

بررسی آثار و تبعات حذف کنکور سراسری بر نظام آموزش عمومی زبان انگلیسی کشور:

دیدگاه صاحب نظران و معلمان زبان انگلیسی

پروانه شایسته فر (استادیار آموزش زبان انگلیسی، دانشگاه فرهنگیان، تهران، ایران، نویسنده مسئول)

parishayeste@yahoo.com

غلامرضا کیانی (دانشیار آموزش زبان انگلیسی، دانشگاه تربیت مدرس، تهران، ایران)

rezakiany@yahoo.com

منیره نوروزی (دانشجوی دکتری آموزش زبان انگلیسی، دانشگاه تربیت مدرس، تهران، ایران)

norouzimonireh@gmail.com

چکیده

اصلاحات آزمون محور اغلب به عنوان ابزاری راهبردی برای القا و پیشبرد تغییراتی مطلوب صورت می گیرند. با وجود این، ادبیات مرتبط حاکی از آن است که موفقیت، اغلب آن طور که امید آن می رود حاصل نمی گردد. دلایل چنین ناکامی هایی معمولاً در مراحل پایانی برنامه تحول یافته، یعنی ارزشیابی مؤخر بر تجربه بررسی گردیده و ارزشیابی حین تجربه همچنان مورد غفلت مانده است. در کشور ما، سیاست گذاری مربوط به اصلاح کنکور سراسری موج اصلاحاتی را در نظام دیرینه سنجش و پذیرش دانشجو به بار آورد. در حوزه زبان انگلیسی، این اصلاحات بر جایگزینی تدریجی بخش زبان انگلیسی کنکور با آزمون های سراسری زبان دبیرستان تأکید دارد. به دنبال چنین تغییری، ارزیابی تبعات و پیامدهای برنامه جدید الزامی به نظر می رسد. به همین منظور، برای جمع آوری داده از جمعی از ذی نفعان و ذی ربطان این تحول از مدل ترکیبی کیفی کمی استفاده شد. ابتدا، دیدگاه استادان، مدرسان و سرگروهان (۱۶ نفر) از طریق مصاحبه های نیمه ساختار یافته استخراج و سپس از ۴۰۶ معلم زبان انگلیسی نظر سنجی به عمل آمد. روایی ابزار محقق ساخته از طریق تحلیل عاملی تأییدی-AMOS و اکتشافی-SPSS انجام و پایایی آن نیز محاسبه شد. نتایج تحلیل محتوایی مصاحبه ها و تحلیل داده های کمی حاکی از آن است که بین نظرات مدرسان و

معلمان با نظرات سیاست‌گذاران تفاوت‌هایی مشاهده می‌شود، به‌ویژه، در مقولات عدالت در آزمون، اهداف آموزشی/یادگیری و طراحی، تدریس و برنامه‌ریزی. به‌طور کلی، نتایج گویای آن است که با توجه به شرایط فعلی آموزش و یادگیری زبان انگلیسی در دبیرستان‌های کشور و چالش‌های موجود، محتمل است که طرح تغییر، در شکل کنونی خود، به ابزاری نامناسب جهت پیشبرد اهداف ایدئال و بلندپروازانه تبدیل شود.

کلیدواژه‌ها: آثار و پیامدهای آزمون، آزمون‌های سراسری، اصلاحات آزمون‌محور، سابقه تحصیلی، نظام سنجش و پذیرش دانشجو

۱. مقدمه

در عصر حاضر، نیاز مبرم به پاسخگویی معتبر و تضمین کیفیت، موج اصلاحاتی را در نظام دیرینه سنجش و پذیرش دانشجو به بار آورده و توجه متخصصان آموزشی را به روش‌ها و شیوه‌های سنجش که ارتقا و بهبود سطح کیفی یادگیری را به ارمغان آورند، جلب نموده است. این امر نه تنها منجر به انتخاب یا طراحی آزمون‌های استاندارد^۱ شده بلکه اهمیت روزافزون این آزمون‌ها را دو چندان نموده است. در ایران، همچون بسیاری از کشورهای مشابه، مادامی که دست یافتن به جایگاه والای اجتماعی و تضمین شغلی منوط بر عملکرد مطلوب افراد در آزمون‌های ورودی دانشگاه‌ها باشد، این آزمون‌های سرنوشت‌ساز (پر وجه)^۲ بی‌شک تأثیر غیرقابل انکار خود را بر ارزش‌های اجتماعی، حسن رقابت و بر کل نظام آموزشی به‌جای خواهند گذاشت. در میان یافته‌های تحقیقاتی که اخیراً صورت گرفته، می‌توان شواهد بیشتری را در تأیید وجود آثار چنین آزمون‌هایی به‌ویژه در حوزه‌های تدریس و یادگیری یافت (برای مثال رجوع شود به دکر و بلت^۳، ۲۰۰۸؛ کیکوچی و برانه^۴، ۲۰۰۹؛ وال و هوراک^۵، ۲۰۰۸). با وجود آثار و پیامدهایی از این قبیل، یا به‌عبارتی قدرت

1. Standardized tests
2. High-stakes tests
3. Decker & Bolt
4. Kichuchi & Browne
5. Wall & Horak

تأثیرگذاری این آزمون‌ها بر حوزه‌های متفاوت آموزش و یادگیری (شوهامی^۱، ۱۹۹۳؛ وال^۲، ۲۰۰۰؛ وال و آلدسون^۳، ۱۹۹۳).

چندان جای تعجب نیست که سیاست‌گذاری‌های آزمونی با هدف اصلاح و ارتقای سطح کیفی آموزش و یادگیری و در راستای پاسخگویی معتبر بنیان نهاده شوند (دکر و بلت، ۲۰۰۸؛ چپمن و اسنایدر^۴، ۲۰۰۰؛ وال، ۲۰۰۰).

در سال‌های اخیر، مسئله پاسخگو بودن نظام‌های ارزشیابی، ضرورت توجه به اثربخشی و کارآمدی آن‌ها در بهبود و تعالی کیفیت تدریس و یادگیری متأثر از آن را بیش از پیش برجسته ساخته است. از این‌رو، ارزیابی تضمین کیفیت این نظام‌ها و بررسی تأثیر و پیامدهای آن‌ها امری است اجتناب‌ناپذیر. علی‌رغم این مهم، هنوز تأثیر و پیامدهای سیاست‌های اصلاح‌طلبانه آزمونی بر کیفیت آموزش و یادگیری از دیدگاه ذی‌ربطان و نقش‌آفرینان محیط‌های واقعی یادگیری و تدریس (مربیان و متریان) چندان مورد بررسی قرار نگرفته است. این در حالی است که مبانی نظری ارزیابی آزمون‌ها اکنون سنجش روایی آزمون را در چارچوب مفهومی گسترده‌تری در نظر می‌گیرد. طبق این چارچوب، علاوه بر سنجش روایی سازه‌ای، روایی صوری و روایی محتوایی، بررسی روایی آثار و تبعات یا اعتبارسنجی^۵ پیامدها و تبعات آزمون امری ضروری است (استولر^۶، ۱۹۹۴؛ کندی^۷، ۱۹۸۸؛ هنریکسن^۸، ۱۹۸۹). لذا، نباید از این نکته غافل بود که سنجش در مقیاس وسیع ملی می‌تواند پیامدهای متفاوتی را برای افراد ذی‌ربط و ارکان مختلفی که در لایه‌های متفاوت این برنامه معین دخیل‌اند در بر داشته باشد.

1. Shohamy
2. Wall
3. Wall & Alderson
4. Chapman & Snyder
5. Validation
6. Stoller
7. Kennedy
8. Henrichsen

هنگامی که وجوه و حساسیت‌های ویژه‌ای وجود دارد، علی‌الخصوص هنگامی که کاربری اصلی نتایج آزمون، تعیین‌کننده فرصت‌های آتی و اساس مجوز ورود به مراحل بالاتر یا دس‌یابی به فرصت شغلی باشد، تعداد عواقب و پیامدهای این آزمون به مراتب بیشتر خواهد شد. نگاهی کلی به اصلاح برنامه کنکور - محور موجود در ایران به خوبی گویای این مطلب است که پیامدهای حساس چنین تغییری را نمی‌توان بدون عواقب مهم آن برای برنامه درسی، ارزش‌های اجتماعی، و آموزش و یادگیری تصور نمود. پژوهش پیش رو در تلاش است از دیدگاه قشری از ذی‌ربطان نظام اصلاح‌شده سنجش و پذیرش دانشجو (همچون مدرسان و معلمان آموزش زبان انگلیسی) به بررسی تبعات و پیامدهای ناشی از بخش زبان انگلیسی این نظام تحول‌یافته بر آموزش و یادگیری زبان انگلیسی دبیرستان‌های ایران بپردازد. به عبارتی این پژوهش در صدد است روایی آثار و تبعات آزمون‌های نهایی زبان انگلیسی دبیرستان را بر حسب میزان اثربخشی آن‌ها در شکل‌دهی و بهبود کیفی تدریس معلمان، یادگیری دانش‌آموزان و برقراری اصل عدالت از دیدگاه افرادی که به‌ندرت در سیاست‌گذاری‌های اصلاحی آزمونی مشارکت داشته یا نظرشان شنیده می‌شود، بررسی نماید.

سیاست‌های بومی مرتبط با تصویب قانون جدید پذیرش دانشجو

از دهه پنجاه تا کنون، ورود به مرحله تحصیلات عالی به‌ویژه دانشگاه‌های دولتی کشور، همواره عرصه رقابت بوده است؛ رقابتی که پس از پیروزی انقلاب اسلامی ایران به دلیل عدم تناسب میان تعداد داوطلبان و ظرفیت دانشگاه‌های دولتی شدت یافته است. سالانه بالغ بر یک میلیون دانش‌آموز در آزمون‌های چندگزینه‌ای کنکور سراسری با یکدیگر به رقابت می‌پردازند. ارزشیابی از دانش شرکت‌کنندگان تنها بر پایه عملکردشان در این آزمون صورت گرفته و عملکرد مستمر دبیرستان یا سوابق تحصیلی داوطلبان در گزینش و تعیین سطح نهایی آنان بی‌اثر بوده است. در چنین شرایطی، اصلاح نظام منحصراً آزمون‌محور کنکور سراسری در دستور کار تحول بنیادین کشور قرار گرفت. در اسناد ملی اخیر کشور، به‌ویژه در بخش‌هایی

که به سنجش و ارزشیابی پیشرفت تحصیلی اختصاص یافته، نیاز فوری به تغییر ملاک‌ها و مقیاس‌های رایج در نظام منحصراً آزمون‌محور کنکور کاملاً مشهود است. علی‌الخصوص، رهنامه نظام تربیت رسمی و عمومی کشور و نیز سند ملی برنامه درسی حاوی بخش‌های جداگانه‌ای پیرامون سیاست‌های آزمون‌های هستند که پیشنهاد تغییر نظام آزمون‌محور وزارت‌خانه‌های آموزش و پرورش و علوم، تحقیقات و فناوری به‌طور ضمنی در آن‌ها مطرح شده است. نکته‌ای که باید به آن اشاره کرد این است که گرچه طی دو دهه اخیر گزارش‌های تأمل‌برانگیزی درباره ضعف نسبی نظام منحصراً آزمون‌محور ارائه شده، پیشنهاد تغییر نظام آزمون‌های موجود برای ورود به دانشگاه تا قبل از سال ۱۳۸۱ به‌طور رسمی مورد بررسی قرار نگرفته بود. این پیشنهاد را اولین بار صاحب‌نظران و پژوهشگران سنجش و متخصصان آموزشی در نشستی سه‌روزه با عنوان «سمینار بررسی روش‌ها و مسائل آزمون‌های ورودی دانشگاه‌ها» مطرح کردند در مهرماه ۱۳۸۱ در دانشگاه اصفهان برگزار شد.

نتیجه مباحث و تحلیل‌های صورت‌گرفته درباره کنکور در این نشست، تقدیم طرح نهایی چگونگی پذیرش دانشجو برای ورود به دانشگاه‌ها و مؤسسات آموزش عالی به مرکز پژوهش‌های مجلس شورای اسلامی بود. پس از بررسی‌ها و پژوهش‌های لازم (برای مثال، قلی‌پور و آقابزرگی، ۱۳۸۴؛ کیا و بزرگی، ۱۳۸۵) پیامد سیاستی مجلس در این مورد تصویب قانون پذیرش دانشجو در دانشگاه‌ها و مراکز آموزش عالی کشور (سال ۱۳۸۶) بود. قانون مصوب سال ۱۳۸۶ که شامل شش ماده و شش تبصره است و به تأیید شورای نگهبان نیز رسیده مقدمه‌ای برای سیاست‌ها و اعمال وابسته به قراین کشور است. طبق این قانون، سابقه تحصیلی و نتایج عملکرد چهارساله در امتحانات سراسری دبیرستان و پیش‌دانشگاهی (همچون امتحانات فعلی سال سوم متوسطه) که به‌صورت نهایی برگزار می‌گردند، به‌تدریج جایگزین عملکرد در آزمون‌های چندگزینه‌ای کنکور سراسری می‌شود که در حال حاضر سالی یک بار سازمان سنجش برگزار می‌کند.

به نظر سیاست‌گذاران بومی، این سیاست جدید مطابق با کاربرد پاسخگویانه آزمون‌ها (کیامنش، ۱۳۷۹؛ ناگی^۱، ۲۰۰۰) است که در مقیاس وسیع، می‌تواند تأثیرات مثبتی را درباره دانش، انگیزه، تلاش فردی، تسلط بر مهارت‌ها و نیز عملکرد دانش‌آموزان به دنبال داشته باشد. در بعد تعلیمی، این آزمون‌ها به معلمان کمک می‌کند نیازها و نقاط قوت و ضعف دانش‌آموزان را تشخیص داده و آموزش اثرگذارتری ارائه نمایند و از طرفی، آموزش را هم‌تراز استانداردهای آموزشی قرار داده و در کارشان به پیشرفت حرفه‌ای قابل ملاحظه‌ای دست یابند. به نظر می‌رسد در زمان تصویب قانون جدید سنجش پذیرش دانشجو، سیاست‌گذاران از نگرش کلی که معتقد است راه اصلاح و بهبود آموزش و یادگیری از تغییر آزمون‌ها می‌گذرد (رک، چنگ و همکاران^۲، ۲۰۱۱؛ هارگریوز و همکاران^۳، ۲۰۰۲؛ وال، ۲۰۰۰) الهام گرفته‌اند.

با این حال سیاست‌گذاری‌های آموزشی در کشورهای متفاوت نشان می‌دهد که تحولات آموزشی همیشه چندان هم موفق نبوده و گاه پیامدهای منفی (روانشناختی و آموزشی) سابق در شرایط جدید همچنان به قوت خود باقی مانده‌اند. ظاهراً عامل اصلی که به این مسئله غامض دامن می‌زند غفلت در اتخاذ ارزشیابی بهنگام و مناسب، تأخیر در ارزیابی و نیز عدم توجه به تحلیل پیامدها و تبعات نظام اصلاح‌یافته است. در کشور ما نیز رویارویی با چالش‌های نظام جدید سنجش و پذیرش دانشجو در مقیاس وسیع کنونی آن، کاهش مشکلات احتمالی چنین برنامه‌ای، افزایش مزایا و پتانسیل‌های پیش‌بینی‌شده در مرحله سیاست‌گذاری آن و پر کردن خلأ بین سطوح سیاست و عمل در این برنامه، همه و همه مستلزم این است که این تغییر علاوه بر مراحل پایانی خود در مراحل آغازین و اجرایی نیز مورد بررسی و بازنگری شفاف قرار گیرد. به بیان دیگر، از آنجا که ممکن است بخشی از اطلاعات مهم و مفید مربوط به اصلاح نظام کنکور محور سنجش و پذیرش دانشجو در مسیر سیاست به عمل، یعنی مسیر سیاست‌گذاری - اجرا - عمل/تکمیل کم‌رنگ گردد، اتخاذ روند

1. Nagy

2. Cheng et al.

3. Hargreaves et al.

ارزیابی از همان مراحل آغازین این برنامه امری ضروری و حتمی به نظر می‌رسد تا از خسارت ناشی از برنامه ارزشیابی نشده و خام که منجر به برنامه‌ای ناکارآمد می‌گردد، پیشگیری شود.

۲. پیشینه پژوهش

روش‌های سنجش و اندازه‌گیری سنتی رایج، مهارت‌ها و اجزای متفاوت زبان همچون دانش واژگان و دستور زبان را به صورت گسسته می‌سنجند، و نه به صورت اجزا و مهارت‌های مرتبط و درهم‌تنیده و کاربردی. این روش‌ها از بعد اجتماعی کاربرد زبان که به شدت مورد تأکید رویکرد ارتباطی زبان^۱ قرار گرفته، غافل شده‌اند. در این رویکرد، استفاده از زبان به عنوان ابزار ارتباطی در محیط‌های اجتماعی، تنها یک بعد از ابعاد اجتماعی دانش زبانی است. به همین دلیل، کین^۲ (۲۰۱۲) و مسیک^۳ (۱۹۸۹) معتقد به وجود ابعاد اجتماعی برای ابزارهای سنجش و اندازه‌گیری دانش زبانی، همچون آزمون‌ها هستند. آن‌ها می‌گویند برای درک کامل‌تر ابعاد اجتماعی آزمون‌های زبان می‌بایست علاوه بر کاربردهای اجتماعی این آزمون‌های خاص، پیامدهای اجتماعی آن‌ها را نیز مورد توجه و بررسی قرار داد. در بعد کاربردی، یکی از متداول‌ترین کاربری‌های آزمون‌های زبان مربوط به نقش آن‌ها به عنوان ابزارهای کنترل ورود به مرحله بالاتر، برنامه، شغل یا موقعیتی متفاوت‌تر است. آزمون‌هایی که با این هدف به خصوص کاربری ویژه دارند، اغلب توسط سیاست‌گذاران و مقامات رده‌بالای آموزشی تبیین و مطرح می‌شوند و متعاقباً، پیامدهایی جدی برای افراد و نهادهایی که در این تصمیم‌گیری دخالتی نداشته، خواهند داشت و بی‌شک بر برنامه‌های آموزشی مرتبط تأثیر خواهند گذاشت. برای مثال، از آن‌جا که دست‌یافتن به جایگاه اجتماعی والا و تضمین شغلی در ایران با دسترسی به تحصیلات عالی و حرفه‌ای شدن ارتباط مستقیمی دارد، آزمون‌های ورودی (همچون کنکور) نه تنها برای

1. Communicative Competence Approach

2. Kane

3. Messick

نظام آموزشی کشور پیامدهایی در بردارند، بلکه ارزش‌های اجتماعی و حس رقابت داوطلبان را نیز تحت‌الشعاع قرار می‌دهند. وجود همین آثار و تبعات اجتماعی مهم بر لایه‌های متفاوت جامعه است که ضرورت تأمل بیشتر بر ابعاد اجتماعی آزمون‌ها را دو چندان می‌نماید.

از منظر روایی آزمون، دیدگاه مک‌نامارا و روور^۱ (۲۰۰۶) با تئوری کلاسیک مسیک (۱۹۸۹) مبنی بر ضرورت ارزیابی آزمون از بعد اجتماعی آن همخوانی دارد. یکی از جنبه‌های حائز اهمیت در این مدل روایی‌سنجی، پرداختن به آثار و تبعات اجتماعی آزمون بر جامعه کاربر آن است. بررسی روایی آزمون با توجه به ابعاد اجتماعی آن اهمیت دو نکته را برجسته می‌سازد؛ اول اینکه چگونه ماده مورد آزمون (سازه و سنجه آزمون) توسط رویکردها، نظرات و استانداردهایی که ریشه در ارزش‌های اجتماعی و فرهنگی جامعه داشته تبیین و طراحی می‌شود. دوم اینکه چگونه آزمون‌ها می‌توانند برای افراد و جوامع کاربر یا محیط‌های اجتماعی که در آن‌جا مورد استفاده قرار می‌گیرند، پیامدهای واقعی داشته باشند. پژوهش‌های اخیر با الهام از مدل کلاسیک مسیک، ضمن برجسته‌سازی ارزش‌های فرهنگی و اجتماعی که شالوده سازه‌های آزمون‌ها را تشکیل می‌دهند، به بررسی پیامدهای پرداخته‌اند که به دنبال کاربرد آزمون‌ها در محیط‌های واقعی‌اند (برای مثال، بکمن و پالمر^۲، ۲۰۱۰؛ چنگ، ۲۰۰۵؛ چنگ و سان^۳، ۲۰۱۵؛ شوهای و همکاران، ۱۹۹۶؛ وال و هوراک، ۲۰۰۸، ۲۰۱۱).

مضافاً اینکه، این پژوهش‌ها با بیان یکی دیگر از زیرساخت‌های چارچوب نظری مدل کلاسیک مسیک، یعنی شرط عدالت در آزمون به‌عنوان یکی از پیش‌شرط‌های اصلی برای احراز روایی آزمون، اهمیت و جامع بودن این مدل را بیشتر می‌نمایانند. در این مدل علاوه بر سنجش روایی سازه، روایی محتوا، روایی صوری، و روایی آثار و تبعات یک آزمون، میزان احراز عدالت اجتماعی توسط آن آزمون نیز بررسی می‌گردد

1. MacNamara & Roever

2. Backman & Palmer

3. Cheng & Sun

(رک. کامیلی^۱، ۲۰۰۶؛ کین، ۲۰۰۲؛ مک‌نامارا و روور، ۲۰۰۶). مک‌نامارا و روور (۲۰۰۶) اذعان داشته‌اند که شرط عدالت اجتماعی در نظام آزمون‌محور تنها از طریق اتخاذ رویکرد صحیح و مناسب سنجش، طراحی دقیق و محاسبه‌شده روش‌های سنجش، برنامه‌ریزی دقیق و آگاهانه برای ساخت یا انتخاب ابزار مناسب سنجش، ایجاد شرایط برابر برای همه آزمون‌دهندگان و علاوه بر این‌ها، انجام قضاوت منصفانه، علمی، منطقی و عاری از سونگری و برداشت‌های شخصی در مورد عملکرد آزمون‌دهنده‌ها، برقرار می‌گردد. از این زاویه که بنگریم، می‌بینیم که برای داشتن تصویری جامع از اعتبارسنجی آزمون‌های زبان پیش‌شرط عدالت در آزمون نیز پیش‌شرطی ضروری است.

ارزیابی اصلاحات آزمون‌های زبان‌های خارجی: بررسی آثار آزمون‌های سرنوشت‌ساز/پر وجه

طی پانزده سال اخیر پژوهش‌های قابل توجهی درباره آثار و تبعات آزمون‌های زبان که در ادبیات رایج آزمون‌سازی بازشویی آزمون^۲ نامیده می‌شود، صورت گرفته است (برای مثال، پن و نیوفیلدز^۳، ۲۰۱۲؛ چنگ و همکاران، ۲۰۰۴؛ شوهای و همکاران، ۱۹۹۶، مونوز و آلوارز^۴، ۲۰۱۰؛ واتاناب^۵، ۲۰۰۱؛ وال و آلدرسون، ۱۹۹۳؛ وینکه^۶، ۲۰۱۱). این پژوهش‌ها معمولاً یا در زمینه تأثیر آزمون‌های زبان انگلیسی بر قلمرو آموزش انجام شده (چپمن و اسنایدر، ۲۰۰۰، وال و هوراک، ۲۰۱۱) یا بر قلمرو یادگیری و بافت مدرسه یا حتی فراتر از آن (اسپرات^۷، ۲۰۰۵؛ چنگ و همکاران، ۲۰۱۱؛ دکر و بلت، ۲۰۰۸؛ شوهای، ۱۹۹۳). شواهد موجود حاکی از آن است که اکثر پژوهش‌هایی که با محوریت قدرت بازشویی آزمون انجام شده‌اند بیشتر بر آزمون‌هایی که برای مدت‌های مدیدی در جامعه هدف مورد استفاده قرار گرفته‌اند

1. Camilli
2. Test Washback
3. Pan & Newfields
4. Munoz & Alvarez
5. Watanabe
6. Winke
7. Spratt

متمرکز شده و تنها تعداد انگشت‌شماری از این پژوهش‌ها به بررسی بازسویی آزمون‌های اصلاح‌شده یا آزمون‌های تازه طراحی‌شده اختصاص یافته‌اند. حتی فراتر از این، هنوز در مورد مبانی سیاستی و سیاست‌گذاری آزمون‌ی یا ارزش‌های سیاسی و اجتماعی که سنگ بنای اصلاحات آزمون‌محور را تشکیل می‌دهند، پژوهش‌چندانی انجام نشده است.

پیشینه اندک اما مفید موجود نشان می‌دهد که پژوهشگران اخیراً با الهام از مبانی اساسی حیطه نوآوری‌های آموزشی و انتقال مفاهیم این حوزه عمومی به حوزه تخصصی آزمون‌سازی در زبان، پای ارزیابی اصلاحات و تحولات آزمون‌ی را به عرصه زبان‌های خارجی باز کرده و بر ارزیابی مراحل سیاست‌گذاری آزمون، فرآیند اجرایی آزمون و همه مراحل مربوط به عملیاتی‌کردن آن تأکید کرده‌اند (برای مثال، چنگ، ۲۰۰۴؛ وال و هوراک، ۲۰۱۱؛ هنریکسن، ۱۹۸۹). برای مثال، طبق یافته‌های وال (۲۰۰۰، ۱۹۹۹)، وال و هوراک (۲۰۱۱) و چنگ (۲۰۰۴)، تبعات و قدرت تأثیرگذاری آزمون اصلاح‌یافته یا جدیدالاجرا می‌بایست از همان مراحل آغازین این برنامه اصلاح‌طلبانه بررسی و نظارت شود و نتایج حاصل از این مرحله باید به‌گونه‌ای منعکس شود که اطلاعات مورد نیاز «برای تصمیم‌گیری درباره تناسب آزمون با برنامه درسی، تناسب روش‌های آزمون، تخصیص منابع کافی، طراحی برای آموزش افراد ذی‌ربط و ایجاد تغییرات در سیاست خرد و کلان مرتبط با آزمون» (وال، ۲۰۰۰، ص. ۵۰۷) فراهم گردد.

مشابه این نتایج را کندی (۱۹۸۸) گزارش کرد. او با انتقال برخی از مفاهیم تئوری نوآوری‌های آموزشی به حوزه ارزیابی نظام آموزشی، ابتدا به توصیف مدل سه‌مرحله‌ای خود پرداخت که شامل ارزیابی مرحله سیاست‌گذاری، مرحله فرآیند اجرایی/عملیاتی‌کردن و مرحله نتایج و برون‌داد است. سپس به‌طور خاص بر این نکته تأکید ورزید که در ارزیابی طرح تغییر نباید تنها به ارزیابی برون‌دادها و بازده نهایی برنامه تغییر یافته (مثلاً محصول یا رفتار خاصی) بسنده کرد بلکه ارزیابی دو مرحله

دیگر یعنی مرحله سیاست‌گذاری و مرحله روند اجرایی و گذار برنامه، قبل از مرحله نهایی نیز از الزامات ارزیابی کامل و اصولی است.

چنین به نظر می‌رسد که بین مدل ارزیابی سه مرحله‌ای کندی با مدل هنریکسن (۱۹۸۹)، معروف به مدل ترکیبی انتشار تغییر و نوآوری^۱، وجه تشابه وجود دارد. ترکیبی هنریکسن در واقع بیانگر چگونگی اشاعه و نشر یک تغییر یا نوآوری به نظام و بافت خود می‌باشد. در این مدل، هنریکسن با تکیه بر سه مؤلفه اصلی نظام آموزشی اصلاح شده به توصیف فرآیند تغییر و نوآوری می‌پردازد. این مؤلفه‌ها عبارت‌اند از بافت و بستر تغییر و نوآوری^۲، فرآیند اشاعه و نفوذ تغییر به بافت مورد نظر^۳، بروندادها، نتایج و پیامدهای تغییر و نوآوری^۴. هر یک از این مؤلفه‌ها از چندین زیرمؤلفه یا زیرمجموعه تشکیل شده که هر یک، در جای خود، در موفقیت یا شکست پروژه تغییر و نوآوری نقش مهمی را ایفا می‌کند. به‌عنوان مثال، در ارزیابی مرحله دوم یا مرحله فرآیند اشاعه و نفوذ تغییر، زیرمجموعه‌هایی همچون منبع صدور تغییر (مثلاً عوامل سیاست‌گذار و برنامه‌ریزان، قوانین و سیاست‌های دولت)، نظام کاربران (مثلاً معلمان و دانش‌آموزان) کانال ارتباطی (یعنی مجاری مخابراتی پیام تغییر از منبع صادرکننده آن به نظام کاربران) از جمله عوامل بسیار مهم بوده که باید مورد بررسی و ارزیابی دقیق قرار گیرند.

بر اساس مدل ترکیبی هنریچسن، به‌محض مخابراتی پیام تغییر به کاربران و گیرندگان آن، نگرش به تغییر در آنان شکل می‌گیرد و همین نگرش، مثبت یا منفی، مبنای تصمیم‌گیری برای پذیرش یا رد تغییر مخابراتی شده است. بنابراین اگر بنای یک پروژه اصلاح طلبانه آزمونی بر این قرار گیرد که به قهقرا نرود و طعم شکست را نچشد لازم است که همه زیرمؤلفه‌ها و عوامل دخیل در این نظام آزمونی در همان ابتدای اصلاح و تغییر، هرچه زودتر، مورد بررسی و ارزیابی منظم و مستمر قرار

1. Hybrid Diffusion Model (HDM)
2. Antecedents
3. Process
4. Consequences

گیرند. وال (۲۰۰۰) خاطر نشان می‌سازد که در چنین فرآیند ارزیابی محوری باید افرادی از همه اقشار ذی‌ربط و ذی‌نفع نظام آزمون‌محور شرکت داده شوند تا بتوان هم از نگرش آن‌ها نسبت به پروژه تغییر مطلع شد و هم به کارآیی، مفید بودن، عملی و کاربردی بودن اصلاحات مورد نظر پی برد. از این‌رو، بررسی ارزش‌های اجتماعی نهفته در پروژه تغییر آزمون و همچنین ارزیابی تبعات و تأثیرات این تغییر بر حوزه‌های آموزش و یادگیری از همان ابتدای امر می‌بایست با رعایت ملاحظات نگرشی ذی‌ربطان، ذی‌نفعان و ذی‌صلاحان این نظام انجام شود.

به‌دنبال اصلاحاتی که در آزمون‌های رایج تافل دنیا و تبدیل آن‌ها به تافل اینترنتی^۱ به‌جای تافل سنتی کاغذی یا کامپیوتری رخ داد، وال و هوراک (۲۰۰۸، ۲۰۱۱) در پژوهشی طولی به بررسی تبعات این تغییر مهم در حوزه آموزش و آمادگی برای این آزمون‌های سرنوشت‌ساز و پر وجه پرداختند و برای این منظور از ترکیبی از دو مدل فوق استفاده کردند. این پژوهش در قالب چهار فاز اصلی و در شش کشور اروپای مرکزی و شرقی انجام شد. در آخرین فاز این پژوهش (فاز ۴، سال ۲۰۱۱)، وال و هوراک شاهد تغییرات چشمگیری در میزان تمرکز مدرسان بر مهارت‌هایی بودند که توسط این آزمون سنجیده می‌شوند. آن‌ها در پایان گزارش خود، به تشریح نقش آزمون جدید و محتوای آن در تغییر روش‌های تدریس و یادگیری که در کلاس‌های آمادگی این آزمون رخ می‌داد، پرداختند.

در تحقیق دیگری، چنگ (۲۰۰۴) به‌دنبال بررسی این بود که بیند آیا تغییر امتحانات زبان انگلیسی پایانی مدارس در هنگ‌کنگ (یعنی آزمون صلاحیت زبان انگلیسی HKCEE-Hong Kong Certificate Examination) از آزمون‌های سنتی به آزمون‌هایی که بیشتر فعالیت‌محور و ارتباطی باشند بر نگرش و ادراک معلمان زبان انگلیسی نسبت به آموزش و یادگیری زبان انگلیسی تأثیرگذار خواهد بود یا نه. چنگ متوجه تشابه نگرش و ادراک معلمان با نگرش و گرایش سیاست‌گذاران و برنامه‌ریزان آزمون صلاحیت زبان انگلیسی هنگ‌کنگ شد و فعالیت‌های معلمان را در

1. TOEFL iBT

راستای سیاست‌گذاری‌های آنان دید، اما پی برد که تغییراتی که به دنبال این تغییر در نظام آموزش زبان انگلیسی دبیرستان‌های این کشور ایجاد شده تغییرات چندان عمیق و ریشه‌داری نبوده بلکه سطحی و ناچیز به نظر رسیدند. اگرچه آزمون‌های جدید با هدف بهینه‌سازی و اصلاح آموزش و یادگیری زبان انگلیسی هنگ‌کنگ طراحی شده بودند، میزان این تغییر محدود بود. تدریس همانند قبل، فقط در راستای آزمون بوده و معلمان همچون سابق همچنان آزمون‌محور و آزمون‌گرا بودند.

نتایج پژوهش پزارو^۱ (۲۰۱۰) در زمینه تأثیر آزمون‌های انگلیسی (ETs) بر برنامه درسی، مطالب درسی، روش تدریس، و همچنین نگرش و درک معلمان زبان انگلیسی دوره متوسطه در اسپانیا نشان داد که این آزمون‌ها به طور قطع بر برنامه درسی، مطالب درسی و روش تدریس تأثیرگذارند. وی در پژوهش خود به بررسی دیدگاه و نظرات معلمان در خصوص بخش‌های اضافه‌شده در آزمون‌های جدید زبان انگلیسی، یعنی دو مهارت شنیداری و گفتاری، نیز پرداخت. یافته‌های پزارو بیانگر دیدگاه مثبت معلمان به اضافه‌شدن این دو مهارت مهم به آزمون‌های جدید بود؛ چراکه آنان به پرداختن صرف به مهارت‌های کتبی اعتقاد نداشتند.

شواهد به دست آمده از پیکره پژوهش‌های ذکر شده در این بخش و مطالعات مشابه موجود در ادبیات مرتبط، اهمیت و ضرورت بررسی نگرش معلمان نسبت به تأثیر ماهیت، محتوا، قالب یا چگونگی اجرای آزمون‌ها بر فعالیت‌ها و رفتارهای تدریس و یادگیری آن‌ها را آشکار می‌سازد. به‌طور ویژه، آنچه در این پیکره پژوهشی مشهود است این است که معلمان به‌وضوح آثار و تبعات آزمون‌های ملی را در حیطه‌های عملکردی زیر درک می‌کنند: فعالیت‌ها و تمرین‌های ارائه‌شده در کلاس، مدیریت زمانی، محتوا و مطالب کمک‌درسی و آمادگی برای آزمون، طرح درس و نحوه ارزیابی یادگیری دانش‌آموزان.

پژوهش حاضر به دنبال بررسی روایی آثار و تبعات نظام اصلاح‌شده سنجش و پذیرش دانشجو در بستر اجتماعی نظام آموزش زبان انگلیسی دبیرستان‌های کشور

1. Pizarro

است. به عبارتی، به دنبال بررسی پیامدهای امتحانات نهایی زبان انگلیسی وزارت آموزش و پرورش از بعد ادراکی و نگرشی مدرسان آموزش زبان انگلیسی کشور است. با الهام از دو مدل یاد شده در حوزه ارزیابی پروژه تغییر (مدل کندی و مدل هنریکسن)، ضرورت بررسی نظام متحول شده سنجش و پذیرش دانشجو در ایران از دیدگاه کاربران اصلی آن بیش از پیش احساس می‌شود. شاهد محکم دیگری برای این ادعا را می‌توان از هیوز^۱ (۱۹۹۳) نقل قول کرد که معتقد است آزمون‌های سرنوشت‌ساز و پر وجه در وهله اول نگرش و دیدگاه کاربران آن را تحت تأثیر قرار می‌دهد و سپس عکس‌العمل، رفتار و یا فعالیت‌های آنان را تحت‌الشعاع قرار خواهد داد؛ چرا که رفتار برخاسته از نگرش و تفکر است.

۳. روش پژوهش

۳.۱. سؤال پژوهش

تعبیر مدرسان آموزش زبان انگلیسی کشور (استادان و معلمان) از امتحانات نهایی دبیرستان برای پذیرش دانشجویان به دانشگاه‌ها چیست؟
به عبارتی دیگر:

از دیدگاه ادراکی مدرسان آموزش زبان انگلیسی کشور، برنامه جدید سنجش و پذیرش دانشجو چگونه تسهیل‌دهنده یا مانع سیاست‌های مورد نظر این برنامه خواهد بود؟

۳.۲. جامعه آماری و شیوه نمونه‌گیری

اعضای هیئت علمی دانشگاه‌های وابسته به وزارت علوم، تحقیقات و فناوری و معلمان زبان انگلیسی شاغل در وزارت آموزش و پرورش در دو فاز این تحقیق شرکت کردند. در ابتدا محققان تصمیم گرفتند به ملاقات و مصاحبه با آن دسته از افرادی که درگیر نظام جدید سنجش و پذیرش دانشجو بوده و علاقمندی خود را به شرکت در این پژوهش نشان داده بپردازند. از این‌رو، با هشت استاد دانشگاه (۷ مرد

و یک زن) از هشت دانشگاه از طریق پست الکترونیکی یا تلفن تماس حاصل شد. مبنای انتخاب آن‌ها علاوه بر رضایت‌مندی و صلاحیت، تدریس در سطوح کارشناسی بود که داوطلبان آن‌ها بر اساس معیارهای جدید پذیرفته شده بودند. شرکت‌کنندگان به‌جز یک نفر دارای مدرک دکترای تخصصی از دانشگاه‌های بین‌المللی (مثل دانشگاه‌های ایالات متحده آمریکا و بریتانیا) یا دانشگاه‌های دولتی ایران بودند. آن‌ها در حوزه‌های مختلفی از جمله پژوهش در زمینه قدرت بازشویی آزمون‌ها، ارزیابی و یادگیری زبان دوم گرفته تا زبان‌شناسی و تربیت معلم تخصص داشتند و تجربه کاری آن‌ها از سه تا بیست و چهار سال بود. به‌منظور محرمانه‌ماندن هویت شرکت‌کنندگان، از کدهای UP1 (استاد دانشگاه-۱) تا UP8 (استاد دانشگاه-۸) استفاده شد. علاوه بر اساتید دانشگاه، از هشت سرگروه زبان انگلیسی شاغل در وزارت آموزش و پرورش (۴ مرد و ۴ زن)، با سابقه کاری ۵ تا ۳۰ سال، برای شرکت در فاز کیفی تحقیق دعوت به عمل آمد. این معلمان که دارای مدرک کارشناسی یا کارشناسی ارشد بودند، در سه استان تهران، اصفهان و مازندران مشغول به تدریس بودند. برای داده‌های مربوط به این معلمان از کدهای ET1 (معلم زبان انگلیسی-۱) تا ET8 (معلم زبان انگلیسی-۲) استفاده شد.

در بخش دوم پژوهش، برای دستیابی به جمعیت آماری بیشتر با استفاده از دو روش نمونه‌گیری خوشه‌ای^۱ و نمونه‌گیری در دسترس^۲ نمایندگان از معلمان زبان انگلیسی تهران، اصفهان و مازندران انتخاب شد. دلیل انتخاب این سه استان پرجمعیت، علاوه بر اهمیت سیاسی، اقتصادی و آموزشی، نحوه اطلاع‌رسانی و انتقال اطلاعات به معلمان یا به‌قول هنریکسن (۱۹۸۹) و وال (۲۰۰۰) کانال ارتباطی ارسال پیام تغییر به کاربران بود (محققان با سرگروه‌های این مناطق در ارتباط و از فعالیت‌های آنان مطلع بودند). در مقایسه با استان‌ها و شهرهای دورتر، در این سه استان پیام‌ها از طریق ابزار ارتباط جمعی، از جمله وبسایت‌ها، سمینارها، نشریه‌ها،

1. Cluster Sampling Procedure
2. Convenient Sampling Procedure

یا نمایندگان استانی، یا از طریق کانال‌های ارتباطی بین مدیر مدرسه، همکاران و دانش‌آموزان در اسرع وقت و واضح مخابره می‌شود.

در مجموع ۴۶۰ پرسش‌نامه بین معلم‌ها توزیع شد، اما فقط ۴۰۶ پرسش‌نامه (۸۹٪) تکمیل و بازگردانده شد. نمونه‌هایی شامل ۲۷۸ (۶۸٪) معلم زن و ۱۲۸ (۳۲٪) معلم مرد، با محدوده سنی ۳۰ تا ۵۵ سال بود. از کل این جمعیت، ۲۹۸ (۷۳٪) نفر مدرک کارشناسی، ۱۰۵ (۲۶٪) نفر مدرک کارشناسی ارشد و ۴ نفر (۱٪) دارای مدرک دکترای تخصصی بودند.

۳.۳. رویه اجرا و ابزار مورد نیاز

این پژوهش به روش تلفیقی کیفی کمی انجام شد و از آنجایی که جزو اولین پژوهش‌های انجام‌شده پس از اجرای نظام جدید سنجش و پذیرش دانشجو است، پژوهشگران ابزار لازم برای جمع‌آوری داده‌های مورد نیاز خود را در دسترس نداشتند، لذا ناگزیر باید به طراحی و تهیه ابزار لازم می‌پرداختند. این اقدام با اتکا به منابع معتبر موجود و نیز منابع زیر انجام شد.

۳.۳.۱. پروتکل مصاحبه با استادان و سرگروهان زبان انگلیسی

برای جمع‌آوری داده‌های فاز کیفی، از مصاحبه فردی استفاده شد. در طراحی و ساخت پروتکل مصاحبه منابع زیر مورد استفاده قرار گرفت:

- قانون پذیرش دانشجو در دانشگاه‌ها و مراکز آموزش عالی کشور (معروف به قانون حذف کنکور)؛
- اسناد بالادستی کشور، از جمله سند ملی برنامه درسی (۲۰۱۰) و نقشه جامع علمی کشور (۲۰۰۹)؛
- مستندات مکتوب و دستورالعمل‌های مرتبط با حذف کنکور سراسری، و دانش و تجربه شخصی محققان حاضر پیرامون نظام کنکور محور و نظام مبتنی بر آزمون‌های سراسری دبیرستان و آثار و تبعات دیرینه هر دو نظام.

مصاحبه‌های این پژوهش با دو هدف انجام شد: ۱. دستیابی به نگرش، دیدگاه و تعابیر مدرسان آموزش زبان انگلیسی از نظام جدید؛ ۲. طراحی و ساخت ابزاری معتبر برای جمع‌آوری داده‌های فاز کمی.

در فاز کیفی، ابتدا پروتکل مصاحبه‌ها بر مبنای سوژه‌های زیر تهیه شد: تبعات برنامه جدید برای حوزه آموزش و یادگیری زبان انگلیسی، رعایت اصل برابری در سنجش و پذیرش و برداشت کلی از کل برنامه جدید. قبل از انجام مصاحبه‌ها روایی صوری^۱ و روایی محتوایی^۲ پروتکل مصاحبه توسط یک متخصص آزمون‌ساز و یک مدرس زبان انگلیسی بررسی شد. مصاحبه‌های انجام‌شده از نوع نیمه‌ساختاریافته بوده و زمان مصاحبه‌ها بین ۵۰ تا ۷۵ دقیقه به طول انجامید. مصاحبه‌کننده همزمان با ضبط صدای پاسخ‌دهندگان، یادداشت‌برداری نیز می‌کرد. البته چهار شرکت‌کننده ترجیح دادند که پاسخ‌های خود را به صورت کتبی به پژوهشگر بدهند.

۳.۲. مجموعه پرسش‌نامه آثار و تبعات تغییر آزمون: نسخه معلمان

برای دستیابی به نظرات و داده‌های ادراکی بیشتر، پرسش‌نامه آثار و تبعات تغییر (معرفی آزمون‌های جدید در این پژوهش) طراحی و ساخته شد و سپس بین ۴۶۰ معلم زبان توزیع شد. برای ساخت این ابزار، بررسی پیشینه موجود در اولویت قرار گرفت تا مدل‌های تغییر/اصلاح و ارزیابی آزمون‌های سرنوشت‌ساز جستجو و مورد بررسی قرار گیرند. در این پژوهش، پژوهشگران به بیش از یک صد مقوله رفتاری دست یافتند (مثل شدت استفاده از مطالب و مواد آزمون‌ها علی‌الخصوص به‌هنگام نزدیک‌شدن به زمان آزمون؛ افزایش توجه معلمان به مهارت‌ها و بخش‌هایی که توسط این آزمون‌ها سنجیده می‌شوند و مواردی از این قبیل). تحلیل و چینه‌ساز معنادار این داده‌ها منجر به پیدایش یک مدل فرضی برای بررسی تأثیر آزمون اصلاح‌شده یا تغییر یافته شد. این مدل حوزه‌هایی را که مکرراً به‌طور چشمگیری تحت تأثیر آزمون‌های پروژه قرار می‌گیرند، آشکار ساخت که عبارت‌اند از: ۱. برنامه درسی

1. Face validity

2. Content validity

(چنگ، ۱۹۹۸؛ هیوز، ۱۹۹۳)؛ ۲. نقش هدایتی و مربی‌گری معلمان در آماده‌سازی آزمون‌دهندگان (آلدرسون و همپ-لیونز، ۱۹۹۶؛ واتاناب، ۲۰۰۰)؛ ۳. محتوای مطالب ارائه‌شده (چنگ، ۱۹۹۸؛ وال و آلدرسون، ۱۹۹۳)؛ ۴. شیوه‌ها و روش‌های ارزشیابی یادگیرندگان (دکر و بلت، ۲۰۰۸؛ کورنهابر^۲، ۲۰۰۴؛ شوهای و همکاران، ۱۹۹۶؛ وال و آلدرسون ۱۹۹۳) و ۶. نگرش و دیدگاه معلمان و شاگردان (چنگ، ۱۹۹۸، واتاناب، ۲۰۰۰).

این اطلاعات مفید سپس با سوژه‌ها/تم‌های به‌دست‌آمده از تحلیل مصاحبه‌های فاز کیفی، تحلیل سیاست‌های برنامه حذف کنکور سراسری (رک. کیانی، شایسته‌فر، غفارثمر و اکبری، ۲۰۱۲) و سیاست‌های کلی نظام آموزش عمومی و آموزش عالی (مندرج در اسناد بالادستی کشور)، تلفیق و نتیجه آن منجر به تشکیل پایگاه اطلاعات و نهایتاً ساخت مجموعه پرسش‌نامه آثار و تبعات تغییر آزمون شد. سازه‌های تشکیل‌دهنده این پرسش‌نامه نهایتاً عبارت بودند از: ۱. فعالیت‌های انجام شده در کلاس (۱۳ سنجه)؛ ۲. روش تدریس (۱۰ سنجه)؛ ۳. مواد و مطالب درسی (۹ سنجه)؛ ۵. طرح درس (۹ سنجه)؛ ۵. محتوای ارائه‌شده (۱۰ سنجه)، ۶. ارزشیابی یادگیری (۱۱ سنجه)؛ ۷. اهداف آموزشی: نوع شناختی (۱۰ سنجه)؛ ۸. اهداف آموزشی: نوع عاطفی (۴ سنجه)؛ ۹. عدالت در آزمون (۶ سنجه).

روایی صوری و محتوایی این ابزار ۸۲ سؤالی توسط دو استاد و یک معلم زبان انگلیسی بررسی شد. در نسخه نهایی پرسش‌نامه، ۹ مؤلفه فوق‌الذکر در قالب ۳ مقیاس/پرسش‌نامه اصلی دسته‌بندی شدند (به پیوست مراجعه کنید): بخش الف. طراح، برنامه‌ریزی و تدریس؛ بخش ب. اهداف آموزشی/یادگیری؛ و بخش ج. عدالت در آزمون. این بخش‌ها در قالب طیف لیکرتی زیر ارائه شدند: کاملاً مخالفم = ۱، مخالفم = ۲، کمی مخالفم = ۳، کمی موافقم = ۴، موافقم = ۵، کاملاً موافقم = ۶.

1. Alderson & Hamp-Lyons

2. Kornhaber

۳. ۴. تحلیل داده‌ها

۳. ۴. ۱. فاز کیفی: تحلیل مصاحبه‌های استادان و سرگروهان زبان انگلیسی

محقق به دفعات به داده‌های ضبط شده گوش فرا داده و با دقت به بازنویسی آن‌ها پرداخت. بر اساس سه رویه متمایز و در عین حال هم‌پوشاننده تحلیلی به کار رفته در این پژوهش، یعنی کدگذاری باز، کدگذاری محوری، و کدگذاری انتخابی/گزینشی (استراوس و گرین، ۱۹۹۹)، به‌عنوان وسیله‌ای جهت شناسایی تم‌ها/سوژه‌های موجود در داده‌ها، چندین تم/سوژه اصلی و فرعی به دست آمد که نشان‌دهنده درک شرکت‌کنندگان از تحول نظام کنکور محور و چشم‌اندازشان درباره آثار و تبعات آن بر نظام آموزش و یادگیری زبان انگلیسی دبیرستان بود. بر این اساس، یک پایگاه اولیه داده کیفی شکل گرفت و در گام‌های بعدی، محقق کوشش نمود با مضمون‌یابی و استخراج معانی مندرج در پاسخ‌ها موضوعات مرتبط را تولید و تبیین نماید. بر این اساس، ۲۱۲ تم تحت ۱۴ گروه رده‌بندی شدند. ضمناً بین کدگذاران (محقق این پژوهش و متخصص دیگری که با تحلیل کیفی و کدگذاری آشنا بود) ۹۵٪ توافق نظر حاصل شد.

۳. ۴. ۲. فاز کمی: تحلیل داده‌های پرسش‌نامه معلمان زبان انگلیسی

با توزیع پرسش‌نامه‌های طراحی شده بین معلمان، پایلوت اولیه انجام شد. پاسخ‌های مربوط به سؤالات منفی معکوس شدند (مثل سؤال‌های ۲۷ و ۶۱). ضریب پایایی به دست آمده برای مقیاس‌های طراحی، برنامه‌ریزی و تدریس، اهداف آموزشی و اصل عدالت به ترتیب ۰/۹۶، ۰/۹۰ و ۰/۹۲ بود. پس از محاسبه پایایی و روایی محتوا و صوری، پرسش‌نامه‌های توزیع شده بین ۴۶۰ نفر تحت پایلوت نهایی قرار گرفتند. از این تعداد، ۴۰۶ پرسش‌نامه بازگردانده شد (میزان بازگشت: ۸۶,۵٪).

به منظور بررسی روایی سازه، تأیید عاملی از نوع اکتشافی^۱ (SPSS، نسخه ۱۸) و نوع تاییدی^۲ (AMOS، نسخه ۱۸، آروباکل، ۲۰۰۹) انجام شد.

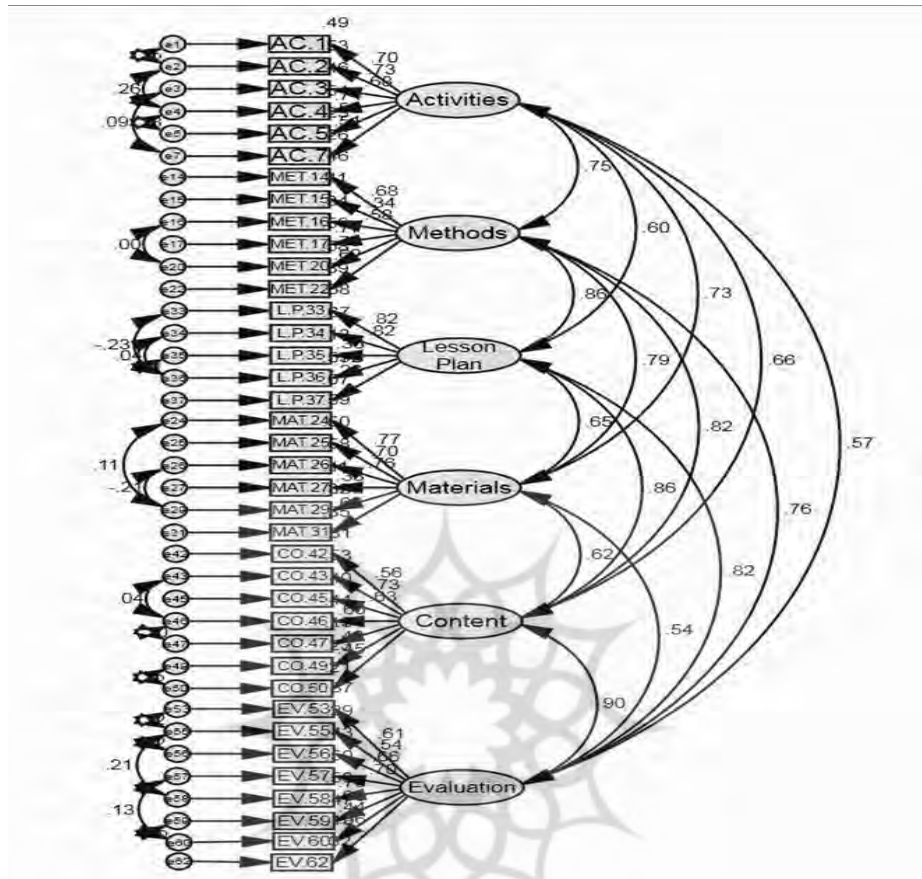
مقیاس طراحی، برنامه‌ریزی و تدریس

برای استخراج مؤلفه‌های این پرسش‌نامه، ابتدا از تحلیل عاملی اکتشافی و روش تحلیل مؤلفه اصلی (Component Analysis Principal) استفاده شد که نتایج آن نشان‌دهنده ظهور ۱۰ عامل اصلی برای مقیاس طراحی، برنامه‌ریزی و تدریس، با ارزش آیینگی (eigenvalue) بالاتر از ۱ بود که جملگی با هم حدود ۵۸٪ از کل واریانس را توضیح می‌دهند. بررسی دقیق‌تر اطلاعات این مقیاس نشان داد که نهایتاً می‌توان شش خرده‌مقیاس (شکل ۱) را برای آن متصور نمود. بارهای عاملی متغیرهای این خرده‌مقیاس‌ها در پیوست گزارش شده‌اند. بارهای عاملی مربوط به این مقیاس نشان داد که از ۶۲ سنجه اولیه، ۳۸ سنجه توسط ۶ مؤلفه اصلی توصیف می‌شوند.

شاخص‌های نیکویی برازش مقیاس طراحی، برنامه‌ریزی و تدریس

به منظور تأیید ساختار این مقیاس و خرده‌مقیاس‌های آن، از تحلیل عاملی تأییدی AMOS استفاده شد. هر ۳۸ سنجه بارگذاری شده بر ۶ عامل اصلی وارد این تحلیل شدند. برآورد شاخص کای-اسکوئر و مقادیر مرتبط با آن حاکی از مقدار ۲/۵۱ برای کای اسکوئر بهنجار شده (CMIN/DF) است که در دامنه مطلوب ۳ < و > ۱ قرار دارد. شاخص نیکویی برازش ۰/۸۸ (GFI=0.88)، و ریشه میانگین مربعات خطای برآورد ۰/۶۰ (RMSEA=0.60) مشاهده شد. این شاخص‌ها نشان از برازش خوب داده‌ها با مقیاس طراحی شده دارد.

1. Exploratory Factor Analysis
2. Confirmatory Factor Analysis



شکل ۱. نتیجه تحلیل عاملی تأییدی (AMOS) مقیاس طراحی، برنامه‌ریزی و تدریس

توضیح: ۱. فعالیت‌های انجام‌شده در کلاس (Activities)؛ ۲. متدولوژی/روش تدریس (Methods)؛ ۳. مواد و مطالب درسی (Materials)؛ ۴. طرح درس (Lesson Plan)؛ ۵. محتوای ارائه‌شده (Content)؛ ۶. ارزشیابی از یادگیری (Evaluation).

شاخص‌های جزئی برآزش، یعنی سطح معناداری (p-value) و نسبت بحرانی (Critical Ratio)، نشان داد که همه بارهای عاملی دارای تفاوت معناداری با صفر ($p < 0.05$) می‌باشند. برآورد تمامی پارامترهای مربوط به این مقیاس نشان داد که از ۶۲ سنج‌سازنده آن، ۳۸ سنج‌سازنده توسط ۶ مؤلفه اصلی آن توصیف و تأیید می‌شوند (شکل ۱).

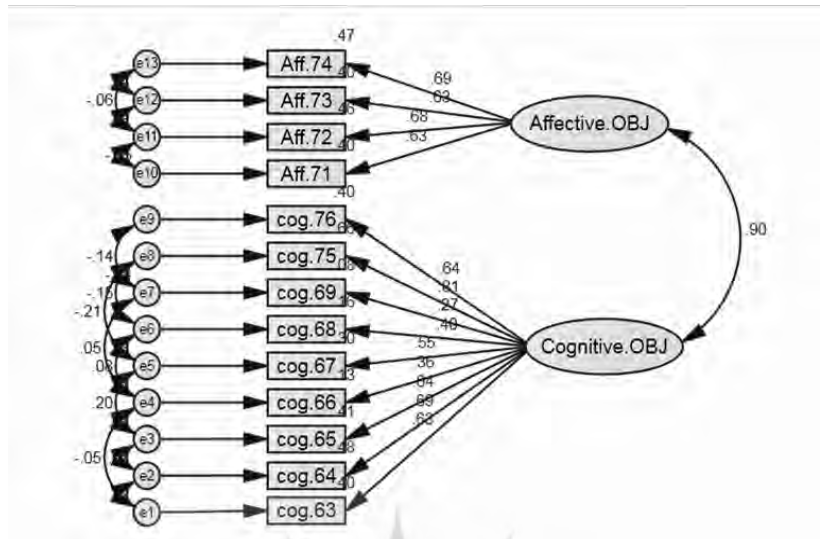
ضریب پایایی خرده‌مقیاس‌های مقیاس طراحی، برنامه‌ریزی و تدریس برای فعالیت‌های انجام‌شده در کلاس، روش تدریس، مواد و مطالب درسی، طرح درس، محتوای ارائه‌شده و ارزشیابی به‌ترتیب عبارت‌اند از: ۰/۸۲، ۰/۷۸، ۰/۵۹، ۰/۷۰ و ۰/۸۴.

مقیاس اهداف آموزشی/یادگیری

نتایج حاصل از تحلیل عاملی اکتشافی و تأییدی مقیاس اهداف آموزشی/یادگیری حاکی از عامل‌پذیری معنی‌دار این مقیاس با ۴ سؤال مربوط به اهداف عاطفی و ۱۰ سؤال مربوط به اهداف شناختی بود. در ابتدا، نتایج تحلیل عاملی اکتشافی حاکی از استخراج ۳ فاکتور بود که پس از بررسی بارهای عاملی مشاهده شده نهایتاً ۸ سنجه بر یک فاکتور (فاکتور شناختی) و ۴ سنجه دیگر بر فاکتور دیگر (فاکتور عاطفی) بارگذاری شدند. بارهای عاملی متغیرهای این خرده‌مقیاس‌ها در پیوست گزارش شده‌اند.

شاخص‌های نیکویی برازش مقیاس اهداف آموزشی/یادگیری

نتایج حاصل از تحلیل عاملی با AMOS تأیید کننده نتایج مرحله اکتشافی بود. فقط سؤال ۷۰ (امتحانات نهائی باعث می‌شود که دانش آموزان به جای اتخاذ مراحل رسیدن به هدف یا محصول به خود محصول و نیل به آن ترغیب شوند) بارگذاری بسیار ضعیفی (۰/۰۳) بر عامل خود داشت (شکل ۲). مقدار شاخص کای-اسکوئر بهنجار شده ۲/۷۹ مشاهده شد که در دامنه مطلوب ۳ < و ۱ > قرار دارد، شاخص نیکویی برازش ۰/۹۸ (GFI=0.98)، و ریشه میانگین مربعات خطای برآورد ۰/۰۴ (RMSEA=0.04) است. این شاخص‌ها نشان از برازش بسیار خوب داده‌ها با مقیاس طراحی شده دارد. ضریب پایایی خرده‌مقیاس‌های اهداف شناختی و اهداف عاطفی به‌ترتیب عبارت‌اند از ۰/۷۷ و ۰/۷۶.



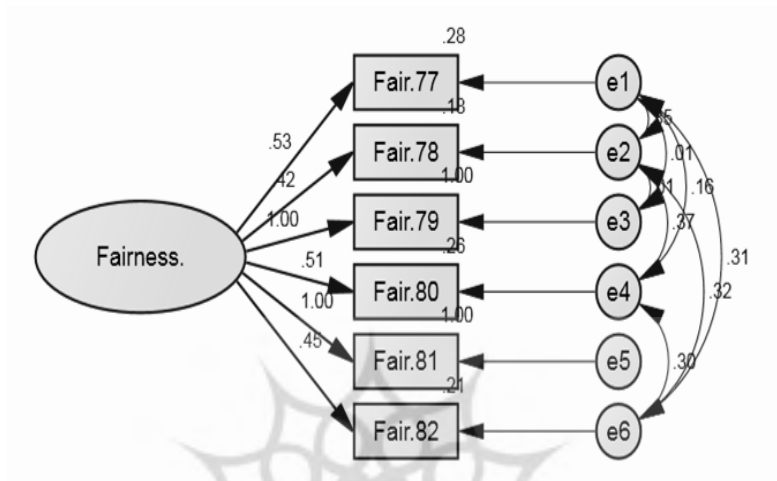
شکل ۲. تحلیل عاملی تأییدی (AMOS) مقیاس اهداف آموزشی

توضیح: اهداف آموزشی: نوع شناختی (Cognitive.OBJ)؛ ۸. اهداف آموزشی: نوع عاطفی (Affective.OBJ)

مقیاس عدالت در آزمون و شاخص‌های نیکویی برآزش آن

مقیاس عدالت در آزمون (یا عدالت در سنجش و پذیرش) یکی از دیگر از مقیاس‌های تشکیل‌دهنده ابزار کمی این پژوهش بود. این مقیاس که بر پایه یک سازه واحد طراحی شد به بررسی نگرش شرکت‌کنندگان نسبت به مقوله عدالت و برابری در سنجش و پذیرش در نظام مبتنی بر امتحانات نهایی سراسری پرداخت. ۶ سنجه تشکیل‌دهنده سازه اصلی بودند مانند: «در مقایسه با کنکور، امتحانات نهائی ملاک عادلانه‌ای برای سنجش پذیرش دانشجو برای آموزش عالی خواهند بود» (سؤال ۸۱) و «امتحانات نهایی در زمینه دستیابی به منابع/مطالب آمادگی برای آزمون فرصت‌های عادلانه‌ای را برای تمامی دانش‌آموزان فراهم می‌کنند» (سؤال ۷۹). در مرحله تحلیل عاملی اکتشافی یک فاکتور اصلی مشاهده شد که ۵۹٪ از واریانس مشاهده شده را توضیح می‌داد. نتایج تحلیل عاملی تأییدی با AMOS تأییدکننده بارگذاری معنی‌دار هر ۶ سؤال بر فاکتور اصلی بود. شاخص کای-اسکوئر بهنجار شده = ۱،۱۴؛ شاخص

نیکویی برآزش ۰/۹۹ و ریشه میانگین مربعات خطای برآورد ۰/۰۱ بود. ضریب پایایی مقیاس عدالت در آزمون ۸۵. گزارش شد.



شکل ۳. نتیجه تحلیل عاملی تأییدی (AMOS) مقیاس عدالت در آزمون. توضیح: عدالت در آزمون (Fairness)

۴. یافته‌ها و بحث

۴.۱. فاز کیفی

بازخوانی، کدگذاری و طبقه‌بندی داده‌های حاصل از مصاحبه‌ها منجر به استخراج ۲۱۲ تم شد که بر اساس محور معنایی نهایتاً در قالب ۱۴ مقوله اصلی طبقه‌بندی شدند (جدول ۱). بیشترین بسامد در مقولاتی همچون کفایت ملاک جدید (بسامد منفی=۳۹ بار) مشاهده شد که مفاهیمی همچون اعتبار، کیفیت، و مناسب بودن آزمون‌های نهایی زبان انگلیسی به‌عنوان معیار سنجش داوطلبان را در بر می‌گیرد. در ارتباط با این مقوله بسیار خطیر، اساتید و سرگروهان زبان انگلیسی نگرانی خود را از قدرت تمیز این آزمون‌ها (با مقیاس نمره‌دهی ۲۰-۰) در مقایسه با آزمون‌های چهارگزینه‌ای کنکور سراسری (با مقیاس نمره‌دهی ۱۰۰-۰) ابراز نمودند. از دیگر

نگرانی‌های مشهود آنان می‌توان به احتمال نمره‌دهی به‌ویژه افزایش نمرات مستمر دانش‌آموزان، خطر اعمال نظر شخصی و سلیقه‌ای عمل کردن معلمان، و استاندارد نبودن آزمون‌های نهایی برای سنجش دانش و توانایی واقعی دانش‌آموزان اشاره نمود. به عنوان مثال، EE7 اظهار داشت که «مهارت‌ها و محتوای سنجیده شده در این آزمون‌ها و همچنین منابعی که در طراحی سؤالات آن‌ها به‌کار می‌روند نه تنها کافی نیستند بلکه نمی‌توانند شاخص مناسبی از توانایی واقعی دانش‌آموزان باشند». نکته جالب این بود که سرگروهان بیشتر از اساتید بی‌کفایتی ملاک و معیار جدید را زیر سؤال بردند (بسامد منفی ۲۲ در مقابل ۱۷).

عدالت در آزمون (با بسامد کل ۲۱) سوژه دیگری بود که نظر نه‌چندان مساعد شرکت‌کنندگان را نمایان ساخت؛ چراکه تقریباً همه آن‌ها معتقد بودند آزمون‌های جدید از حیث معیارهای سنجش و اندازه‌گیری، شیوه‌های نمره‌دهی و همچنین استاندارد بودن از دقت و صحت کافی ابزاری برخوردار نبوده و در نتیجه نمی‌توانند فرد اصلاح را از غیر اصلاح تمییز دهند. یکی از شرکت‌کنندگان معتقد بود:

«اگر برقراری عدالت در سنجش و پذیرش از جمله اهداف اصلی سیاست‌گذاری حذف کنکور سراسری باشد، نظام اصلاح‌شده هم لزوماً شیوه منصفانه‌ای نیست، زیرا شرکت در آزمون‌های سراسری کشوری نظیر امتحانات نهایی دبیرستان خیلی متفاوت با شرکت در کنکور سراسری نخواهد بود... اگر به دنبال برقراری اصل عدالت و برابری در آزمون هستیم، باید همزمان چندین و چند اصلاح اساسی در این طرح ایجاد کنیم، یعنی اصلاحاتی که آنقدر اصولی، علمی و عمیق باشند که بتوانند بر بافت آموزش و پرورش تاثیرگذار بوده و آن را متحول سازند» (ET7).

جدول ۱. فراوانی آثار و تبعات ملاک جدید سنجش و پذیرش دانشجو بر نظام آموزش زبان انگلیسی دبیرستان: دیدگاه استادان، مدرسان و سرگروهان زبان انگلیسی

شماره	سوژه‌ها/تم‌ها	بسامد به تفکیک دو گروه				بسامد کل	
		سرگروهان		استادان		(+)	(-)
		(+)	(-)	(+)	(-)	(+)	(-)
۱	روش تدریس معلم	۳	۲	۴	۲	۷	۴
۲	کلاس‌های آمادگی برای آزمون	۷	۰	۱۲	۳	۱۹	۳
۳	فعالیت‌های ارائه‌شده توسط معلم	۳	۱	۳	۱	۶	۲
۴	مدیریت زمان توسط معلم	۱	۲	۳	۱	۴	۳
۵	حالات عاطفی/روحی-روانی دانش‌آموز	۱	۱	۵	۴	۶	۵
۶	حالات عاطفی/روحی-روانی معلم	۱	۱	۰	۴	۱	۵
۷	استفاده موکدمعلم از محتوا و مهارت‌های به‌خصوصی	۳	۲	۲	۱	۵	۳
۸	ارزشیابی دانش‌آموزان توسط معلم	۰	۴	۳	۳	۳	۷
۹	کیفیت خروجی و برونداد کلاس‌های زبان انگلیسی دبیرستان	۵	۱	۹	۳	۱۴	۴
۱۰	عدالت در سنجش و پذیرش	۹	۱	۱۲	۳	۲۱	۴
۱۱	فعالیت‌های یادگیری دانش‌آموزان	۵	۲	۳	۷	۸	۹
۱۲	استفاده مؤکدمعلم از مواد درسی و یا کمک درسی به‌خصوصی	۲	۱	۲	۱	۴	۲
۱۳	کیفیت و اعتبار ملاک‌ها و آزمون‌های جدید	۲۲	۱	۱۷	۲	۳۹	۳
۱۴	کیفیت و عمق دانش دانش‌آموز	۶	۰	۱۴	۰	۲۰	۰
کل		۶۸	۱۹	۸۹	۳۶	۸۹	۵۵
		۲۱۲					

توضیح: (+) اثرات مثبت و موردنظر و (-) اثرات منفی و نامطلوب آزمون‌های پایانی دبیرستان

کیفیت و عمق دانش و مهارت‌های کسب شده توسط دانش‌آموزان (با بسامد منفی ۲۰ در مقابل بسامد مثبت ۰) و همچنین کیفیت خروجی و برونداد کلاس‌های زبان انگلیسی دبیرستان (با بسامد منفی ۱۴ در مقابل بسامد مثبت ۴) دو مقوله مهم دیگری هستند که به‌نظر شرکت‌کنندگان به‌طور منفی تحت تأثیر برنامه جدید قرار می‌گیرند.

در این خصوص، نارضایتی نسبی آنان کاملاً محسوس بود. همه این افراد بر این باورند که خروجی نهایی وزارت آموزش و پرورش از حیث کیفیت، کمیت، میزان و عمق یادگیری آنچنان که سیاست‌گذاران پیش‌بینی می‌کنند، نخواهد بود. برخی از دلایل دخیل در این موضوع را با استناد به وضعیت کتاب‌ها و مطالب درسی دبیرستان، زمان کم یادگیری زبان در مدرسه، آموزش و یادگیری صرفاً آزمون‌گرا نه سازنده‌گرا و یادگیری محور، یادگیری طوطی‌وار و عدم وجود استانداردهای قوی و دقیق در سیستم فعلی آموزش و پرورش برای سنجش و اندازه‌گیری یادگیری دانش‌آموزان، مطرح نمودند. این عوامل می‌توانند در تشخیص نادرست توانایی واقعی زبانی دانش‌آموزان سهیم باشند و چه بسا دانش‌آموزی علی‌الظاهر دانش و مهارت‌های لازم زبان انگلیسی را کسب کرده باشد، اما در واقع توانایی واقعی او در همان سطوح مقدماتی باشد. یکی از پاسخ‌دهندگان (EB5) از یک پارادوکس سخن به میان آورد، پارادوکس بین استانداردهای بالای دانشگاه و انتظاراتی که از دانشجویان زبان انگلیسی می‌رود با استانداردهای نسبتاً متوسط یا پایینی که از نظر عملکردی برای دانش‌آموزان مقطع متوسطه وجود دارد.

از دیگر آثار منفی برنامه مبتنی بر امتحانات نهایی بازخیز همان بازار کلاس‌های آمادگی برای آزمون است که در گذشته برای موفقیت در کنکور و ورود به دانشگاه‌ها بسیار رایج بود. شرکت‌کنندگان به کرات به این موضوع اشاره کردند. همچنین بیان داشتند که انگیزه بالابردن نمرات امتحانات نهایی معلمان را وا می‌دارد فقط آن دسته از مهارت‌ها و مؤلفه‌هایی را که در این امتحانات مورد سنجش قرار می‌گیرند، در کلاس ارائه، تمرین و تأکید نمایند، خواه کلاس عادی باشد خواه کلاس تقویتی و یا کلاس آمادگی برای آزمون. همین آزمون‌محور بودن کافی است تا تعادل و توازن آموزش و یادگیری مهارت‌های اصلی زبان انگلیسی را بر هم زند. ET7 بر این باور بود که:

«...از آنجایی که معلمان سعی دارند نمرات امتحانات نهایی دبیرستان را بالا ببرند، در نتیجه تمرکز آن‌ها بیشتر بر نمونه سؤالات سال‌های پیش امتحانات نهایی قرا

می‌گیرد و بخش مهمی از وقت کلاس را به انجام این نمونه سؤالات اختصاص می‌دهند... فی الواقع، برای هر مهارت یا مؤلفه دیگری که هدف این آزمون‌ها نیست، همچون مهارت‌های ارتباطی و تعاملی گفتاری یا شنیداری، انگیزه و علاقه چندان‌ی نشان نمی‌دهند».

از دید شرکت‌کنندگان، بهبود در بخش‌های زیر نیز بر اساس نظر خوش‌بینانه سیاست‌گذاران پیش نخواهد رفت: روش تدریس معلم (بسامد ۷)، فعالیت‌های ارائه شده توسط معلم (بسامد ۶)، استفاده مؤکد معلم از محتوا و مهارت‌های به‌خصوصی (بسامد ۵) و مدیریت زمان (بسامد ۴). با اینکه این تم‌ها به میزان عدالت در سنجش و پذیرش، کیفیت و عمق دانش و مهارت کسب‌شده توسط دانش‌آموزان و همچنین کیفیت برون‌داد کلاس‌های زبان انگلیسی دبیرستان تکرار نشد، تعابیر منفی‌ای که برای هر یک از این‌ها مشاهده شد به مراتب بیشتر از تعابیر مثبت بود. این چهار تم، چهار شاخص تشکیل‌دهنده مقوله عام‌تر طراحی، برنامه‌ریزی و تدریس معلم هستند که شرکت‌کنندگان صحبت از تأثیرپذیری منفی آن‌ها به میان آوردند.

نکته قابل توجه درباره تم‌های فعالیت‌های یادگیری دانش‌آموزان و ارزشیابی دانش‌آموزان توسط معلم مشاهده شد. بسامد تأثیرات مثبت و منفی برای تم اول تقریباً مشابه بود (۹ برای اثر مثبت و ۸ برای اثر منفی) و این ممکن است بیانگر عدم قطعیت نظر پاسخگویان در مورد پیامدهای برنامه بر مؤلفه فعالیت‌های یادگیری باشد. اما بررسی دقیق داده‌های مصاحبه نشان می‌دهد که ایجاد تأثیرات مثبت در حوزه یادگیری بیشتر مربوط به افزایش سرعت و میزان/تعداد تمارین و فعالیت‌های انجام‌شده توسط دانش‌آموزان است نه عمق و کیفیت یادگیری. همان‌طور که EE1 توضیح داد «وقتی عملکرد فعلی دانش‌آموزان در امتحانات نهایی دبیرستان تعیین‌کننده ورود آنان به مقطع آینده باشد، در نتیجه، انگیزه یادگیری و انجام تمارین و فعالیت‌های امتحانات نهایی که بر اساس کتاب و مطالب درسی آموزش و پرورش طراحی می‌شوند، در آن‌ها افزایش می‌یابد»، وی ادامه داد «... و هر چه انگیزه بیشتری داشته باشند، بیشتر درگیر تمارین و فعالیت‌های این آزمون‌ها می‌شوند... و اوقات

کلاس گاه صرف فعالیت‌های تولیدی، البته شبیه آنچه در امتحانات نهایی هست، می‌شود تا صرفاً صرف فعالیت‌های دریافتی و کتبی که در کنکور سراسری وجود داشت». البته سرگروهان اثرات مثبت برنامه مبتنی بر امتحانات نهایی را چندان قابل توجه نمی‌دانند و اثرات منفی آن را بیشتر از اثرات سازنده و مثبت گزارش کردند.

وقتی صحبت از ارزشیابی از یادگیری دانش آموزان شد تقریباً همه پاسخگوها تصویری مثبت داشتند. از نظر آن‌ها طرح جدید سنجش و پذیرش، به‌طور کلی، به معلمان کمک می‌کند که برای سنجش عملکرد مستمر یا پایانی دانش‌آموزان آزمون‌ها و کوئیزهای متفاوتی را اجرا کنند یا روش‌های متفاوتی را تجربه کنند و به‌کار ببرند و همین امر معلمان را به سمت کسب مهارت‌های متفاوت ارزشیابی سوق می‌دهد. آن‌ها معتقد بودند که لحاظ سابقه عملکردی دانش‌آموزان به‌عنوان ملاکی تعیین‌کننده در ورود آن‌ها به آموزش عالی می‌تواند اثرات مثبتی بر فرهنگ ارزشیابی آموزشی داشته باشد، چراکه درک ضرورت و اهمیت این نوع ارزشیابی القاکننده ارزیابی تکوینی و مستمر خواهد بود که به‌جای خود بازخوردهای مثبتی برای یادگیری می‌دهد.

۴.۲. فاز کمی

برای دستیابی به تصویر کامل‌تری از تبعات و پیامدهای طرح جدید از افراد بیشتری از جامعه معلمان زبان انگلیسی آموزش و پرورش نظرسنجی به‌عمل آمد. برای این منظور، در فاز کمی پژوهش حاضر از مجموعه پرسش‌نامه آثار و تبعات تغییر آزمون (بخش ۲، ۲، ۳) استفاده شد. آمار توصیفی مربوط به مقیاس طراحی، برنامه‌ریزی و تدریس این ابزار حاکی از باور معلمان به تأثیرگذاری ملاک‌های جدید سنجش و پذیرش دانشجو بر بعضی از مؤلفه‌های این حوزه است. به‌طور مشخص، پاسخ‌های آنان نشان از تأثیرپذیری دو مؤلفه تمارین و فعالیت‌های انجام شده در کلاس و همچنین مواد و مطالب درسی بود. همان‌طور که در جدول ۲ مشاهده می‌شود میانگین پاسخ‌های مربوط به فعالیت‌های کلاسی ۴،۴۰ (انحراف معیار ۰/۹۶) و میانگین پاسخ‌های مربوط به مواد و مطالب درسی ۴/۳۵ (انحراف معیار ۰/۷۹) بود که هر دو درجاتی از نگرش مثبت شرکت‌کنندگان را نشان می‌دهد.

تحلیل درون‌مؤلفه‌ای نشان از شباهت و همخوانی پاسخ‌های سؤالات هر یک از این دو مؤلفه دارد. به‌عنوان مثال در مؤلفه فعالیت‌های کلاسی، نمره به‌دست‌آمده برای سؤال ۱ (نوع و ساختار تمارین و فعالیت‌های کلاسی)، سؤال ۲ (مهارت‌های آزمون‌دهی)، سؤال ۳ (صرف زمان بیشتر برای دستور زبان) و سؤال ۴ (تمارین به‌خصوصی که افزایش نمره در آزمون را به‌دنبال داشته باشد)، در حدود ۴/۵ گزارش شده که بیانگر نظرات موافق یا کاملاً موافق است. همین وضعیت در مورد مؤلفه مطالب درسی نیز مشاهده شد. معلمان بر این باور بودند که طرح جدید سنجش و پذیرش دانشجو تأثیر مثبت بر استفاده آنان از مواد و مطالب درسی و کمک درسی دارد و آگاهی و حساسیتشان را نسبت به «نقش کتاب‌های درسی و کمک درسی وزارت آموزش و پرورش در بالا بردن نمرات آزمون‌های نهایی» افزایش داده است. به‌طور مشخص، دو سؤال ۲۴ و ۳۱ گویای این مطلب‌اند: «از آنجائی که وزارت آموزش و پرورش طراح اصلی سؤالات آزمون‌های نهایی است، معلمان نیز آن دسته از مطالب و منابع درسی که توسط این وزارت‌خانه و سازمان‌های مربوطه تأیید و منتشر شده را به‌کار می‌برند؛ زیرا این مطالب به احتمال بسیار زیاد موضوعات و محتوای امتحانات نهایی را پوشش می‌دهند» (سؤال ۲۴) و «امتحانات نهایی معلمان و دانش‌آموزان را به استفاده از جزوات و مطالب آمادگی برای آزمون‌های نهایی دبیرستان (مثل کتاب‌های تست و جزوات گرامری رایج در بازار) ترغیب می‌کند» (سؤال ۳۱). ۷۰٪ از معلمان با انتخاب گزینه «موافقم» موافقت خود را در این زمینه اعلام نمودند.

از سوی دیگر، تأمل بیشتر در این دو مؤلفه ممکن است نتیجه‌ای متفاوت‌تر از تأثیرپذیری مثبت را بنمایاند. در واقع، بر خلاف اینکه میانگین گزارش شده برای این دو مؤلفه از میانگین سایر مؤلفه‌ها بیشتر است، اما شاید بیشتر بیانگر تأثیر سطحی یا بازشویی منفی ملاک‌های جدید سنجش و پذیرش باشد تا تأثیر عمیقاً مثبتی که مورد نظر سیاست‌گذاران این طرح بوده است. به‌عبارتی، اتکای بیش از حد معلمان به کتاب‌های درسی دبیرستان، آزمون‌های آزمایشی و نمونه‌سؤالات طراحی شده به همان

شکل سابق همچنان به قوت خود باقی است و تغییری نکرده است. بنابراین، نمی‌توان به‌سادگی ادعا کرد که فعالیت‌های کلاسی معلمان و دانش‌آموزان موافق با اهداف واقعی سیاست‌گذاری طرح پیش می‌رود و چه بسا با این نوع بازشویی یادگیری مادام‌العمر، آموزش و یادگیری با کیفیت و کسب مهارت‌های کاربردی و ارتباطی که مورد نظر این سیاست‌گذاری بوده است، به‌راحتی حاصل نشود (شایسته‌فر و همکاران، ۲۰۱۳).

جدول ۲. درک معلمان زبان انگلیسی از تأثیر آزمون‌های نهایی دبیرستان بر مؤلفه‌های طراحی، برنامه‌ریزی و تدریس، عدالت در سنجش و پذیرش و اهداف آموزشی/یادگیری

مؤلفه	میانگین	انحراف معیار	مینیمم	ماکزیمم
فعالیت‌های انجام شده در کلاس	۴,۴۰	.۹۶	۱	۶
محتوای ارائه شده	۳,۵۹	.۷۶	۱	۶
متدولوژی/روش تدریس	۴,۰۱	.۸۱	۱	۶
طرح درس	۳,۴۸	.۶۸	۱	۶
مواد و مطالب درسی	۴,۳۵	.۷۹	۱	۶
ارزشیابی از یادگیری	۳,۴۸	.۹۵	۱	۶
عدالت در آزمون	۲,۵۹	۱,۰۱	۱	۶
اهداف آموزشی: شناختی	۲,۹۱	.۹۲	۱	۶
اهداف آموزشی: عاطفی	۲,۸۹	۱,۱۲	۱	۶
طراحی، برنامه‌ریزی و تدریس	۳,۹۹	.۷۳	۲	۶

ناهم‌آهنگی بین اهداف سیاست‌گذاری و آنچه واقعاً در عمل وجود دارد در مؤلفه‌های محتوای تدریس، روش تدریس و ارزشیابی یادگیری دانش‌آموزان مشاهده شد (به ترتیب با میانگین ۳/۵۹، ۴/۰ و ۳/۴۸). در ارتباط با تأثیرپذیری روش تدریس، به‌عنوان مثال، دو سوم از شرکت‌کنندگان گفتند که روش تدریس خود را عمدتاً در راستای تقویت مهارت‌هایی که توسط آزمون سنجیده می‌شوند (سؤال ۱۷)، ارتقای نمرات دانش‌آموزان در آزمون (سؤال ۱۴) و آماده‌سازی دانش‌آموزان برای امتحان

(سؤال ۱۶) تغییر می‌دهند. این پاسخ‌ها و نظراتی از این جنس نشان می‌دهد که چگونه معلمان آموزش و پرورش ترغیب می‌شوند متدلوژی خود و فعالیت‌های کلاسی را صرفاً در جهت امتحانات نهایی تغییر دهند نه در جهت نیل به اهداف یادگیری.

به‌طور کلی، همین آزمون‌محور بودن، به احتمال زیاد، باعث شده این افراد تمایل چندانی به تغییر طرح درس، محتوای تدریس و شیوه‌ها و ابزارهای ارزشیابی یادگیری دانش‌آموزان یا حتی فرآیند این ارزشیابی نداشته باشند. میانگین سؤالات این سه مؤلفه ۳/۵ بود که با استناد به طیف لیکرت موجود نشان از نگرشی بی‌طرفانه نسبت به اثر امتحانات نهایی سراسری بر این سه مؤلفه دارد. این نگرش بیشتر درباره مؤلفه ارزشیابی از یادگیری دانش‌آموزان صادق است. درباره مهارت‌های تولیدی، نیمی از معلمان اظهار داشتند که به هر دو مهارت شنیداری (سؤال ۵۶) و همچنین مهارت گفتاری (سؤال ۶۲) اهمیت داده و آن‌ها را هم همچون مهارت‌های دریافتی مورد سنجش قرار می‌دهند، در حالی که نیمی دیگر به ضرورت سنجش این دو مهارت فقط به دلیل اینکه در آزمون‌های نهایی وجود ندارند، اعتقاد چندانی نداشتند. علاوه بر این، برخی از معلمان نظر موافق خود را با سؤال ۵۹ اعلام داشتند که می‌گوید: «از آنجایی که سؤالات امتحانات نهایی شامل سؤالات سطوح پائین هرم یادگیری (مثل به‌یادآوری مطالب، تشخیص بخشی یا جزئی در سؤال، و ...) است تا سطوح بالاتر هرم، من نیز سؤالات امتحانات خود را به‌همین صورت طراحی می‌کنم». برخی دیگر نظری مخالف داشتند. شایان توجه است که دیدگاه معلمان درباره تأثیرپذیری این مؤلفه متفاوت با دیدگاه اساتید و مدرسان (مصاحبه‌شوندگان) است؛ همان‌طور که پیش‌تر مطرح شد اساتدان و مدرسان به پیامدهای مثبت طرح جدید بر مؤلفه «ارزشیابی یادگیری دانش‌آموزان» اشاره نمودند.

علاوه بر پیامدهای درک شده در حوزه تدریس و مؤلفه‌های مرتبط با آن، بررسی نظرات بیان شده در خصوص عدالت در آزمون و فاکتورهای مرتبط با آن نیز حاکی از عدم همخوانی بین پیامدها و تبعات مشاهده شده با پیامدها و تبعات مطلوب و

مورد نظر سیاست‌گذاران طرح است. به‌طور کلی، ارزیابی معلمان از عدالت در سنجش و پذیرش مشابه ارزیابی استادان و مدرسان بود. یعنی آن‌ها نیز در مورد برقراری شرط عدالت و برابری برای همه داوطلبان چندان خوش بین نبودند ($2/59 < 3/5$ میانگین). بیش از نیمی از این معلمان به تأمین شرط دسترسی برابر به آموزش عالی (سؤال ۷۷) امیدوار نبودند و دو سوم آن‌ها نیز نسبت به اعتبار و قدرت امتحانات نهایی برای سنجش توانایی واقعی زبان انگلیسی دانش‌آموزان (سؤال ۷۸) دیدگاه مثبتی نداشتند.

در خصوص تأثیرپذیری اهداف آموزشی/یادگیری، معلمان نظری بی‌طرفانه داشتند، به‌ویژه در مورد اهداف عاطفی ($2/90$ میانگین). تحلیل درون‌گروهی نشان داد که اکثریت قریب به اتفاق پاسخگویان ($75/4\%$) نسبت به کاهش سطح فشار روانی دانش‌آموزان به‌دنبال جایگزینی کنکور با امتحانات نهایی دبیرستان (سؤال ۷۴) دیدگاه مثبتی نداشتند. وقتی از این معلمان درباره اهداف برنامه از جمله پوشش علایق متفاوت دانش‌آموزان (سؤال ۷۱) و تقویت اعتماد به‌نفس دانش‌آموزان (سؤال ۷۳) نظرسنجی شد، دو سوم آن‌ها (67%) نه در طیف موافق قرار گرفتند نه در طیف مخالف بلکه پاسخ‌هایشان حاکی از بی‌طرفی بود.

همین دیدگاه درباره تأثیرپذیری جنبه دیگر اهداف آموزشی یعنی اهداف شناختی مشاهده شد. بیش از نیمی از معلمان (219 نفر) به تقویت اهداف شناختی در اثر معیار قرار دادن امتحانات نهایی اعتقاد چندانی نداشتند یا حداقل تصور می‌کردند این تغییر بر همه جوانب شناختی یادگیری دانش‌آموزان اثر چشمگیری بر جای نخواهد گذاشت. برای مثال، تحلیل در سطح سؤالات این مؤلفه نشان از آن دارد که بیش از دو سوم معلمان (68%) بعید می‌دانند که برنامه جدید موجبات تقویت و پیشرفت خلاقیت دانش‌آموزان (سؤال ۶۴) را فراهم نماید و وقتی سخن از یادگیری مادام‌العمر و عمیق (سؤال ۷۵) به میان آمد، تعداد بیشتری از شرکت‌کنندگان (71%) درباره ایجاد چنین تغییر مطلوبی توسط برنامه جدید مردد بودند. در خصوص تقویت مهارت‌های حل مسئله (سؤال ۶۵)، یادگیری توأم با درک و فهم عمیق (سؤال ۶۷) و همچنین

تقویت یادگیری به شیوه همیاری و هم‌آموزی نه رقابتی (سؤال ۶۶)، پاسخ‌های بیش از دو سوم معلمان مقیاس «نسبتاً موافقم» بود.

۵. بحث و نتیجه‌گیری

پژوهش پیش رو، با الهام از دو مدل معروف ارزیابی پروژه تغییر، یعنی مدل کندی و مدل هنریکسن که تضمین کیفیت برنامه تغییر یافته را مستلزم ارزیابی بهنگام آن و نه دیرنگام یا در مراحل پایانی می‌دانند، به بررسی پیامدهای نظام تحول‌یافته سنجش و پذیرش دانشجو بر حوزه‌های آموزش و یادگیری زبان انگلیسی دبیرستان پرداخته است.

سیاست‌های طرح تغییر نظام سنجش و پذیرش دانشجو گواه آن است که نظام آموزشی و تربیتی کشور برای تکرر و تنوع روش‌ها و ابزار سنجش دانش و توانایی دانش‌آموزان مشروعیت قائل شده است و اصرار دارد فقط سنجش عده‌ای از مهارت‌های خاص تحت یک سری شرایط و موقعیت‌های خاص، الگو نشود. به عبارت دیگر، این نوع حرکت از سنت اندازه‌گیری و روان‌سنجی محض^۱ به سوی فرهنگ ارزشیابی (پارادایم ارزشیابی)^۲ (سیگرز و دوچی، ۲۰۰۱) نشان از آن است که سیاست‌گذاران و متولیان تعلیم و تربیت و سنجش و ارزشیابی به تدریج از انحصار نگاه تک‌بعدی در سنجش و پذیرش دست می‌شویند. مقولاتی همچون وارد کردن روش‌های کیفی نظیر مصاحبه و ارتباط شفاهی برای رشته‌های خاص مانند زبان انگلیسی، ثبت مستمر عملکرد دانش‌آموزان از ابتدای دبیرستان تا فارغ‌التحصیلی‌شان، به‌کارگیری روش‌های منظم ارزشیابی تکوینی به جای استفاده صرف از روش‌های تلخیصی، استفاده از فناوری اطلاعات در سنجش یادگیری فراگیران، لحاظ‌کردن فعالیت‌های فوق برنامه، خواه در سطح ملی یا بین‌المللی و طراحی و اجرای چندباره

1. Psychometric Pradigm

2. Edumetrics Culture

3. Segers & Dochy

آزمون‌های استاندارد و پیشرفت تحصیلی، برای مثال، به جای آزمون کنکور یک بار در سال همگی بر پیشرفت فوق‌الذکر صحنه می‌گذارند.

علی‌رغم محاسن یاد شده، بررسی طرح جدید در بستر عمومی آموزش و یادگیری زبان انگلیسی حاکی از آن است که بازیگران اصلی این عرصه، یعنی استادان، مدرسان و معلمان برخلاف سیاست‌گذاران طرح که ندای اصلاحات آزمون‌محور برای القای تبعات مطلوب در حوزه‌های یادگیری و آموزش را سر داده‌اند، چندان خوش‌بینانه به این قضیه نمی‌نگرند. واگرایی نسبی بین سیاست و عمل در برخی از مؤلفه‌های عرصه آموزش و یادگیری زبان انگلیسی (همچون محتوای ارائه‌شده، طرح درس،...)، همگرایی مشاهده‌شده در مورد درگیری دانش‌آموزان در تمرین‌ها و فعالیت‌های یادگیری کلاسی و هم‌سویی سطحی استفاده از مواد و مطالب درسی با سیاست‌های طرح جدید را می‌توان با استناد به دو دیدگاه تشریح نمود: تحقق اهداف یادگیری/آموزشی که صراحتاً توسط سیاست‌گذاران طرح بیان شد؛ مفهوم روایی سیستمی (همستی)^۱ که به صورت ضمنی در کلام آن‌ها بازتاب داشت.

طبق تعریف فردریکسن و کالینز^۲ (۱۹۸۹) و کونان^۳ (۲۰۰۴)، روایی همستی فرآیندی است که تبعات و پیامدهای مطلوب و سازنده‌ای را برای آموزش و یادگیری به دنبال خواهد داشت. در حقیقت، در روایی همستی، بر اساس ارتباط نزدیک بین آزمون و یادگیری و تأثیرگذاری آزمون بر یادگیری، معرفی یک آزمون به‌عنوان روندی پویا مقدمه‌ای خواهد بود برای تغییر نظام موجود. در شرایط فعلی، شاید چندان اغراق‌آمیز نباشد اگر بگوییم نیل به چنین اعتبار سیستمی (همستی) از طریق جایگزین کردن آزمون‌های کنکور سراسری با دیگر ابزارهای سنجش با ماهیتی

1. Systemic Validity
2. Fredriksen & Cullins
3. Kunnan

متفاوت تر که یادگیری محور و تا حدی سازنده‌گرا^۱ و مبتنی بر یادگیری فعال بوده، میسر می‌گردد.

از سویی دیگر، نتایج این مطالعه نشان داد که در بستر عمومی آموزش و یادگیری زبان انگلیسی کشور روایی همستی و اهداف یادگیری، آنچنان که منادیان آن انتظار داشتند به‌طور کامل حاصل نشده‌اند. یافته‌های مطالعاتی آلدرسن و همپ-لیونز (۱۹۹۶) و چنگ (۲۰۰۴) نشان می‌دهد که بافت و ذات آموزش و یادگیری زبان انگلیسی در عمل بسیار پیچیده‌تر و فراتر از ایده به‌ظاهر دست‌یافتنی ایجاد اصلاحات مطلوب از طریق سیاست‌های اصلاح‌طلبانه است. در پژوهش حاضر، گزارش عاملان عرصه تدریس حاکی از افزایش میزان و مدت زمان پرداختن به فعالیت‌های یادگیری و درگیری دانش‌آموزان در تمارینی است که کاملاً از سؤالات امتحانات نهایی الگو گرفته (مثل جمله‌سازی با کلمات به‌هم‌ریخته، شکل صحیح واژگان فعلی، صفتی و قیدی، پرکردن جای خالی تک‌کلمه‌ای یا نهایتاً دو کلمه‌ای، و دستور زبان)، اما نگاهی موشکافانه‌تر به آنچه عملاً در کلاس رخ می‌دهد، چه از جانب معلمان و چه از جانب دانش‌آموزان، حاکی از آن است که حتی طرح جدید هم فعلاً نتوانسته عمیقاً و اساساً تدریس معلمان و یادگیری دانش‌آموزان را بهبود بخشد. اگرچه اصلاحات مورد نظر، اهدافی غایی مانند تدریس معنی‌دار با رویکرد ارتباطی، یادگیری عمیق و معنی‌دار با رویکرد ارتباطی و فعالانه نه منفعلانه را دنبال نموده اما ادراکات، اظهارات و مشاهدات شرکت‌کنندگان حاضر مؤید این مطلب نیست. از این‌رو، تعبیر نتایج باید با دقت نظر صورت گیرد نه سطحی و شتاب‌زده. از دید چنگ (۱۹۹۸) تغییراتی شبیه آنچه در بالا بدان اشاره شد تغییراتی سطحی‌اند و به‌قول چن (۲۰۰۲) ممکن است فقط در چپستی تدریس و یادگیری کارگر افتند نه در چگونگی و ذات آن. این امر پیچیدگی ماهیت اعتبار همستی یا تأثیرات واقعی بر آموزش و یادگیری زبان انگلیسی را بیش از پیش می‌نمایاند.

یافته‌های مربوط به رفتار آموزشی روزانه معلمان، به‌طور ویژه، مؤید دیدگاه آلدرسن و همپ‌لیونز (۱۹۹۶) است که معتقدند تأثیرات بازشویی مطلوب ممکن است یکسان و مشابه رخ ندهند، بلکه بر اساس جنبه‌های مختلف تدریس و بافت ویژه آن به اشکال متفاوت و غیر قابل انتظار جلوه می‌کنند. در پژوهش حاضر، مؤلفه‌های روش تدریس و مواد و مطالب درسی صرفاً در راستای امتحانات نهایی دبیرستان تغییر کرد نه در جهت سیاست‌های مطلوب فوق‌الذکر. تأکید و تمرکز معلمان بر منابع و نمونه‌سؤالات و آزمون‌های آزمایشی که ناشران و مؤلفان وزارت آموزش و پرورش یا سازمان‌های همکار طراحی و تألیف کرده‌اند نشان از شدت تأثیر امتحانات نهایی بر انتخاب مواد و مطالب درسی آموزش و پرورش (همچون کتاب‌ها و جزوات دبیرستانی) توسط معلمان و محدود شدن آن‌ها به این منابع دارد. به‌علاوه، به‌دلیل هم‌سویی طرح درس و ارزشیابی یادگیری دانش‌آموزان (چه تکوینی/تلخیصی یا تشخیصی) با برنامه امتحانات نهایی سراسری، نادیده‌گرفتن مهارت‌های شنیداری و ارتباطی در این دو مؤلفه همچنان به قوت خود باقی و بدون تغییر است. این نتایج با یافته‌های مکتوب درباره پیامدهای آزمون‌های مهم مطابقت دارد. طبق این یافته‌ها، آزمون‌های سرنوشت‌ساز و پر وجه، شیوه‌های تدریس را محدود می‌کنند و توانایی معلمان در تدریس محتوا و کاربست روش‌ها و مطالب درسی ناسازگار با قالب آزمون‌های استاندارد شده را کاهش می‌دهند (اسمیت^۱، ۱۹۹۱).

از زاویه تحقق اهداف یادگیری/آموزشی که بنگریم، شاهد ادراکات و نگرش منفی شرکت‌کنندگان نسبت به رفع فشارهای روانی همچون استرس و اضطراب آزمون هستیم. به‌عبارتی، آن‌ها فکر نمی‌کنند که جانشینی سابقه تحصیلی و آزمون‌های نهایی دبیرستان به‌جای کنکور سراسری، استرس و اضطراب ناشی از رقابت برای ورود به دانشگاه را از بین ببرد؛ بلکه برعکس همان آسیب‌های روحی‌روانی که سالیان سال در مورد نظام کنکور محور وجود داشت همچنان به قوت خود باقی است. اگر چه ممکن است شکل آن به‌ظاهر تغییر یابد. در خصوص ایجاد پیامدهای مثبت، همچون افزایش

1. Smith

علاقه و انگیزه یادگیری در دانش‌آموزان، شرکت‌کنندگان نظر کاملاً مساعد یا کاملاً مخالف نداشته و در این زمینه مطمئن نبودند. نتایج گواه آن است که این افراد برنامه جدید در شکل فعلی‌اش را فاقد پتانسیل کافی برای تقویت همه‌جانبه ابعاد روحی‌روانی و عاطفی دانش‌آموزان می‌دانند.

نگرش و ادراکات منفی معلمان زبان انگلیسی هشدار می‌دهد که خبر از عدم تمایل آن‌ها به ایجاد تغییرات مورد انتظار می‌دهد. شاید بتوان این اکراه معلمان را با تردید آن‌ها درباره مقبولیت، اعتبار، کفایت و مناسب بودن امتحانات نهایی برای تصمیم‌گیری‌های مهم از جمله ورود به عرصه آموزش عالی توجیه نمود. دلیل دیگر بروز چنین تردیدهایی این است که معلمان زبان انگلیسی با نظام امتحانات نهایی کاملاً آشنا بوده و از تجربه کافی در این زمینه برخوردارند و تنها تفاوتی که وجود دارد این است که این آزمون‌ها در گذشته، ملاک تصمیم‌گیری برای ورود به دانشگاه‌های کشور نبوده‌اند. درجاتی از هم‌پوشانی مشاهده‌شده بین نظرات استادان و مدرسان (داده‌های کیفی) و معلمان (داده‌های کمی) نشان می‌دهد امتحانات نهایی سراسری در شکل کنونی خود، اهرمی کارآمد و قدرتمند برای ایجاد تغییرات مورد نظر سیاست‌گذاران طرح نمی‌باشند. شاید به همین دلیل است که به‌طور کلی در حوزه طراحی، برنامه‌ریزی و تدریس، تغییرات چشمگیری گزارش نشد. دیدگاه نه‌چندان مثبت نسبت به برقراری شرط عدالت در سنجش و پذیرش نیز شاهدی دیگر بر این ادعاست.

نکته شایان توجه این است که طرح جدید سنجش و پذیرش دانشجو از قدرت و شفافیت لازم برای ایجاد نگرش مثبت در افراد ذی‌ربط و ذی‌نفع همچون استادان، مدرسان و معلمان زبان انگلیسی برخوردار نیست. ادبیات موجود درباره مدیریت و اجرای طرح تغییر نشان می‌دهد که چگونه درک و نگرش معلمان نسبت به کارایی، سودمندی و عملی و قابل اجرا بودن یک تغییر بر تمایل آن‌ها به پذیرش و اجرای تغییر، یا به عبارتی بهتر، تغییرپذیری معلمان تأثیر می‌گذارد (کارلس، ۱۹۹۷). اگرچه

رابطه بین درک و نگرش افراد با عملکرد آنان همیشه مستقیم و خطی نیست (کندی و کندی^۱، ۱۹۹۶)، احتمال مقاومت معلمان در برابر تغییر زمانی محتمل خواهد بود که این تغییر با دیدگاه و درک آنان سازگار نباشد. این بدان معناست که اعتقاد معلمان به اصلاح آموزش و یادگیری آزمون‌محور تغییر نمی‌کند مگر اینکه نگرش و درک آنان از تبعات و پیامدهای احتمالی آزمون تغییر کند (چنگ، ۲۰۰۴؛ چنگ و کورتیس^۲، ۱۹۹۴؛ وال و آلدسون، ۱۹۹۳). بر اساس این شواهد و قرائن می‌توان نتیجه گرفت تا زمانی که افراد ذی‌ربط و ذی‌نفع برنامه‌ای اصلاح‌شده دیدگاه مثبتی نسبت به اصلاحات صورت‌گرفته نداشته باشند، هر گونه سیاستی برای تغییر رفتار و عملکرد آنان بی نتیجه خواهد ماند.

نگرانی‌های مهم مطرح‌شده در این پژوهش بخشی از چالش‌های آینده را به تصویر می‌کشد. شاید این نگرانی‌ها بازتاب این تفکر است که می‌گویند مادامی که چالش‌های مهم برنامه به صورت جدی مورد مذاقه و بررسی واقع نشوند، سیاست‌های طراحی‌شده به صورت کامل و مطلوب جامه عمل نخواهند پوشید. این مسئله زمانی پررنگ‌تر شد که پاسخ‌دهندگان احتمال بازخیز آسیب‌های روانی و فشار و اضطراب آزمون‌ی و توزیع آن‌ها در کل دوران چهار ساله دبیرستان را مطرح نمودند. حضور فیلترهای عاطفی نامطلوب همچون فشار روانی و اضطراب کافی است تا شالوده نگرش‌های منفی نسبت به عملکرد برنامه جدید را شکل داده و مقبولیت برنامه را تضعیف نماید. از دیگر مسائل چالش‌برانگیزی که مورد توجه شرکت‌کنندگان حاضر قرار گرفت اعمال نظرهای شخصی معلمان و سلیقه‌ای عمل کردن، جانبداری و تعصب آن‌ها در قضاوت درباره دانش‌آموزان و همچنین تداوم همان مقوله آموزش برای آزمون می‌باشد. در این موارد، ذکر این نکته کافی است که بگوییم «بهبود عملکرد قبلی» (مارکی^۳، ۱۹۹۰؛ راجرز^۴، ۱۹۸۳؛ استلر،

1. Kennedy & Kennedy
2. Cheng & Curtis
3. Markee
4. Rogers

۱۹۹۴) مهم‌ترین شاخص قضاوت درباره میزان مقبولیت هر نوآوری، تغییر یا تحولی همچون تغییر کنکور در ایران است.

تحول نظام سنجش و پذیرش در بطن اجرایی خود مواردی همچون سازماندهی، ساختارسازی، الگوهای راهبردی و مهارت‌های توسعه و اجرای تکنیک‌های نوین را در بر می‌گیرد. از این منظر، تغییر نظام کنکور در ایران تغییر پیچیده‌ای است که نیازمند مهارت و خبرگی لازم برای ثبت و ضبط پیچیدگی‌های عملکرد دانش‌آموزان یا داوطلبان آتی است. این موارد و شمار دیگری از پیشنهادها و چالش‌هایی که در بالا مطرح شد، ماهیت پیچیده تبعات و پیامدهای اجتماعی آزمونی مهم و سرنوشت ساز را باز می‌تاباند. بررسی کارایی و کارآمدی آزمون و پیامدهای اجتماعی متفاوتی که در بر دارد در رأس مقوله اعتبارسنجی آزمون از بعد اجتماعی قرار دارد. بررسی آزمون‌ها از بعد اجتماعی آن‌ها زمینه مشارکت ذی‌ربطان، ذی‌نفعان و ذی‌صلاحان را در یک شبکه ارزیابی مشارکتی (کینگ^۱، ۲۰۰۵) فراهم می‌نماید که از طریق آن با یکدیگر به اشتراک نظر پرداخته و محملی را برای انتقال نتایج ارزیابی (همچون نقاط قوت، نقاط ضعف، چالش‌ها و تهدیدهای مشاهده‌شده) به مقامات سیاست‌گذار و تصمیم‌گیرنده فراهم می‌آورند.

در کل، آنچه زنگ هشدار را برای سیاست‌گذاران و برنامه‌ریزان کشور به صدا در می‌آورد این است که برخی عوامل می‌توانند هرگونه سیاست مدون مطلوب را تضعیف نمایند و در عمل، موجب پیدایش چالش‌ها و تهدیدهای ناشناخته شوند. بنابراین هر فاز تصمیم‌گیری و سیاست‌گذاری یا هر فاز اجرایی همراه رسالت ذاتی‌اش نیازمند آن است که از طریق لنز دقیق ارزیابی بررسی شود تا بتواند حرکت نوآورانه و تحول‌گر را مفید گرداند. نوع استفاده‌ای که از نتایج آزمون‌های کنکور و تفسیر نمرات آن‌ها می‌شود و چالش‌های مرتبط با شرایط کنونی کشور بر این دلالت دارند که کماکان همان نقش‌گزینشی^۲ است که جامعه را به سمت معیار قرار دادن

1. King

2 Selective Function

امتحانات نهایی و سابقه تحصیلی سوق داده است. چنین انتخابی احتمالاً مشکلات دیرینه ناشی از عواقب نامطلوب آزمونی را همانند قبل تقویت می‌کند، چرا که رقابت برای کسب نمرات بالاتر همچنان به قوت خود باقی است. این مسئله مسلماً نقش برنامه جدید سنجش را به‌عنوان ابزاری برای پیشبرد، القا و احصای تغییرات مطلوب در آموزش و یادگیری زبان انگلیسی مقطع متوسطه تضعیف خواهد نمود. بنابراین، تضمین کاملی وجود نخواهد داشت که پیامدهای برنامه جانشین کنکور به‌طور کامل با سیاست‌های اتخاذ شده و مورد نظر فاز سیاست‌گذاری هم‌سو و منطبق باشند، مگر این‌که تغییرات بزرگی از این قبیل از چشم‌اندازها و زوایای مختلف به‌دقت مورد بررسی و ارزشیابی زود هنگام و مناسب قرار گیرند.

کتابنامه

۱. شورای عالی آموزش و پرورش. (۱۳۸۹). *برنامه درسی ملی جمهوری اسلامی ایران*، نگاشت سوم. سازمان پژوهش و برنامه‌ریزی آموزشی.
۲. شورای عالی انقلاب فرهنگی. (۱۳۸۹). *نقشه جامع علمی کشور*.
۳. قلی‌پور، ر.، و آقابزرگی، م. (۱۳۸۴). *گزارش بررسی پیشنهاد لایحه گزینش دانشجو در ایران*. دفتر فرهنگی - کمیته آموزش و تحقیقات مجلس شورای اسلامی.
۴. کیا، ع.، و بزرگی، ک. (۱۳۸۵). *اظهار نظر کارشناسی درباره طرح نحوه پذیرش در دانشگاه‌ها و مراکز آموزش عالی کشور*. مرکز پژوهش‌های مجلس شورای اسلامی.
۵. کیامنش، ع. (۱۳۷۹). *ارزشیابی آموزشی*. تهران: پیام نور.
۶. مجلس شورای اسلامی. (۱۳۹۲). *قانون پذیرش دانشجو در دانشگاه‌ها و مراکز آموزش عالی کشور*. مجلس شورای اسلامی.
7. Alderson, J. C., & Hamp-Lyons, L. (1996). TOEFL preparation courses: A study of washback. *Language Testing*, 13(3), 280-297.
8. Arbuckle, J. L. (2009). *IBM SPSS Amos 18 user's guide*. Chicago, IL: IBM.
9. Bachman, L., & Palmer, A. (2010). *Language assessment in practice*. Oxford: Oxford University Press.
10. Camilli, G. (2006). Test fairness. In R. Brennan (Ed.), *Educational measurement* (4th ed., pp. 221-256). Westport, CT: American Council on Education and Praeger.

11. Carless, D. (1997). Managing systematic change: A critical analysis of Hong Kong's target-oriented curriculum initiative. *International Review of Education*, 43(4), 349-366.
12. Chapman, D., & Snyder, C. W. (2000). Can high stakes national testing improve instruction: Reexamining conventional wisdom. *International Journal of Educational Development*, 20, 457-474.
13. Cheng, L. (1998). Impact of public English examination change on students' perceptions and attitudes toward their English learning. *Studies in Educational Evaluation*, 24(3), 279-301.
14. Cheng, L. (2002). *Taiwanese junior high school English teachers' perceptions of the washback effect of the basic competence test in English* (Unpublished doctoral dissertation). The Ohio State University, Ohio, USA.
15. Cheng, L. (2004). The washback effect of a public examination changes on teachers' perceptions toward their classroom teaching. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 146-170). Mahwah, NJ: Lawrence Erlbaum Associates.
16. Cheng, L. (2005). *Changing language teaching through language testing: A washback study*. Cambridge: Cambridge University Press.
17. Cheng, L., & Curtis, A. (2004). Washback or backwash: A review of the impact of testing on teaching and learning. In L. Cheng, Y. Watanabe, & A. Curtis (Eds.), *Washback in language testing: Research contexts and methods* (pp. 3-17). Mahwah, New Jersey: Lawrence Erlbaum.
18. Cheng, L., & Sun, Y. (2015). Interpreting the impact of the Ontario Secondary School Literacy Test on second language students within argument-based validation framework. *Language Assessment Quarterly*, 12(1), 50-66.
19. Cheng, L., Watanabe, Y., & Curtis, A. (2004). *Washback in language testing: Research contexts and methods*. London: Laurence Erlbaum Associates, Publishers.
20. Cheng, L., Andrews, S., & Yu, Y. (2011). Impact and consequences of school-based assessment (SBA): Students' and parents' views of SBA in Hong Kong. *Language Testing*, 28(2), 221-249.
21. Decker, D. M., & Bolt, S. E. (2008). Challenges and opportunities for promoting student achievement through large-scale assessment results: Research, reflections, and future directions. *Assessment for Effective Intervention*, 34(1), 43-51.
22. Fredriksen, J., & Collins, A. (1989). A system approach to educational testing. *Educational Researcher*, 18, 27-32.
23. Hargreaves, A. Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal*, 39(1), 69-95.
24. Henrichsen, L. E. (1989). *Diffusion of innovations in English language teaching: The ELEC effort in Japan 1956-1968*. New York: Greenwood Press.

25. Hughes, A. (1988). Introducing a need based test of English language proficiency into an English medium university in Turkey. In A. Hughes (Ed.), *Testing English for university study* (pp. 134-153). London: Modern English Publications.
26. Kane, M. (2002). Validating high-stakes testing program. *Educational Measurement: Issues and Practices*, 21(1), 31-41.
27. Kane, M. (2011). Validating score interpretations and uses. *Language Testing*, 29(1), 3-7.
28. Kennedy, C. (1988). Evaluation of the management of change in ELT projects. *Applied Linguistics*, 9(4), 329-342.
29. Kennedy, C., & Kennedy, J. (1996). Teacher attitudes and change implementation. *System*, 24(3), 351-360.
30. Kiany, Gh., Shayestefar, P., Ghafar Samar, R., & Akbari (2013). High-rank stakeholders' perspectives on high-stakes university entrance examinations reform: Priorities and problems. *Higher Education*, 65(3), 325-340.
31. Kikuchi, K., & Browne, C. M. (2009). English educational policy for high schools in Japan: Ideals vs. reality. *RELC*, 40, 172-191.
32. King, J. A. (2005). Participatory evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 291-294). Thousand Oaks, CA: Sage.
33. Kornhaber, M. L. (2004). Appropriate and inappropriate forms of testing, assessment, and accountability. *Educational Policy*, 18(1), 45-70.
34. Kunnan, A. J. (2004). Regarding language assessment. *Language Assessment Quarterly*, 1(1), 235-253.
35. Markee, N. (1990, March 10th). *The diffusion of communicative innovations and classroom culture: An ethnographic study*. Paper presented at the 24th Annual TESOL Convention, San Francisco.
36. McNamara, T., & Roever, C. (2006). *Language testing: The social dimension*. Oxford: Blackwell Publishing.
37. Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York: American Council on Education and Macmillan.
38. Munoz, A. P., & Alvarez, M. E. (2010). Washback of an oral assessment system in the EFL classroom. *Language Testing*, 27(1), 33-49.
40. Nagy, P. (2000). The three roles of assessment: Gate keeping, accountability, and instructional diagnosis. *Canadian Journal of Education*, 25(4), 262-279.
41. Pan, Y-Ch., & Newfields, T. (2012). Tertiary EFL proficiency graduation requirements in Taiwan: A study of washback on learning. *Electronic Journal of Foreign Language Teaching*, 9(1), 108-122.
42. Pizzaro, M. A. (2010). Exploring the washback effects of a high-stakes English test on the teaching of English in Spanish upper secondary schools. *Revista Alicantina de Estudios Ingleses*, 23, 149-170.
43. Rogers, E. M. (1983). *The diffusion of innovations* (3rd ed.). London: Macmillan.

44. Segers, M., & Dochy, F. (2001). New assessment forms in problem-based learning: The value added of students' perspective. *Studies in Higher Education*, 26(3), 327-343.
45. Shayestefar, P., Kiany, G., Ghafar Samar, R., & Akbari, R. (2012). Reform in high-stakes university entrance examinations of Iran: A critical base-line study. *International Journal of Humanities*, 19(2), 97-122.
46. Shohamy, E. (1993). *The power of tests: The impact of language tests on teaching and learning*. NFLC Occasional Paper. Washington, D.C.: National Foreign Language Center.
47. Shohamy, E., Donista-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13, 298-317.
48. Smith, M. L. (1991). Meaning of test preparation. *American Educational Research Journal*, 28, 521-542.
49. Spratt, M. (2005). Washback and the classroom: The implications for teaching and learning of studies of washback from exams. *Language Teaching Research*, 9(1), 5-29.
50. Stoller, F. (1994). The diffusion of innovations in intensive ESL program. *Applied Linguistics*, 15(30), 300-327.
51. Strauss, A. & Corbin, J. M. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
52. Wall, D. (2000). The impact of high-stakes testing on teaching and learning: Can this be predicted or controlled? *System*, 28, 499-509.
53. Wall, D., & Alderson, J. C. (1993). Examining washback: The Sri Lankan impact study. *Language Testing*, 10, 41-69.
54. Wall, D., & Horak, T. (2008). *The impact of changes in the TOEFL examination on teaching and learning in central and eastern Europe: Phase 2, Coping with change*. TOEFL iBT Research Service Report No. TOEFL Ibt-05, Princeton, NJ: Educational Testing Service.
55. Wall, D., & Horak, T. (2011). *The impact of changes in the TOEFL exam on teaching in a sample of countries in Europe: Phase 3, The role of the course book. Phase 4, describing change*. TOEFL iBT Research Service Report No. TOEFL RR-11-41, Princeton, NJ: Educational Testing Service.
56. Watanabe, Y. (2001). Does the university entrance examination motivate learners? A case study of learner interviews. In Akita Association of English Studies (Eds.), *Trans-equator exchanges: A collection of academic papers in honor of Professor David Ingram* (pp. 100-110). Akita, Japan: Author.
57. Winke, P. (2011). Evaluating the validity of a high-stakes ESL test: Why teachers' perceptions matter. *TESOL Quarterly*, 45(4), 628-660.