دهد، ارزشیابی شامل مجموعه پیچیده‌ای از مفاهیم و روش‌ها است و در نتیجه این خطر وجود دارد که معلمان بخواهند آن را برای خود ساده‌تر کنند و به عقیده پورن^۸، شرایطی که اکثر معلمان در آن هستند این خطر را به همراه دارد که آن‌ها دچار مشکل تلاش برای کاهش پیچیدگی ارزشیابی شوند و او این مشکل را متوجه معلمان، متخصصان آموزشی که معلمان را تربیت می‌کنند و خود موسسات آموزشی می‌داند (پورن، ۲۰۰۱). پس می‌بایست به این چالش‌ها بصورت علمی و صحیح پرداخته شود و راه‌حل‌های عملی برای معلمان حوزه آموزش، خصوصاً آموزش زبان خارجی که حوزه اصلی مدنظر ما در این مطالعه است ارائه شود، تا آنها بتوانند بدون تسهیل غیراصولی و کلاستن از پیچیدگی ذاتی عمل ارزشیابی، آن را پیاده‌سازی کنند.

چالش‌های مربوط به ارزیابی کننده را می‌توان به دو دسته تقسیم کرد:

۱- اثرات روانشناختی محل ارزشیابی

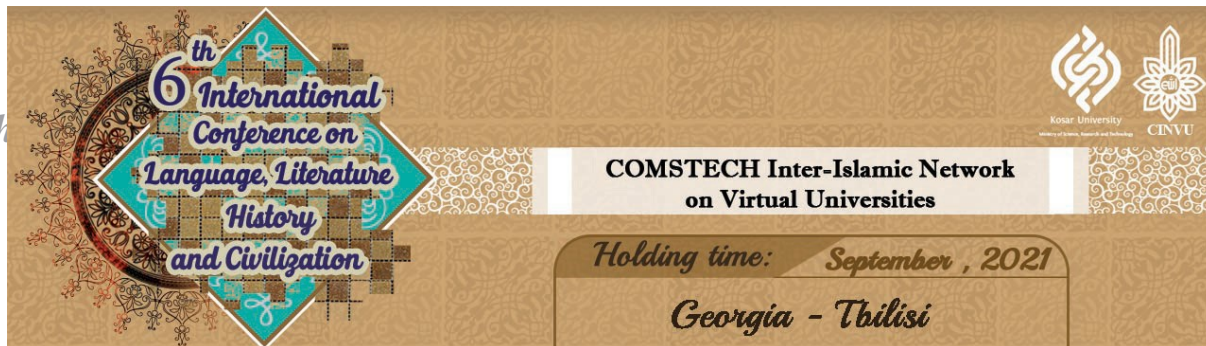
۲- دیدگاه ارزیابی کننده در مورد عدالت آموزشی

در ادامه به هر یک از این دو مقوله خواهیم پرداخت.

⁶ Morrow

⁷ Peretti

⁸ Puren

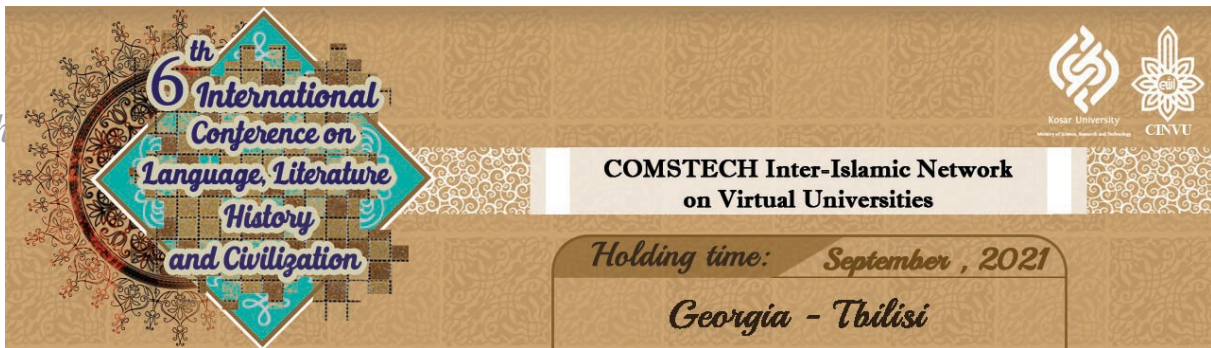


یکی از مهم ترین جاهایی که چالش های درونی بروز پیدا می کنند در هنگام انجام ارزشیابی، مثلاً در زمان تصحیح برگه های امتحانی و نمره دهی است. بنابر آنچه تحقیقات علمی نشان داده است، برخی اثرات پارازیتی^۹ وجود دارند که در هنگام انجام ارزشیابی می توانند بر ارزیابی کننده اثر بگذارند و دیدگاه وی در نمره دهی را تحت تاثیر مثبت یا منفی قرار دهند و منجر به خطا شوند. برای اینکه معلم از این خطاها اجتناب کند، باید نسبت به آنها آگاهی داشته باشد. این اثرات در ادامه توصیف شده اند:

۱. اثر خستگی: خستگی و شرایط تصحیح و ارزشیابی آزمون می تواند بر قضاوت فرد ارزیابی کننده تاثیر بگذارد، بعنوان مثال صبح ها معمولاً شرایط مطلوب تری برای انجام ارزشیابی فراهم است.
 ۲. اثر تضاد: اگر یک برگه امتحانی متوسط بلافاصله بعد از یک برگه امتحانی عالی مورد ارزشیابی و تصحیح قرار گیرد، احتمالاً ضعیف تر از آنچه هست قضاوت می شود و بالعکس، برگه ای که دقیقاً بعد از یک برگه امتحانی ضعیف قرار گیرد، احتمالاً نمره بالاتری نسبت به استحقاقش دریافت خواهد کرد.
 ۳. اثر ترتیب قرارگیری: یک ارزیابی کننده بطور معمول در انتهای تصحیح اوراق امتحانی سخت گیری بیشتری به خرج می دهد و احتمالاً برگه های آخر نمرات کمتری نسبت به استحقاق خود دریافت می کنند.
 ۴. تاثیر اولیه: تمایل به ارزیابی افراد بر اساس تاثیر برخورد اول و نه بر اساس تحلیل عملکرد وی در طول دوره را گویند. گاهی اوقات اولین برخورد تصویری را در ذهن فرد ایجاد می کند که در نتیجه ارزشیابی تاثیر می گذارد.
 ۵. اثر شوک: تکرار یک خطا می تواند منجر به پایین رفتن نمره بصورت افراطی شود.
 ۶. اثر قطره آب: در طول آزمون، فرد ارزیابی کننده چندین بار مشاهده خطاهایی نظیر ایراد نوشتاری نظیر املا را تحمل می کند، اما از یک نقطه به بعد به ناگهان تحمل وی به سر آمده و شروع به محاسبه آن خطا می کند و این باعث پایین آمدن نمره می شود. خطاهای رخ داده بعد از آن نقطه در حکم قطره آبی را دارد که گرچه کوچک است اما باعث سرریز شدن صبر ارزیابی کننده می شود. (چارچوب اروپایی مشترک مرجع زبان ها، ۲۰۰۱)
- لازم به ذکر است که صرفاً آگاهی ارزیابی کنندگان و معلمان از اثرات و خطاهای فوق، می تواند تا حد زیادی منجر به کاهش اثرات آنها در هنگام ارزشیابی شود.

اما غیر از اثرات روانشناختی محل ارزشیابی، نوع نگرش ارزیابی کننده و بطور کلی معلم به عدالت در آموزش نیز اهمیت دارد و همین تفاوت در نگرش ها باعث تفاوت های بسیاری در رویکرد معلمان مختلف به ارزشیابی می شود و نتایج متفاوتی را برای ارزیابی شونده ها به همراه دارد و این در حالی است که یکپارچگی و یکسان بودن رویکردها باعث ایجاد نتایج قابل قیاس تر و منصفانه تری می شود. در رابطه با این همسان سازی روش ها و معیارهای ارزشیابی، باید گفت که معیارهای و روش هایی برای ارزشیابی در امر آموزش ارائه شده است که بر اصول و نظریات مختلفی استوار است و حتی پیروی از

⁹ Effets Parasites / Parasite effets



یکی از این روش‌ها باعث غلبه بر این چالش ارزشیابی نمی‌شود. مثلاً، با نگاهی به پیشینه ارزشیابی در می‌یابیم که برای مدت زیادی ارزشیابی ابزاری برای اندازه‌گیری میزان یادگیری دانش بکار می‌رفت. با این حال، اخیراً در تحقیقات زیادی نشان داده شده است که رایج‌ترین شکل‌های ارزشیابی، خصوصاً اغلب سیستم‌های نمره دهی، بیشتر متمرکز بر مقایسه اجتماعی ارزیابی شوندگان با همکلاسی‌های خود هستند تا اینکه متمرکز بر وظیفه‌ای که فرد باید انجام دهد، باشند و این مقایسه از آن جهت تهدید محسوب می‌شود که می‌تواند به ضرر حرکت به سمت خودارزیابی است و باعث آسیب به عملکرد فرد می‌شود، اعتماد به نفس را کاهش داده و منجر به بروز رفتارهایی گاهاً ضداجتماعی می‌شود (بو ترا و همکاران^{۱۰}، ۲۰۱۱). در نتیجه آنچه در حال حاضر در علم ارزشیابی بعنوان اصول ارزشیابی تعریف شده است، منتج شده از نظریات متناقض و مختلف و در نتیجه غیرهمگراست.

امروزه ثابت شده است که معلمان مختلف از نظر خودسازگاری^{۱۱}، سخت‌گیری و نوع نمره‌دهی عناصر مختلف در ارزشیابی متفاوت عمل می‌کنند (شوهامی و سایرین^{۱۲}، ۲۰۱۷) و قطعاً آموزش آنها در راستای یادگیری و به کارگیری یک روش یکتا برای ارزشیابی باعث کاهش این تفاوت‌ها و نزدیک‌تر شدن به عدالت آموزشی می‌شود. می‌بایست کدهای اخلاقی یکسانی به شکل استاندارد تعریف شده و در اختیار همه ارزیابی‌کنندگان آموزشی و معلمان قرار گیرد، چراکه اخلاق حرفه‌ای باعث فراهم شدن قراردادی بین حرفه مورد نظر، فرد و عموم مردم می‌شود و از هر سه مراقبت می‌کند. (داوینز^{۱۳}، ۱۹۹۷) و در نتیجه این قرارداد می‌تواند به شکل کد اخلاقی، کد عمل و یا یک سند استاندارد باشد تا در اختیار هر فردی در آن حرفه قرار گیرد (فولچر و دیویدسون^{۱۴}، ۲۰۰۷).

در رابطه با قضاوت و نگاه فردی به عدالت نمی‌توان به نظر شخصی بسنده کرد چراکه یک فرد به همه جوانب آن آگاه نیست و آنچه در حال حاضر قضاوت می‌کنیم، ممکن است طبق استانداردها و تعاریف آینده نادرست قضاوت شود (رورتی^{۱۵}، ۱۹۹۹). به گفته فولچر و دیویدسون (۲۰۰۷) سال‌ها بعد از مرگ شما، افراد آگاه تر و خبره‌تری احتمالاً کار شما را یک اشتباه فاحش قضاوت می‌کنند و همین‌طور ممکن است باورهای علمی شما در آینده فقط با ارجاع به یک الگوی منسوخ شده قابل فهم باشند. از طرف دیگر، درک و تفاهم جمعی در میان افراد یک جامعه کمک می‌کند تا بهتر تعریف کنیم که چه چیزی درست است و چه چیزی اشتباه (فولچر و دیویدسون، ۲۰۰۷).

این همسان‌سازی روش‌ها و معیارهای قضاوت در ارزشیابی منجر به همگرا شدن نتایج ارزشیابی و در نتیجه عدالت آموزشی بیشتر می‌شود. به عقیده پیرس^{۱۶} (۱۸۷۷) اگر افراد یک جامعه در تحقق اهداف خاصی با هم متعهد باشند و روش یکتایی

¹⁰ Butera et al

¹¹ Self-consistency / auto-cohérence

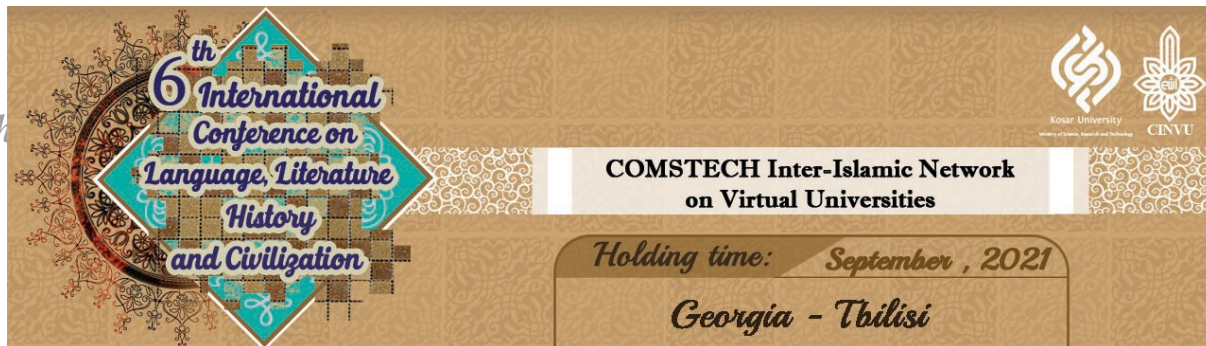
¹² Shohamy et al

¹³ Davies

¹⁴ Fulcher and Davidson

¹⁵ Rorty

¹⁶ Pierce



برای رسیدن به آن در پیش بگیرند، ممکن است در ابتدا به نتایج متفاوتی برسند اما در ادامه با تکمیل کار هر کدام، همگی به سمت نتایج یکنابیی پیش خواهند رفت و نتایج قضاوتشان همگرا خواهند شد. در نهایت، بنابر آنچه گفته شده برای حل چالش‌های مربوط به ارزیابی کننده، می‌توان بطور خلاصه راهکارهایی زیر را در قالب مراحل زیر ارائه کرد:

- ۱- تلاش برای حرکت به سوی حل تناقضات نظری در مورد اصول و روش‌های ارزشیابی
- ۲- تعریف استانداردهای ارزشیابی آموزشی و روش‌های ارزشیابی همگرا بر اساس آن‌ها
- ۳- آموزش مدرسان و ارزیابی کنندگان در رابطه با استانداردها و روش‌های استاندارد ارزشیابی
- ۴- آگاه‌سازی مدرسان و ارزیابی کنندگان از اثرات روانشناختی محل در ارزشیابی

۲- چالش‌های مربوط به محتوای مورد ارزیابی

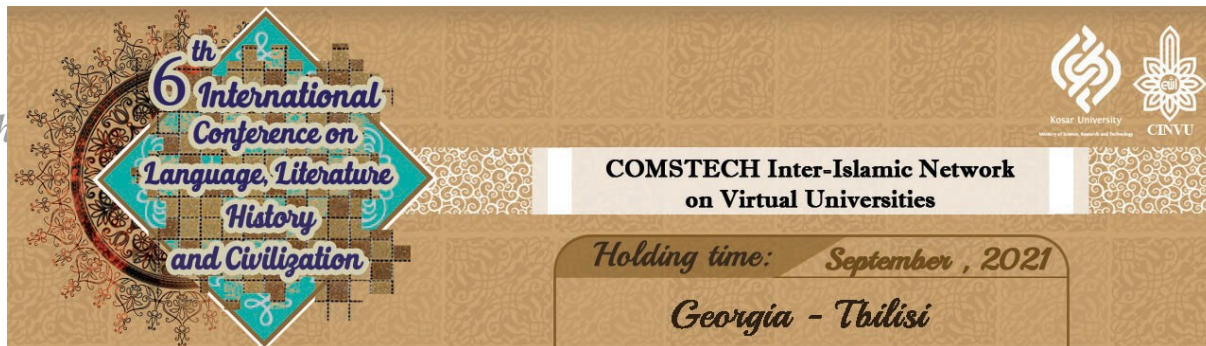
همانطوری که پیش‌تر اشاره شد، وجود تعاریف متعدد و مفاهیم تعریف شده گاهاً متناقض باعث ایجاد ابهام در ذهن ارزیابی کننده می‌شود. در برخی تعاریف با مفاهیمی نظیر محاسبه و ارقام مواجه هستیم که قابل اندازه‌گیری و عینی‌تر هستند و در برخی تعاریف دیگر با مفاهیمی نظیر تخمین تقریبی مواجهیم که حاشیه خطایی برای آن‌ها در نظر بگیریم و معیارهای آنها مبهم‌تر و در نتیجه ذهنی‌تر هستند (تاردیو^{۱۷}، ۲۰۰۶). وجود این دوگانگی باعث می‌شود به این فکر کنیم که آیا واقعا ارزشیابی منصفانه‌ای انجام می‌دهیم یا خیر؟ بنابراین مشکل قابل اندازه‌گیری بودن، در برخی موضوعات نظیر توانش زبانی وجود دارد و باعث ایجاد ابهام و تناقض در ذهن فرد شده و چندگانگی در عمل بوجود می‌آورد. در مورد هدف ارزشیابی و آنچه مورد ارزشیابی قرار می‌گیرد نیز ابهام وجود دارد (تاردیو، ۲۰۰۶). در رابطه با آموزش و یادگیری زبان ما بطور همزمان هم توانش و هم کنش را از طریق عملکرد فرد در زبان مقصد می‌سنجیم. اگر نگاهی تاریخی به متدهایی که تاکنون استفاده شده بیاندازیم، متوجه می‌شویم که این مفاهیم یک واقعیت یکتا را نشان نمی‌دهند و در واقع هر متد آموزشی، عنصر خاصی را مورد ارزشیابی قرار می‌داد.

مثلا در روش شنیداری-دیداری^{۱۸} زبان بصورت مجموعه ساختارهایی تعریف می‌شد که منجر به یادگیری می‌شد و بنابراین آنچه ارزشیابی می‌شد دانش زبانشناختی بود که بصورت مکانیکی آموخته می‌شد و همچنین توانش پیش-ارتباطی (این اصطلاح اولین بار توسط بولتون بکار برده شد) و تولید شفاهی هدایت شده (بایلی^{۱۹}، ۱۹۹۷). اما در رویکرد ارتباطی زبان بعنوان وسیله ارتباط واقعی در نظر گرفته می‌شود و در نتیجه علاوه بر دانش زبانشناختی، توانش ارتباطی شفاهی و کتبی

¹⁷ Tardieu

¹⁸ Audio-Visual method / la méthode Audio-Visuelle

¹⁹ Bailly



برای کسب استقلال زبانی و فراتوانش تفکر در مورد زبان، یادگیری و فرهنگ مورد ارزشیابی قرار می گیرند (بچمن^{۲۰}، ۱۹۹۰).

اما در حال حاضر چگونگی ارزشیابی توانش ارتباطی خود یکی از بزرگترین چالش هاست. در مورد تعریف آن همانطوری که می دانیم، طبق نظر چامسکی^{۲۱} (۱۹۸۱) با دو بخش توانش^{۲۲} و کنش^{۲۳} در یادگیری زبان روبرو هستیم و کنش نحوه بکارگیری آن توانش زبانی را شامل می شود.

اما برای ارزشیابی توانش زبانی، انواع مختلفی از آزمون تعریف شده است که به نظر بولتون^{۲۴} (۱۹۹۱) تنها آزمون سطح توانش و آزمون کنترل، واقعا به ارزشیابی توانش ارتباطی منجر می شوند، چراکه در آنها ارتباط بعنوان استفاده از کلام بطور کاملاً ادغام شده در موقعیت گفتمانی دیده می شود. اما مساله این است که آیا در عمل معلمان و ارزیابی کنندگان، خصوصاً در کلاس های آموزش زبان از چنین آزمون هایی برای سنجش توانش های زبانی فراگیرن بهره می برند یا خیر؟ هر فرد یا موسسه ممکن است رویکرد منحصر به فرد خود را در عمل برای ارزشیابی توانش زبانی به کار گیرند و به همین دلیل است که آزمون های استاندارد نظیر DELF و TCF و ... تعریف شده اند تا بتوان مقیاس سنجش یکتایی برای مقایسه افراد بدست آورد. اما با این وجود، معلمان به دلیل عدم آگاهی و آموزش کافی قادر نیستند در عمل توانش زبانی زبان آموزان را بسنجند و اغلب به بررسی میزان دانش زبانشناختی بسنده می کنند. به همین دلیل آزمون های آنها از روایی و پایایی قابل قبولی برخوردار نیست. بولتون در مورد مشکل روایی و پایایی آزمون های سنجش توانش ارتباطی چنین می گوید:

"اگر به هدف توانش ارتباطی تعاملی توجه کنیم، آزمون ها باید به گونه ای باشند که بیشترین امکان را برای پاسخ فردی خلاقانه به فرد بدهد. اما بالعکس، ما حاشیه تفسیر را در نمره دهی نتایج کتبی و شفاهی افزایش می دهیم و این به آن معناست که پایایی اندازه گیری را کاهش می دهیم."

در واقع هنر فرد ارزیابی کننده، قابل اندازه گیری کردن موضوعات مبهمی نظیر توانش های زبانی در آموزش است (تاردیو، ۲۰۰۶).

درمورد این چالش نیز آموزش معلمان و ارزیابی کنندگان مهم ترین امری است که به غلبه بر چالش مذکور کمک می کند و منجر به طراحی آزمون هایی با اعتبار و پایایی بیشتر برای سنجش توانش زبانی می شود. آگاه سازی معلمان از رویه ها و اصول ارزشیابی توانش های زبانی که توسط چارچوب اروپایی مشترک مرجع برای زبان ها تعریف شده است، یک گزینه خوب در این راستاست.

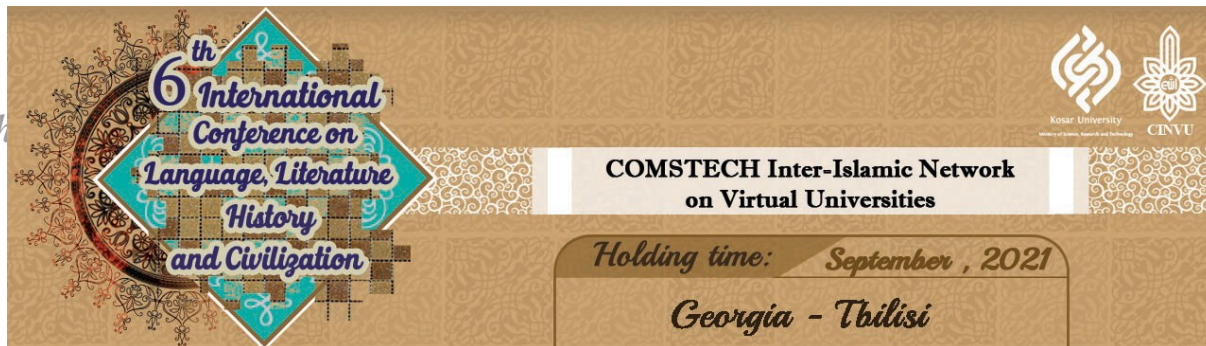
²⁰ Bachmann

²¹ Chomsky

²² Competence / La Compétence

²³ Performance / La Performance

²⁴ Bolton



۳- چالش‌های مربوط به آزمون و فرایند ارزشیابی

با حرکت بسیاری از سیستم‌های آموزشی به سوی نظام ارزشیابی توصیفی، تلاش برای کمرنگ‌تر کردن رقابت و افزایش انگیزه‌های شخصی در آموزش قابل مشاهده است (فولچر و دیویدسون، ۲۰۰۷) و برخی از صاحب‌نظران نیز در همین راستا تلاش کرده‌اند تا آزمون‌ها را هر چه کمرنگ‌تر کنند. اما به قول شوهامی (۲۰۰۱) "آزمون‌ها آمده‌اند که بمانند" و علیرغم اینکه امکان استفاده سوء از آنها هست، رویکردهای اخلاقی و دموکراتیک به آزمون منجر به ایجاد فرصت‌هایی برای آموزش و استخدام شده است (فولچر و دیویدسون، ۲۰۰۷).

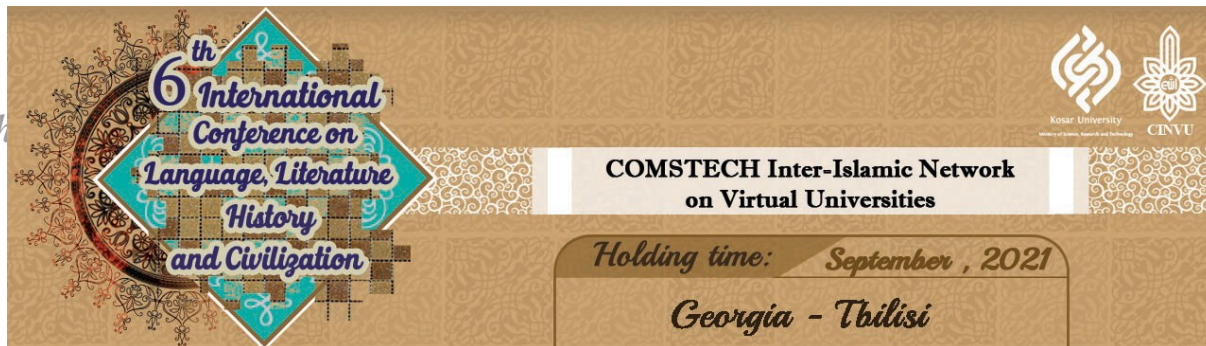
حال که می‌دانیم آزمون‌ها در امر آموزش اجتناب‌ناپذیر هستند، باید در راستای ایجاد شرایطی گام برداریم که بتوان همواره بر نتایج آزمون‌ها تکیه کرد. یعنی ارزیابی‌کنندگان و معلمان باید تلاش کنند که آزمون‌هایی طراحی کنند که قابل اعتماد و پایدار باشند. این دو اصطلاح (روایی و پایایی) همواره در طراحی آزمون‌ها مطرح بوده‌اند اما عدم آگاهی و کم‌توجهی بسیاری از طراحی‌کنندگان آزمون‌ها، باعث بوجود آمدن مشکلات جدی در رابطه با آنها در ارزشیابی آموزشی شده است. در ادامه به بررسی آنها می‌پردازیم:

الف) روایی آزمون: قابل اعتماد بودن آزمون که معمولاً با عنوان روایی شناخته می‌شود بدان معناست که سوالات آزمون به گونه‌ای طراحی شوند که روایتگر هدف آزمون باشند یعنی دقیقاً آنچه قصد سنجیدنش را داریم، مورد سنجش قرار دهند. انواع مختلفی از آزمون‌های روایی را می‌توان مورد استفاده قرار داد. این آزمون‌ها را تحت سه عنوان کلی دسته‌بندی می‌کنیم: روایی محتوا، روایی معیار و روایی سازه (دانایی فرد و همکاران، ۱۳۸۳).

الف - ۱) روایی محتوا:

روایی محتوا بدان معناست که همه ابعاد و مؤلفه‌هایی که می‌توانند مفهوم مورد نظر ما را انعکاس دهند، در آن سنجیده وجود دارد. هر چه وجود این ابعاد و مؤلفه‌ها در سنجیده، در راستای انعکاس مفهوم، بیشتر باشد، روایی محتوا بیشتر می‌شود. نردی^{۲۵} (۲۰۰۳) برای تفهیم این نوع روایی، از مثال "محتوای آزمون رانندگی" استفاده می‌کند، وی می‌گوید که آماده بودن یک شخص برای دریافت گواهینامه رانندگی، بجای متکی بودن به یک یا چند ویژگی جداگانه فرد ارزیابی شونده، به مجموعه ویژگی‌های وی وابسته است و در نتیجه آزمون رانندگی زمانی معتبر است که محتوای آزمون بصورت کلی ارزیابی شود.

²⁵ Nardi



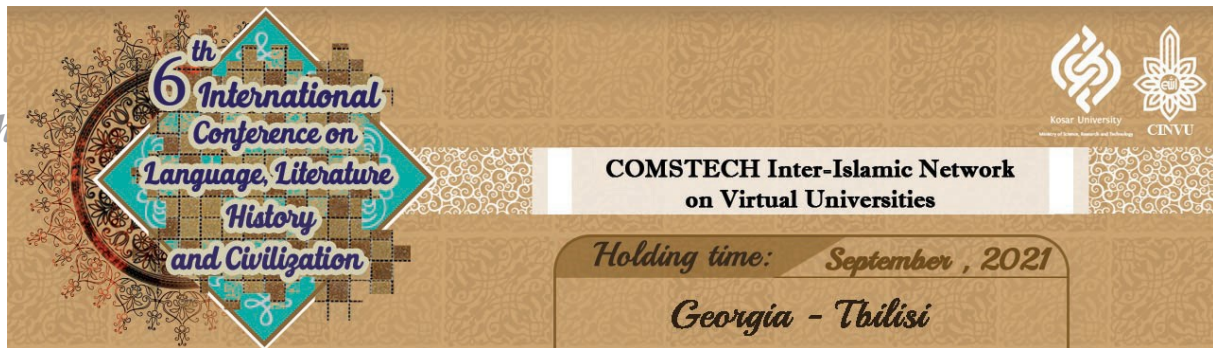
الف-۲) روایی معیار: روایی معیار وقتی ایجاد می‌شود که سنج‌های مورد نظر، افراد را بر اساس معیاری که انتظار پیش بین آن می‌رود، از هم متمایز کند. این امر می‌تواند به وسیله‌ی ایجاد روایی هم‌زمان و روایی پیش بین انجام شود. روایی هم‌زمان موقعی ایجاد می‌گردد که معیار و مقیاس آزمون، افرادی که مشخص است نسبت به یکدیگر متفاوتند را از هم متمایز می‌کند؛ یعنی آنها باید در آزمون، نمرات متفاوتی بگیرند. برای مثال، اگر یک سنج در مورد اخلاق کار تدوین می‌شود و در مورد گروهی از دریافت کنندگان خدمات رفاهی اجرا می‌گردد، مقیاس باید افرادی که مایلند شغلی را بپذیرند و از زیر چتر خدمات رفاهی بیرون آیند را از کسانی که می‌خواهند تحت پوشش خدمات رفاهی باقی بمانند و خواهان شغلی نباشد، متمایز گرداند، زیرا کسانی که برای اخلاق کار ارزش زیاد قائل‌اند، نمی‌خواهند تحت پوشش خدمات رفاهی باشند و خواهان پذیرش شغل هستند. از طرف دیگر کسانی که برای اخلاق کار ارزش کمی قائل‌اند، ممکن است تمایل داشته باشند کماکان از خدمات رفاهی برخوردار باشند ولی اشتغال به کاری نداشته باشند. اگر هر دو نوع از افراد، نمره یکسانی بر اساس مقیاس اخلاق کاری بگیرند، آن‌گاه آزمون نتوانسته است سنج‌های اخلاق کار باشد، بلکه سنج‌های از چیز دیگری بوده است.

اما روایی پیش بین، توانایی ایجاد تفکیکی بین افراد، در خصوص یک سنج است که در آینده رخ می‌دهد. نیومن (۲۰۰۷) آزمون SAT²⁶ را که یک آزمون استاندارد ورود به کالج در آمریکاست، مثال می‌زند. این آزمون سعی دارد تا قبل از ورود افراد به کالج، عملکرد آنها در آینده در کالج را پیش‌بینی کند و در نتیجه روایی پیش بین بدان معناست که افرادی که نمره بالایی در این آزمون می‌گیرند باید در آینده عملکرد بهتری در کالج داشته باشند. پس اگر افراد با نمره بالا در این آزمون در آینده در کالج عملکرد بهتری نسبت به افرادی که نمره پایین‌تری در این آزمون گرفته‌اند داشته باشند، می‌توان گفت که این آزمون از قدرت پیش‌بینی کنندگی و روایی پیش بین قابل قبولی برخوردار است. آزمون‌ها ممکن است از روایی پیش‌بینی کنندگی بالا، متوسط و یا پایینی برخوردار باشند.

الف-۳) روایی سازه: روایی سازه دلالت بر این دارد که نتایج به دست آمده از کاربرد سنج‌ها تا چه حد با تئوری‌هایی که آزمون بر اساس آنها طراحی شده، سازگاری دارد. روایی سازه بیشتر از روایی محتوا جنبه‌ی نظری دارد. بنا به تعریف، یک آزمون در صورتی دارای روایی سازه است که نمرات حاصل از اجرای آن به مفاهیم یا سازه‌های نظریه‌ی مورد نظر مربوط باشند (کرونباخ²⁷، ۱۹۷۰). برای مثال، یک آزمون یا پرسشنامه‌ی اضطراب در صورتی دارای روایی سازه است که نمرات حاصل از آن، با سازه‌هایی که در نظریه‌ی اضطراب آمده‌اند ارتباط داشته باشد. (دانایی فرد و همکاران، ۱۳۸۳)

²⁶ Scholastic Achievement Test

²⁷ Cronbach



به دو روش می توان اعتبار یک آزمون را سنجید. در روش اول، با استفاده از روش های تجربی و فرمول های آماری که بعد از جمع آوری داده ها بکار می بریم، اعتبار آزمون را می سنجیم. در روش دیگر که غیر تجربی است، بر اساس شهود و حس این کار را انجام می دهیم. مثلاً با تعریف کردن سازه یا توانایی ای که باید سنجیده شود، باید مشخص شود آن توانش دقیقاً چه چیزی است و شامل چه چیزی می شود (رایی و میکنی^{۲۸}، ۲۰۱۴). معمولاً بخاطر نبود آموزش کافی در رابطه با آزمون و شیوه های آزمون گیری، معلم ها معمولاً از حس و شهود خود و نیز آنچه خود درست تصور می کنند، استفاده می کنند که البته عاری از خطا نیست. اما در هر حال، آنچه مسلم است این است که آگاهی معامان و ارزیابی کنندگان از مفاهیمی نظیر روایی و پایایی یک اصل اساسی در حرکت به سوی حل چالش ها و معضلات مربوط به محتوای آزمون ها است.

ب) پایایی آزمون: به عقیده هنینگ^{۲۹} یک آزمون در صورتی پایدار است که اگر یک شرکت کننده در آزمون، دوباره در همان آزمون یا آزمون مشابه آن شرکت کند، باز هم همین نتایج را بدست آورد (هنینگ، ۱۹۸۷). مثلاً اگر شخصی را وزن کردیم و وزنش ۶۲ کیلوگرم بود، انتظار داریم در یک آزمون پایدار، بعد از یک ساعت هم اگر او را وزن کنیم به همان وزن برسیم.

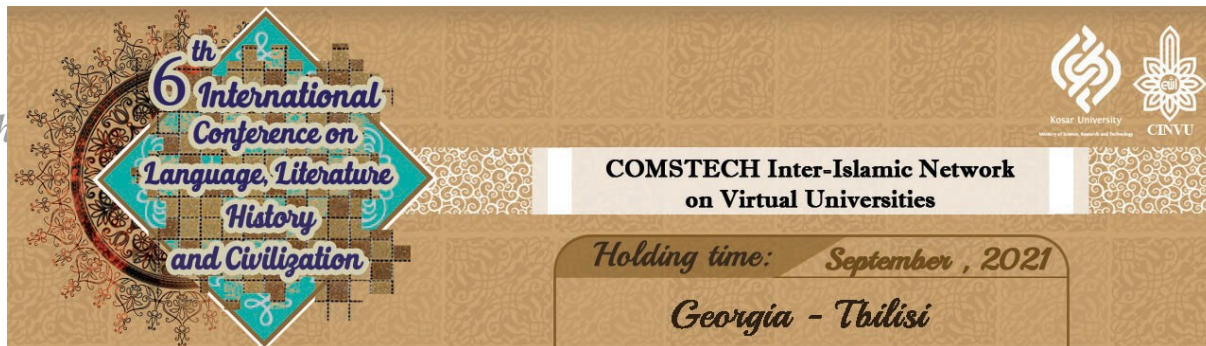
پایداری نشان دهنده دقت آزمون است (رایی و میکنی، ۲۰۱۴). طبیعتاً اندازه گیری پایداری یک آزمون برای بررسی توانش زبانی کاری دشوارتر است به این دلیل که با مفاهیمی انتزاعی سروکار داریم. براون^{۳۰} (۱۹۹۶) سه گروه از عوامل را بر روی پایایی تاثیر گذار می داند: ۱- عوامل محیطی (مثلاً یک محیط پر سروصدا، گرم یا سرد و تنگ و همچنین خستگی یا بیماری اثر منفی بر نمره افراد دارند). ۲- عوامل مربوط به رویه های اجرای آزمون (مثل وقتی که دستورالعمل آزمون غیرواضح باشد و یا زمان آزمون ناکافی باشد). ۳- ویژگی های موارد آزمون (معمولاً آزمون های طولانی تر و پرسوال تر بهتر می توانند شرکت کنندگان سطح های مختلف را متمایز کنند و این منجر به بالاتر رفتن پایداری تست می شود. میزان دشواری آزمون نیز اهمیت دارد و آزمون ها نباید بیش از حد دشوار یا بیش از حد آسان باشند. به عقیده هنینگ اندازه آزمون باید معقول و متناسب باشد اما بیشتر از ۷۵ مورد را ارزیابی نکند. از جمله روش های اطمینان از پایداری آزمون یکی روش آزمون-باز آزمون^{۳۱} است که در آن در فاصله کمتر از دو هفته از یک گروه از شرکت کنندگان، یک آزمون یکسان، دو بار گرفته شود و نتایج مقایسه شود، بدون اینکه در این فاصله به آنها بازخوردی در مورد آزمون اول داده شود و یا تمرینی در مورد آزمونی که داده اند انجام دهند.

²⁸ Rapi & Miconi

²⁹ Henning

³⁰ Brown

³¹ test-retest



یک روش دیگر، استفاده از آزمون‌های موازی است یعنی از یک گروه افراد دو آزمون مشابه (از نظر سطح دشواری) گرفته شود و این دو آزمون بدون دادن بازخورد و تمرین افراد، در فاصله کمتر از دو هفته انجام شوند (رایی و میکنی، ۲۰۱۴).

بحث و نتیجه‌گیری

در مطالعه حاضر سعی کردیم چالش‌هایی را که فرد ارزیابی‌کننده در آموزش، خصوصاً آموزش زبان، با آنها مواجه می‌شود، مورد بررسی قرار دهیم. در آخر باید توجه داشت که احتمالاً نمی‌توان تک تک آموخته‌های فرد زبان‌آموز را مورد ارزیابی قرار داد. البته این بحث که "آیا می‌توان همه چیز را ارزیابی کرد؟" خود همواره یکی از چالش‌های موجود در ارزیابی بوده است.

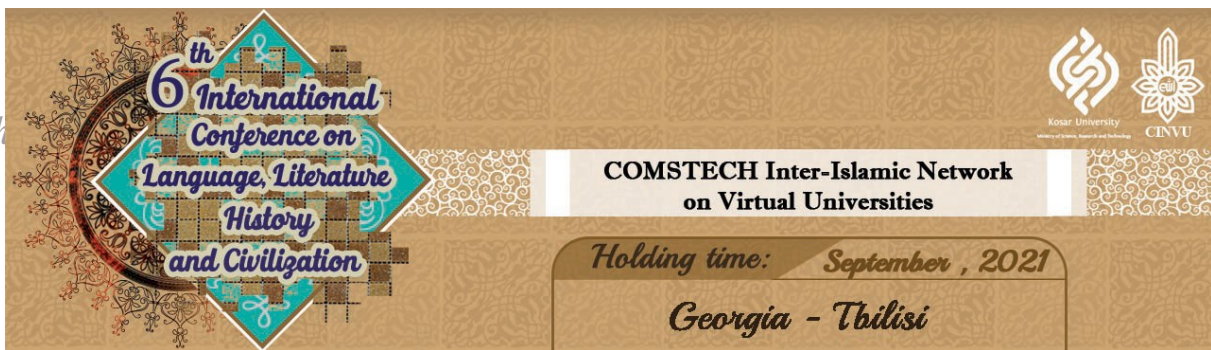
در واقع سوال اینجاست که اصولاً آیا لازم و یا موردپسند است که همه چیز را ارزیابی کنیم؟ در زمینه ارزیابی آموزشی نظرات مختلفی وجود دارد. از نظر دکرولی و بویسه^{۳۲} (۱۹۲۹)، هر آنچه قابل تدریس کردن و در نتیجه قابل یادگیری بوسیله فراگیران است، یعنی هر آنچه مورد آموزش قرار می‌گیرد، می‌تواند مورد ارزیابی نیز قرار گیرد. اما بسیاری از متخصصان ارزیابی با این عقیده مخالف هستند و معتقدند ما نمی‌توانیم و اصولاً لازم نیست که همه چیز را مورد ارزیابی قرار دهیم. اما نکته اینجاست که زمانی که آنها چنین گزاره‌ای را بیان کردند، هنوز توانش زبان‌شناختی اصلی‌ترین و تنها مساله مطرح در آموزش زبان بود و هنوز حرفی از توانش و عملکرد زبانی و زبان بعنوان یک مهارت مطرح نشده بود.

اما امروزه این امر در آموزش زبان مطرح می‌باشد چراکه با ارزیابی توانش‌های مختلف زبانی مواجه هستیم. بعنوان مثال استفانی شرتون^{۳۳} (۲۰۰۵) می‌گوید که توانش شنیداری بسیار پیچیده‌تر از توانش‌های نوشتاری هستند و در نتیجه معیارهای سنجش آنها می‌تواند تعداد بسیار زیادی باشند و عملاً سنجش همه آنها دشوار و غیرممکن است. وی نتیجه می‌گیرد که نمی‌توان و اصولاً نباید هر آنچه فراگیر انجام می‌دهد را مورد ارزیابی قرار داد و در عوض، فرد باید بتواند یک کار یا وظیفه یا تمرین را در کلاس انجام دهد بدون اینکه مجبور به قرار گرفتن در معرض نگاه ارزیابی‌کننده یک نفر باشد.

در واقع به این دلیل که به گفته چامسکی با دو بخش توانش و عملکرد بصورت جدا از هم مواجهیم، ارزیابی با چالش روبرو می‌شود. چون آنچه در درون و ذهن فرد است، یعنی توانش را باید با عملکرد وی بسنجیم و احتمالاً همیشه و برای همه موارد و آموخته‌ها انجام این ارزیابی امکان‌پذیر نیست. از طرف دیگر، یادگیری زبان با شناخت فرهنگ و مباحث منظورشناختی زبان مقصد در ارتباط است که در آنها همه موارد به خودی خود قابل آموزش نیست، و در نتیجه نمی‌توان همه مواردی که فرد بواسطه فعالیت‌های مختلف و حتی به روشی غیر از آموزش فراگرفته است را ارزیابی کرد. در نتیجه، بطور کلی نمی‌توان انتظار داشت که همه موارد فراگرفته شده از زبان مقصد قابل ارزیابی باشد.

³² Decroly & Buyse

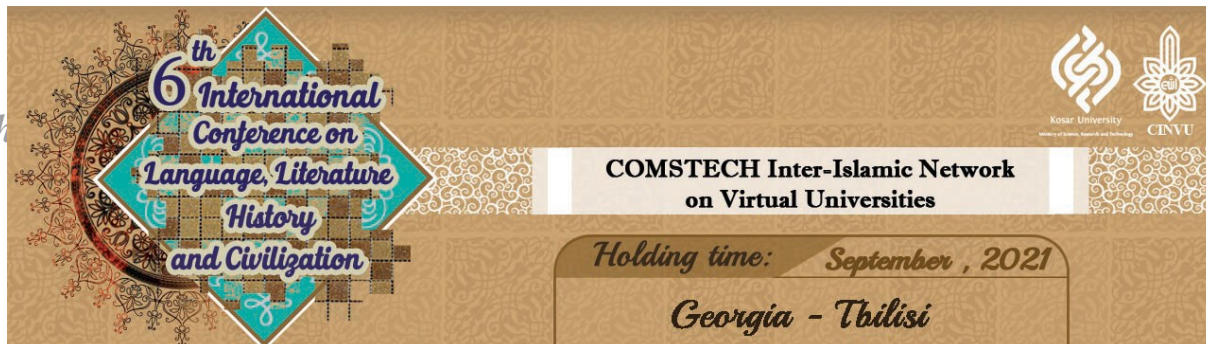
³³ Stéphanie Charton



در آخر باید توجه داشت که اصلی ترین مساله ای که باعث بوجود آمدن تمام چالش های مذکور می شود، کمبود آموزش در رابطه با ارزشیابی است. چراکه اگر فرد ارزیابی کننده و در اینجا مدرس زبان، به خوبی با مسائل مربوط به ارزشیابی آشنا باشد، بسیاری از این چالش ها خود بخود حل می شوند.

منابع

- بازرگان، عباس (۱۳۷۴). ارزیابی درونی دانشگاهی و کاربرد آن در بهبود مستمر کیفیت آموزش عالی "فصلنامه پژوهش و برنامه ریزی آموزش عالی، شماره ۱۱"
- دانایی فرد، حسن، الوانی، مهدی؛ و آذر، عادل، (۱۳۸۳)، روش شناسی پژوهش کمی در مدیریت رویکرد جامع: چاپ اول، جلد اول: تهران انتشارات صفار- اشراقی.
- Bachmann, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
- Baecher, R. E. (1982). *Language Proficiency Assessment: Issues and Definitions*.
- Bailly, D. (1997). *Didactique de l'anglais, 1: Objectifs et contenus de l'enseignement*. Paris: Nathan.
- Bass, M. (2012). Quelles en sont les conditions « impératives » d'une « vraie » évaluation ?, *Air Formation*, Septembre 2012
- Bolton, S. (1991). *Évaluation de la compétence de communication en langue étrangère*. Paris : Didier.
- Brown, J. D. (1996). *Testing in Language Programs*. Prentice Hall Regents.
- Butera, F., Buchs, C., & Darnon, C. (2011). *L'évaluation, une menace ?* Puf.
- Cadre européen commun de référence pour les langues (CECRL). (2001). Conseil de l'Europe. Strasbourg : Unité des Politiques linguistiques.
- Chomsky, N. (1981). On cognitive capacity. *Readings in philosophy of psychology*, 2, 305-323.
- Cronbach, L. J., & Furby, L. (1970). How we should measure "change": Or should we? *Psychological bulletin*, 74(1), 68.
- Davies, A. (1997). Introduction: The limits of ethics in language testing.
- DECROLY, O., & BUYSE, R., 1929, Introduction à la pédagogie quantitative. Éléments de statistique appliqués aux problèmes pédagogiques. Bruxelles : Maurice Lamertin éditeur
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment*. London and New York: Routledge.
- Henning, G. (1987). *A guide to language testing: Development, evaluation, research*.
- Morrow, C. K. (2018). Communicative language testing. *The TESOL encyclopedia of English language teaching*, 1-7.
- Nardi, P.M. (2003). *Doing survey research: A guide to quantitative methods*. Boston, MA: Allyn and Bacon.
- Neuman, W. L. (2007). *Basics of social research: Qualitative and quantitative approaches* (2nd ed.). Boston, MA : Allyn and Bacon.
- Patton, M.Q. (1981). *Creative Evaluation*. London: SAGE Publications.
- Pierce, C. S. (1877). The Fixation of Belief, *Popular Science Monthly*.
- Puren, C. (2001). La problématique de l'évaluation en didactique scolaire des langues. *Langues modernes*, 95(2), 12-29.
- Rapi, L., & Miconi, H. (2014). Theoretical and practical problems in designing tests for foreign language programs. *European Scientific Journal*, 10(29).
- Rorty, R. (1999). *Philosophy and social hope*. Penguin UK.
- Shohamy, E. G. (2001). *The power of tests: A critical perspective on the uses of language tests*. Pearson Education.
- Shohamy, E., Or, I. G., & May, S. (Eds.). (2017). *Language testing and assessment*. Springer.



- Spanou, c. (2017). *Les enjeux organisationnels des entretiens d'évaluation*.
- Tardieu, C. (2006). L'évaluation en langues : quelles perspectives ? *Recherches en didactique des langues et des cultures. Les cahiers de l'Acedle*, (2).
- Van de Kerkhove, A. (2017). Compte rendu de la conférence : 'Evaluation des acquis des élèves', STAPS de Sainte Etienne.