

# ارزشیابی و سنجش: اهرم اصلاح آموزش

تحلیلی بر یافته‌های ارائه شده در کنفرانس ۴۳  
«انجمن بین‌المللی سنجش آموزش»

ابراهیم طلایی

عضو هیئت علمی دانشگاه تربیت مدرس

## مقدمه

چهل و سومین کنفرانس سالانه انجمن بین‌المللی سنجش آموزش که ۱۰ تا ۱۴ مهرماه ۱۳۹۶ در باتومی گرجستان برگزار شد، بهانه‌ای برای توجه مجدد به یکی از عناصر مهم آموزش، یعنی سنجش و ارزشیابی، بود. چهل و چهارمین کنفرانس سالانه مرکز «ارزشیابی آموزشی» دانشگاه آکسفورد در ۹ الی ۱۴ سپتامبر ۲۰۱۸ برگزار شد. نشست بعدی این کنفرانس در ۲۰ الی ۲۷ سپتامبر سال ۲۰۱۹ میلادی در باکو برگزار خواهد شد. این برای من نیز فرصتی فراهم آورد که هم آنچه را در این کشورها می‌گذرد، بشنوم و هم از پیشرفت‌های علمی متخصصان و مراکز تخصصی اندازه‌گیری و سنجش آموزش در دانشگاه‌ها مطلع شوم. در نهایت هم فرصتی فراهم شد تا ساختار اداری و ارتباط منطقی و غیرمنطقی دولت، بخش خصوصی و معلمان را در حوزه سنجش و ارزشیابی بررسی کنم.

لازم است در ابتدا توضیح دهم گرچه سنجش و ارزشیابی و اندازه‌گیری، مفاهیمی متفاوت و قابل تفکیک در گفتمان علوم تربیتی و اجتماعی هستند، در این نوشته این تفکیک مفهومی با دقتی که در کتاب‌های درسی از آن یاد می‌شود، منظور نظر نیست و گهگاه از ارزشیابی و سنجش به جای هم استفاده شده است.

سخنرانان کلیدی کنفرانس امسال، خانم پروفیسور جوان برد<sup>۱</sup> از دانشگاه آکسفورد (رئیس دپارتمان تعلیم و تربیت و رئیس مرکز تحقیقات ارزشیابی آموزش)، پروفیسور<sup>۲</sup> هری هاران سوامی ناتان رئیس دپارتمان روان‌شناسی تربیتی دانشگاه کنکتیکت<sup>۳</sup> و استیور یو کین<sup>۴</sup> رئیس دپارتمان اقتصاد دانشگاه ایلینویز بودند:

یک روز قبل از کنفرانس، سه کارگاه هم به صورت موازی برگزار می‌شد که موضوعات آن جالب توجه و ایده‌بخش بودند: **چگونه می‌توان مهارت‌های تفکر سطح بالا (پیچیده) را سنجید؟ طراحی سؤال‌های کتبی و رایانه‌ای؛ رایانه‌ای کردن سنجش‌های کلان مقیاس و خطیر؛ تجربه‌ها، چالش‌ها و انتظارات، چگونه شواهد خوبی برای روایی و پایایی و عادلانه بودن آزمون فراهم کنیم.**

## هشت مضمون اصلی طرح شده در کنفرانس ۴۳:

**۱) تخصصی و حرفه‌ای بودن تولید سؤال خوب:** برای تولید آزمونی که قابل اطمینان باشد و براساس آن بتوان تصمیم‌گیری کرد (یا به اصطلاح آزمون‌های خطیر) همچون آزمون‌های نهایی دیپلم و کنکور سراسری، مسیر تخصصی و مراحل دوازده گانه خاصی باید طی شود. که این دقت برای تولید آن آزمون هزینه قابل توجهی می‌طلبد و مهم‌تر اینکه، لازم است از در دسترس قرار دادن سهل و سریع چنین آزمون‌های گرانی جلوگیری شود. این آزمون‌ها لزوماً چهار جوابی نیستند یا اصلاً نباید تعداد زیادی از آن‌ها چهار جوابی باشند؛ چرا که سؤالات چهار جوابی محدودیت‌های ویژه‌ای دارند که تشخیص سره از ناسره را سخت می‌کند و در عین حال، امکان‌هایی را فراهم می‌آورد که از یک طرف شانس را در کار دخیل می‌کند و از طرف دیگر، «مسیر رسیدن به هدف» را بی‌ارزش جلوه می‌دهد و نتیجه مهم‌تر از فرایند می‌شود. البته چنین آزمون‌هایی هم گران تولید می‌شوند و هم گران تصحیح و گزارش می‌گردند اما کشورها چاره‌ای جز رفتن به سمت این انتخاب و فرار کردن از

## تلنگرهای ارزشیابی

**یک اصل عمومی برای تعیین زمان مناسب بازخورد، این است که معلم خود را به جای دانش‌آموز بگذارد و دریابد که چه زمانی مایل است معلم به او بازخورد دهد. بدیهی است بازخورد زمانی مفید واقع می‌شود که ذهن دانش‌آموز به شدت درگیر کار است و هنوز فرصتی برای انجام دادن کار یا اصلاح آن دارد.**

لحظه‌ای از این زندگی در آخر سال و در یک امتحان یک ساعته؛

روند رو به افزایش نیاز به ارزشیابی‌های بیشتر و بیشتر برای مقاصد مختلف (ورود به دانشگاه، دیپلم گرفتن، اشتغال، تأمین نیروی انسانی...)؛ بنابراین امروزه تعداد بیشتری از آدم‌ها در آزمون شرکت می‌کنند و لازم است این آزمون‌ها «اعتبار پیش‌بین»<sup>۵</sup> مطلوبی داشته باشند؛ چرا که همه موارد فوق برای پیش‌بینی است.

**۴) فساد در ارزشیابی:** گرجستان در چند سال گذشته چالشی جدی برای برگزاری آزمون نهایی دیپلم داشته است: اعتبار سؤالات آزمون، فساد ایجاد شده در آزمون (هم در بعد لو رفتن سؤالات و هم در بعد ایجاد بازار نمره توسط معلمان خبره!)، و میانگین پایین کشوری در دروس اصلی دیپلم. بنابراین، وزارت آموزش و پرورش تصمیم گرفت آزمون را با حمایت بخش‌های خصوصی و مؤسسات علمی - بین‌المللی برگزار نماید؛ به طوری که فساد ایجاد شده جمع گردد، دروس بیشتری را شامل شود، در زمان کوتاهی در داخل مدارس اجرا گردد، و نسبت هزینه به فایده آن منطقی باشد.

**۵) ارزشیابی به عنوان اهرم سیاست‌گذاری:** این عبارت بسیار گفته و شنیده می‌شود اما در عمل این گونه نشده است و سیاست‌گذاران نتایج ارزشیابی را هم مصادره به مطلوب می‌کنند. بنابراین، روشنفکران برخی کشورها دست به شورش اجتماعی و تحرکی ملی در این خصوص زده‌اند. نارضایتی از مدارس و تجمع و اخبار تلویزیون و روزنامه‌ها نه به خاطر شهریه، عدم امکانات مدرسه‌ای و کمبود معلم، بلکه به خاطر نارضایتی از کیفیت خروجی‌های مدارس - یعنی دانش‌آموزان - است. چنین مطالباتی، نه تنها مطلوب و ممدوح است بلکه چراغی

ورطه هولناک سؤالات چهار جوابی ندارند. برای همین، چنین آزمون‌هایی نباید به راحتی و بلافاصله در دسترس قرار گیرند و امکان خرید و فروش پیدا کنند.

## ۲) فناوری‌های پیشرفته در حوزه سنجش، تصحیح

اوراق الکترونیکی، گزارش‌دهی الکترونیکی و به خصوص تصحیح و نمره‌دهی آزمون‌های تشریحی در دنیا کشف و خلق شده است آموزش و پرورش و بخش‌های خصوصی از آن‌ها برای تولید و تنوع بخشی به آزمون‌ها و همچنین تصحیح در تعداد کلان بهره می‌برند. این فناوری‌ها گرچه به نهایت صحت و بهره‌وری نرسیده‌اند، تحقیقات نشان می‌دهد که هم‌راستایی و همبستگی بالایی بین آزمون‌های تشریحی تصحیح شده توسط ماشین و انسان وجود دارد و آن میزان غیرهم‌راستایی را می‌توان در قبال آورده‌های ذی‌قیمت آن روش، نادیده گرفت و آن را با تصحیح انسانی تکمیل کرد. کشور هلند در این زمینه پیشرفت‌های قابل ملاحظه و در خور توجهی داشته است.

یکی از درس‌آموخته‌های هلندی‌ها این است که نمی‌توان «ساختار سازمانی و کیفیت نیروی انسانی» سازمانی را که تاکنون آزمون قلم و کاغذی تولید و اجرا و تصحیح و گزارش می‌کرده، برای سازمانی که قرار است ارزشیابی الکترونیکی (e - assessment) برگزار نماید، حفظ کرد؛ چرا که اساساً این سازمان جدید فقط متشکل از یک گروه یکدست آموزشی (معلم) نیست بلکه کار گروهی است و متخصصان آموزش، فناوری، نرم‌افزار، امنیت، چند رسانه‌ای، مدارس، وزارت آموزش و پرورش، و بخش خصوصی و دانشگاهی در کنار هم این هدف را محقق می‌سازند و نیازمند تغییر الگوی سازمانی است. تجربه هلندی‌ها نشان می‌دهد که اگرچه آزمون‌های الکترونیکی در ابتدا زمان‌بر و انرژی‌بر هستند، در بلندمدت هم حجم کاری را کاهش می‌دهند و هم امکان سنجش شایستگی‌هایی را فراهم می‌آورند که ارزش افزوده آموزش و پرورش است.

**۳) روندها:** در کشورها بعضی روندها قابل مشاهده است؛ در محتواهای ارائه شده بیشتر کشورهای شرکت‌کننده در این کنفرانس، از این چرخش‌ها در سنجش و اندازه‌گیری نام برده می‌شد که برخی از آن‌ها قابل تأمل‌اند؛ مثلاً:

۱) چرخش از سنجش حقایق و دانسته‌ها به سنجش مهارت‌ها و عملکردها؛

۲) چرخش به سنجش محتوا و مهارت‌هایی که در کتاب‌های درسی کمتر به آن‌ها توجه می‌شود؛

۳) چرخش به سنجش «چیزهایی» که خود معلم به آن‌ها مجهز نیست؛

۴) تغییر نقش ارزیاب و ارزیابی، نقش بازرسی و بازرسی، نقش ناظر و نظارت؛

۵) چرخش به ارزشیابی عملکردی و سیاهه‌ای (فیلم گرفتن طولانی مدت از زندگی یادگیری بچه‌ها به جای عکس گرفتن

نمی‌توان «ساختار سازمانی و کیفیت نیروی انسانی» سازمانی را که تاکنون آزمون قلم و کاغذی تولید و اجرا و تصحیح و گزارش می‌کرده، برای سازمانی که قرار است ارزشیابی الکترونیکی (e - assessment) برگزار نماید، حفظ کرد

ارزشیابی عملکرد معلم در مقایسه با دانش آموز.

## پنج درس آموخته برای ایران:

### ۱) در حوزه ارزشیابی و سنجش آموزش انجمن علمی

**فعال در ایران وجود ندارد.** فعال بودن یک انجمن تا حد زیادی به حضور و فعالیت چند نفر انگشت‌شمار و بسیار آسیب‌پذیر وابسته است. یک پیشنهاد می‌تواند این باشد که انجمن‌های پراکنده علوم تربیتی، با هم یکی شوند و توان‌های پراکنده خویش را یک کاسه کنند و یک انجمن علوم تربیتی قوی و هماهنگ ایجاد نمایند که سالانه یک کنفرانس بزرگ، وزین و پرستار داشته باشد و در طول سال هم در قالب «گروه‌های هم‌علاقه» به حوزه‌های زیرمجموعه علوم یادگیری و آموزشی بپردازند. نمونه‌های بین‌المللی چنین انجمن‌هایی هم وجود دارد و توفیقشان نشان داده شده است.

### ۲) سنجش و ارزشیابی آموزشی در کشور ما به یک

#### حوزه تخصصی مبدل نشده است؛ یعنی ارزشیابی در

نظام دولتی آموزش و پرورش تماماً معادل امتحانات نهایی (در مرکز سنجش) و آزمون تیمز و پرلز در پژوهشگاه مطالعات آموزش و پرورش شده است. در مدرسه هم، انگاره غالب از ارزشیابی، امتحانات میان ترم و پایان ترم است و اخیراً هم ارزشیابی توصیفی را به زبان می‌آورند، اما در بخش خصوصی ایران، ارزشیابی رشد (و نه توسعه) عجیبی یافته؛ یعنی هم پول آور است و هم تغییر جهت‌دهنده. بخش خصوصی ارزشیابی در ایران تقریباً پیشروترین بخش خصوصی آموزش در کشور و شاید در بیرون از این کشور باشد. منظور از پیشرو همان رشد کمی است (تعداد دانش‌آموزانی که در آزمون‌های مرحله‌ای شرکت می‌کنند و میزان درآمدی که کسب می‌شود).

دو مجله علمی - پژوهشی هم در کشور با موضوع اندازه‌گیری، سنجش و ارزشیابی آموزش چاپ می‌شوند. بررسی مقالات آن‌ها نشان می‌دهد که نویسندگانشان به تبدیل «ارزشیابی

است کم سو از اینکه مطالبات والدین / جامعه پوست‌اندازی کرده و از حد نمره در آزمون و رتبه در کنکور گذشته است. سیاست دولت‌ها هیچ‌گاه بدون مطالبه «درست» مردمی اصلاح نمی‌شود.

### ۶) اعتماد عمومی به نمره به‌عنوان حکم قاضی (معلم)

**است** و هر گاه این اعتماد، سلب شود، کل نظام قضایی (ارزشیابی) زیر سؤال می‌رود.

اگر فساد در قضاوت وارد شود و نمره، معرف توان فرد نباشد و جایگاه بعدی فرد نشان دهد که آن حکم و تصمیم (قبولی یا رد در امتحان) نادرست بوده است، دیگر برای رخنه در حکم معلم و گسترش فساد سنگ روی سنگ بند نمی‌شود. کسب مجدد اعتماد مردم در طولانی مدت تحقق می‌پذیرد.

### ۷) سنجش ۱۸۰ درجه: اصلی وجود دارد

که بر مبنای آن، اهداف برنامه درسی اگر در ارزشیابی - با تناظر یک به یک - متجلی نشوند، به احتمال بسیار زیاد معطل می‌مانند؛ تخمین زده می‌شود که سنجش فعلی در کشورها، حدود ۴۵ درجه از ۱۸۰ درجه اهداف را هدف گرفته‌اند و این ۴۵ درجه شامل همان اهدافی است که راحت‌تر به چنگ سنجش و تصحیح و نمره‌دهی می‌آیند و اغلب از جنس اهداف دانشی هستند. کشورهای حاضر در این کنفرانس هم برای این ۱۳۵ درجه مغفول کار را تمام نکرده بلکه آغاز کرده‌اند. این امور با توان بخش دولتی سازگار نیست و معمولاً بخش‌های خصوصی به اتفاق بخش‌های دانشگاهی وارد عرصه شده‌اند و کار را با تعاون با بخش دولتی به جایی رسانده‌اند. در ایران بخش خصوصی سنجش و ارزشیابی کشور توان مالی برای سرمایه‌گذاری در این بخش را دارد اما یا خود را بی‌نیاز از دانشگاه می‌انگارد و یا دولتی‌ها اجازه ورود به آن را نداده‌اند یا تمایلی به این بخش نمی‌بینند؛ چون برگشت مالی در آن مشاهده نمی‌کند؛ بنابراین، انرژی و ثروت خویش را در بخش خصوصی صرف می‌کند که همان ۴۵ درجه اهداف را بیشتر لحاظ می‌کند.

۸) **حوزه‌های جدیدی از ارزشیابی و سنجش که فناوری**

هم به کمک آن‌ها آمده و در این کنفرانس از آن‌ها یاد شد، عبارت‌اند از:

ارزشیابی عملکرد مجازی

آزمون چند مرحله‌ای

ارزشیابی بهبود یافته با فناوری، برای مهارت‌های قرن ۲۱

نمره‌دهی خودکار جستار

استفاده از تحلیل یادگیری و مشاهده دروس برای ارزشیابی

## بخش خصوصی

### سنجش و

ارزشیابی در ایران

می‌تواند ظرفیت

توسعه بخشی را

برای ارزشیابی در

کشور ایجاد نماید

و در کنار درآمدی

که در این حوزه

کسب می‌کند،

به صورت علمی

و تخصصی وارد

این حوزه شود و

حتی دانشگاه و

آموزش و پرورش را

حمایت کند؛ زیرا

منابعی در اختیار

دارد که این دو

بخش ندارند



### تلنگرهای ارزشیابی

تأثیر بازخورد توصیفی زمانی که با نمره و رتبه قضاوتی همراه است، کم می‌شود. گاهی قصد شما توصیف وضع دانش‌آموز است ولی او بیشتر آن را قضاوتی می‌بیند تا توصیفی. این امر به‌خصوص در مورد دانش‌آموزانی که چندان موفق نیستند، محسوس‌تر است. آنان معمولاً از هر نوع بازخوردی، مفهوم «تو نمی‌فهمی» را برداشت می‌کنند. در مورد چنین دانش‌آموزانی، بازخورد شما باید بیشتر بر نقاط قوت و پیشرفت آنان تأکید کند؛ حتی اگر این پیشرفت در حداقل لازم هم نباشد، به همراه آن یکی دو گام بعدی و امکان‌پذیر برای آنان تعیین کنید که در بازخورد بعدی بتواند با توجه به موفقیت‌هایشان، بازخورد امیدوارکننده بدهند.

است و تصور می‌کند صواب این است که با آن‌ها مقابله بریزد، عقل سلیم حکم می‌کند از توان مالی آن‌ها استفاده شود و دانش تخصصی ارزشیابی آموزشی در کشور شکل گیرد. این رشته هم‌گرا است و هم فناوری‌خواه؛ لذا به سرمایه‌گذاری نیاز دارد و از طرفی دولت هم چنین سرمایه‌ای ندارد. پس چه بهتر که در اینجا مشارکتی صورت گیرد و پشتوانه مالی، فناورانه و دانشی بخش خصوصی، به کمک بخش دولتی بیاید؛ یعنی آموزش و پرورش از "عجز ارزشیابی درست به سراغ" حذف این رکن رفته و این ناصواب است.

**۵) ارزشیابی با کمک گرفتن از توان فناوری اطلاعات و ارتباطات:** چنانچه خواهیم در ایران در جهت ارزشیابی الکترونیکی<sup>۸</sup> و کلان داده<sup>۹</sup> حرکت کنیم، چاره‌ای جز پوست‌اندازی و تغییر الگوی ذهنی، هم در بخش خصوصی ارزشیابی و هم در بخش دولتی که شامل آموزش و پرورش و دانشگاه است، نداریم. این پوست‌اندازی در درجه اول، ناظر بر توان کار گروهی و نشستن سر یک میز و گفت و شنود است و در درجات بعد، به بازنگری و بازاندیشی درباره آنچه انجام داده‌ایم و آنچه بدان دامن زده‌ایم، نیاز دارد.

\* پی‌نوشت‌ها

1. Jo Ann \_ Baird
2. Hariharan Swaminathan
3. Connecticut
4. Steve Rivkin
5. predictive validity
6. Special interest group
7. PIRLS and TIMSS
8. e \_ assessment
9. big data

به مطالعه «علاقه چندان ندانند. یعنی تعداد کمی از مقالات به تحلیل داده‌های ثانویه و نمایش روندها در سال‌های متمادی که آزمون‌های کلان برگزار شده، پرداخته‌اند. اخیراً برخی بخش‌های خصوصی حوزه سنجش در کشور اقدام به همکاری‌های بین بخشی با دانشگاه‌ها، استادان و دانشجویان حوزه آموزش سنجی، ارزشیابی و اندازه‌گیری کرده‌اند که به نشست‌های علمی مشترک با دانشگاه‌ها و انجام پروژه‌های مشترک انجامیده است و می‌تواند نوید توسعه هر سه بخش خصوصی، دانشگاهی و سیاست‌گذاری باشد.

### ۳) عضویت در مجامع و انجمن‌های بین‌المللی و شرکت

**فعال در رویدادهای آن‌ها:** خیلی خوب است که بخش‌های خصوصی، دولتی و دانشگاهی ایران هم عضو این مجامع شوند و هم در کنفرانس‌های آنان فعالانه شرکت کنند. منظور از شرکت فعال هم، همان است که از سایر کشورها می‌بینیم: ارائه سخنرانی از یافته‌های بومی و نه اینکه دوباره آنچه را غربی‌ها در کتاب‌هایشان درباره سنجش و ارزشیابی گفته‌اند و ما آن کتاب‌ها را در ایران خوانده‌ایم، در قالب اسلاید و پاورپوینت به خودشان برگردانیم؛ عقد تفاهم‌نامه‌های همکاری؛ دعوت از آن‌ها برای ارائه کارگاه سخنرانی در ایران و هم ایجاد فضایی برای ارائه دستاوردهای آموزش ایران در کشورهای دیگر، تولید آزمون‌های مشترک، انجام دادن مطالعات مشترک روی داده‌های ثانویه، و مطالعات تطبیقی بین‌المللی با استفاده از داده‌های مشابه.

### ۴) بخش خصوصی سنجش و ارزشیابی در ایران

می‌تواند ظرفیت توسعه بخشی را برای ارزشیابی در کشور ایجاد نماید و در کنار درآمدی که در این حوزه کسب می‌کند، به‌صورت علمی و تخصصی وارد این حوزه شود و حتی دانشگاه و آموزش و پرورش را حمایت کند؛ زیرا منابعی در اختیار دارد که این دو بخش ندارند. مخصوصاً در شرایط کنونی که آموزش و پرورش، پرورش، سرستیز با مؤسسات آموزشی برداشته

