

بررسی اعتبار نمره گذاری معلمان

«ارزشیابی تحصیلی»

معرفی مقاله

دکتر حمزه گنجی

هدف از این بررسی نشان دادن تفاوت‌های فردی معلمان در ارزشیابی اوراق امتحانی بوده است. در این مقاله، محقق ابتدا مفاهیم اندازه‌گیری و ارزشیابی را توضیح می‌دهد و اختلاف‌های ارزشیابان را در موقعیتهای واقعی بر می‌شمارد و به مواردی از قبیل اختلاف آنان از نظر درصد قبول‌شدگان، میانگین نمرات، پراکندگی نمرات، طبقه‌بندی اوراق اشاره می‌کند. سپس به علل وجود اختلاف در ارزشیابی می‌پردازد و به مواردی مانند اختلاف در شخصیت ارزشیابها و اختلاف در موقعیت ارزشیابی اشاره می‌نماید. در پایان نیز نویسنده مقاله به تشریح نتایج آزمایشی که در دانشکده روان‌شناسی دانشگاه علامه طباطبایی در مورد تصحیح اوراق زبان انگلیسی انجام داده است می‌پردازد. در این آزمایش نیز تأیید می‌شود که وقتی ارزشیابها در موقعیتهای متفاوت قرار گیرند به صورتهای مختلف قضاوت می‌کنند. ایشان نتیجه‌گیری می‌کنند که تفاوت موقعیتهای می‌تواند یکی از علل وجود اختلاف ارزیابها در امر ارزشیابی باشد. موفقیت این پژوهشگر را در انجام تحقیقات بیشتر آرزو مندیم.

«فصلنامه»

هدف از انجام این بررسی که در بهمن ماه ۱۳۶۶، در سمینار روان‌شناسان شاغل در وزارت بهداشتی ارائه شد، نشان دادن این واقعیت است که معلمان، در ارزشیابی اوراق امتحانی دانش‌آموزان تفاوت‌های فردی قابل ملاحظه‌ای دارند. به سخن دیگر، نمراتی را که معلمان به اوراق امتحانات تشریحی می‌دهند فاقد اعتبار است. قبل از آن که نتایج حاصل از این بررسی ارائه شود لازم است ابتدا معنای دقیق ارزشیابی^۱ را بفهمیم. برای درک معنای دقیق ارزشیابی قبلاً باید مفهوم اندازه‌گیری^۲ را روشن کنیم.

۱ - اندازه‌گیری

اندازه‌گیری یعنی عملی که به وسیله آن تعداد واحدهای موجود در داخل مقدار معینی از جنس همان واحد تعیین می‌شود. مانند عملی که توسط آن تعداد کیلوهای موجود در داخل یک کیسه برنج، تعداد لیترهای موجود در داخل یک بشکه نفت، تعداد مترهای موجود در طول و عرض یک اتاق، تعداد گرمهای موجود در یک دستبند یا در یک انگشتری و... تعیین می‌شود. بنابراین می‌توان گفت که اندازه‌گیری یعنی مقایسه دو مقدار همجنس.

۲ - ارزشیابی

ارزشیابی یعنی عملی که به وسیله آن درباره یک واقعه، یک فرد یا یک شی، با توجه به یک معیار^۳ یا چند معیار قضاوت می‌شود. ارزشیابی معمولاً بعد از اندازه‌گیری به عمل می‌آید و فرقی عمده‌اش با آن این است که دارای قضاوت است. در اندازه‌گیری هیچ نوع قضاوت و اظهار نظر شخصی وارد عمل نمی‌شود. در اندازه‌گیری، با وزن چند کیسه برنج، مساحت چند اتاق یا چند خانه، تعداد پاسخهای صحیح چند ورقه، وزن چند دستبند یا چند انگشتری را تعیین می‌کنیم. اما در ارزشیابی، از میان آنهایی که اندازه یکسانی دارند، بر اساس یک یا چند معیار، یکی را انتخاب می‌کنیم. مثلاً از دو کیسه برنجی که هر دو ۳۰ کیلو وزن دارد، یکی را به دیگری بر اساس معیار یا معیارهایی که داریم ترجیح می‌دهیم. از میان دو اتاقی که مساحت یکسانی دارند، از میان دو ورقه‌ای که نمره یکسانی را برده‌اند و... یکی را بنا به عللی به دیگری ترجیح می‌دهیم. به این ترتیب می‌توان گفت:

ارزشیابی یعنی اندازه‌گیری + قضاوت بر اساس یک یا چند معیار

با توجه به آنچه گفته شد، موضوع قضاوت و معیار ارزشیابی هر چه باشد در نفس عمل ارزشیابی تغییری به وجود نمی‌آید. بنابراین ارزشیابی حوزه بسیار گسترده‌ای دارد که نمونه‌های فراوانی را می‌توان برای آن ذکر کرد: دادن جایزه به فیلمها، انتخاب یک محصول از میان انواع مختلف همان محصول، انتخاب یک ورزشکار از میان چند ورزشکار، انتخاب یک کتاب از میان دهها کتاب، انتخاب یک ورقه از میان صدها ورقه، انتخاب یک دانشجو از میان هزاران دانشجو...

هریک از زمینه‌های فوق و صدها زمینه دیگر را می‌توان مورد بررسی قرار داد و تفاوت‌های افراد را در ارزشیابی آنها مطالعه کرد. اما از آنجا که حوزه فعالیت ما بیشتر زمینه تحصیلی را در بر می‌گیرد، بنابراین تلاش خود را به مطالعه «ارزشیابی تحصیلی»^۴ محدود می‌کنیم و از آنجا که کتاب حاضر «روانشناسی تفاوت‌های فردی» را در بر می‌گیرد. لذا در این بخش نیز سعی می‌کنیم ارزشیابی

تحصیلی را از دیدگاه روان‌شناختی مورد بررسی قرار دهیم و ببینیم مسلمان مدارس، استادان دانشگاهها و بطور کلی کلیه کسانی که درباره نوشته‌ها و آموخته‌های تحصیلی دیگران قضاوت می‌کنند در این مورد چه رفتاری از خود نشان می‌دهند و چه تفاوت‌هایی با یکدیگر دارند. لذا مطالب زیر «روان‌شناسی ارزشیابی تحصیلی» را در بر خواهد گرفت و تفاوت‌های افراد را در این امر نشان خواهد داد.

«روان‌شناسی ارزشیابی تحصیلی» یعنی مطالعه رفتار معلمان به هنگام تصحیح اوراق دانش‌آموزان. به عبارت دیگر، معلمی که ورقه دانش‌آموزی را تصحیح می‌کند یا استاد دانشگاهی که اوراق دانشجویان را نمره می‌دهد، در موقعیتی قرار می‌گیرد که باید رفتاری را از خود نشان دهد. منظور ما در این فصل مطالعه همین رفتار و نشان دادن تفاوت‌های فردی ارزشیابها در همان رفتار است. البته لازم به یادآوری است که در اینجا بیشتر ارزشیابی اوراق سوالات تشریحی مورد نظر است نه اوراق سوالات چند گزینه‌ای.

۳ - اختلافات ارزشیابها در موقعیتهای واقعی

فرض کنید در یکی از حوزه‌های امتحانات نهایی کلاسهای چهارم دبیرستان، که در خرداد ماه هر سال توسط وزارت آموزش و پرورش برگزار می‌شود، ۱۰۰۰ داوطلب وجود دارد که مواد مربوط به رشته تحصیلی خود را امتحان می‌دهند. باز فرض کنید که در همان حوزه برای تصحیح اوراق هر درس ده دبیر حضور دارند. بهترین و منطقی‌ترین راه برای توزیع اوراق بین تصحیح‌کنندگان این است که همه ۱۰۰۰ ورقه را بطور کاملاً تصادفی مخلوط کنند، آنها را به ده قسمت مساوی تقسیم و هر قسمت را در اختیار یک ارزشیاب قرار دهند تا براساس باری که در اختیار دارد آنها را ارزشیابی کند. در اینجا احتمال این که اوراق هر ده ارزشیاب با یکدیگر معادل باشند از نظر آمار بسیار زیاد است. به عبارت دیگر، طبق احتمال آماری، انتظار این است که میانگینهای نمرات ده ارزشیاب نزدیک به هم باشد. بدین معنا که اگر اختلافی هم بین آنها دیده شود، که مسلماً دیده می‌شود. نباید از حد معینی تجاوز کند. حد مجاز این اختلاف را نیز محاسبات آماری می‌تواند در اختیار ما بگذارد. مثلاً اگر میانگین نمرات یکی از ارزشیابها ۱۵ باشد ممکن است محاسبات آماری نشان دهد که نوسان میانگینها بین ۱۴ و ۱۶ مورد قبول خواهد بود.

اگر شما اجازه پیدا کنید و بتوانید در یکی از حوزه‌های امتحانی چنین کاری را به مرحله عمل در آورید خواهید دید که بین میانگینهای نمرات ارزشیابها تفاوت‌های بسیار چشمگیری وجود دارد. این تفاوتها معمولاً در حدی است که نمی‌توان آنها را به تصادف نسبت داد. مثلاً خواهید دید که میانگین نمرات یکی از ارزشیابها مثلاً ۱۲ و آن یکی دیگر ۱۸ است و...

مطالعات بسیار زیادی برای نشان دادن این نوع اختلافات در ارزشیابی اوراق انجام گرفته که ذکر همه آنها در اینجا امکان‌پذیر نیست. بنابراین ابتدا نمونه‌هایی از نتایج تحقیقاتی را که توسط دیگران انجام گرفته ارائه می‌دهیم و آنگاه بررسیهایی را که خود انجام داده‌ایم از نظر شما می‌گذرانیم. برای بررسی وجود تفاوت‌های فردی در ارزشیابی اوراق، همانطور که در بالا اشاره کردیم روش کار به این صورت است که نتایج امتحانات یک حوزه معین را برحسب ارزشیابهای مختلف با

یکدیگر مقایسه می‌کنند. واقعیت این است که وقتی اوراق بطور تصادفی و به تعداد تقریباً مساوی بین ارزشیابها توزیع می‌شود و خود ارزشیابها نیز بطور تصادفی از میان ارزشیابهای متعدد انتخاب می‌گردند، می‌توان بین نمرات ارزشیابهای مختلف حدود نوسان یا، به قول متخصصان آمار، حدود اطمینان را تخمین زد. اگر تفاوت‌های مشاهده شده خارج از حدود تخمین باشد اجباراً باید خود ارزشیابها را مسؤل این تفاوتها دانست.

با استفاده از روش فوق بارها ثابت شده است که بین ارزشیابها، مخصوصاً در مورد درصد قبول شدگان و میانگین نمرات آنها، تفاوت‌های معنی‌داری وجود دارد.

الف - اختلافات ارزشیابها از نظر درصد قبول شدگان: در هر امتحان عدالت حکم می‌کند که امتحان دهندگان، صرف‌نظر از این که اوراق آنها توسط چه کسی ارزشیابی شود، شانسه‌های مساوی برای قبول شدن داشته باشند. بدین معنا که درصدهای قبولی ارزشیابها، اگر هم به طور کامل متعادل نباشد، لااقل باید نزدیک به هم باشد. به عبارت دیگر، تفاوت درصد قبولی دو ارزشیاب نباید از نظر آماری معنی‌دار باشد.

بیرون^۵ یکی از روان‌شناسان فرانسوی، و همکاران او درصد قبولیهای ۱۳ ارزشیاب اوراق امتحانی رشته فلسفه و ۱۷ ارزشیاب اوراق امتحانی رشته ریاضیات را محاسبه کرده‌اند. اعداد به دست آمده در جدول شماره یک نشان داده شده است.

جدول ۱ - بالاترین و پایین‌ترین درصد قبولی توسط ارزشیابهای مختلف

انحراف	پایین‌ترین درصد قبولی	بالاترین درصد قبولی	تعداد ارزشیابها	رشته امتحانی
۲۲	۴۸	۷۰	۱۳	فلسفه
۲۲	۳۱	۵۳	۱۷	ریاضیات

مفهوم اعداد جدول شماره یک این است که در میان ۱۳ ارزشیاب اوراق رشته فلسفه، ارزشیابی بوده که به ۷۰ درصد اوراق نمره قبولی داده و در مقابل ارزشیاب دیگری بوده که به ۴۸ درصد همان اوراق نمره قبولی داده است. در مورد اوراق ریاضی این درصدها بترتیب عبارت است از ۵۳ و ۳۱. در هر دو مورد انحراف بین بالاترین و پایین‌ترین درصد قبولی ۲۲ است. اعداد فوق محققان را به این نتیجه‌گیری وادار می‌کند که نوسانات تصادفی هرگز موجب چنین تفاوت‌های چشمگیری نمی‌شود. تفاوت‌های فوق را باید ناشی از وجود تفاوت‌های فردی بین ارزشیابها دانست.

ب - اختلافات ارزشیابها از نظر میانگین نمرات: بیرون و همکاران او مقایسه جالب دیگری نیز در مورد نمرات ۱۳ ارزشیاب اوراق رشته فلسفه و ۱۷ ارزشیاب اوراق ریاضیات انجام داده‌اند. بدین ترتیب که میانگین کل نمرات هر ارزشیاب را برای هر ماده درسی حساب کرده‌اند. جدول شماره ۲، میانگینهای کل، بالاترین و پایین‌ترین میانگینها و انحراف آنها را نشان می‌دهد. بدیهی است که بعضی از این انحرافات، مخصوصاً در امتحانات شفاهی بسیار چشمگیر است. مثلاً در علوم طبیعی

رشته ریاضیات، انحراف میانگین نمرات سهل گیرترین آزماینده از میانگین نمرات سخت گیرترین آزماینده بیش از ۱۰ است^۶. انحراف بین آزمایندهگان امتحان شفاهی ریاضیات در رشته فلسفه نزدیک به ۹ است (جدول ۲).

جدول ۲ - میانگینهای نمرات کتبی و شفاهی در یکی از حوزه‌های امتحانی شهر باریس

انحراف	میانگینها			رشته فلسفه (۱۳ ارزشیاب)
	بایین ترین	بالا ترین	کل	
				کتبی:
۱/۳۰	۸/۲۰	۹/۵۰	۸/۹۰	فلسفه
۲/۷۰	۸/۷۰	۱۱/۴۰	۱۰/۴۰	فیزیک
۳/۹۰	۷/۷۰	۱۱/۶۰	۱۰/۴۰	علوم طبیعی
				شفاهی:
۲/۶۰	۹/۵۰	۱۲/۱۰	۱۰/۹۰	فلسفه
۸/۸۰	۴/۰۰	۱۲/۸۰	۹/۵۰	ریاضیات
۵/۵۰	۹/۵۰	۱۵/۰۰	۱۲/۶۰	فیزیک
۳/۱۰	۱۰/۴۰	۱۳/۵۰	۱۲/۲۰	تاریخ و جغرافیا
۵/۴۰	۹/۰۰	۱۲/۲۰	۱۲/۲۰	علوم طبیعی
۲/۸۰	۱۰/۶۰	۱۳/۴۰	۱۱/۹۰	زبان خارجه
				رشته ریاضیات (۱۷ ارزشیاب)
				کتبی:
۳/۲۵	۵/۸۱	۹/۰۶	۷/۹۸	ریاضیات
۲/۳۰	۷/۰۰	۹/۳۰	۸/۲۰	فلسفه
۲/۸۷	۷/۶۳	۱۰/۵۰	۹/۳۷	فیزیک
				شفاهی:
۲/۹۰	۹/۶۰	۱۲/۵۰	۱۱/۰۰	ریاضیات
۴/۷۰	۸/۳۰	۱۳/۰۰	۱۱/۳۰	فیزیک
۶/۲۰	۸/۱۰	۱۴/۳۰	۱۰/۷۰	فلسفه
۵/۱۰	۷/۸۰	۱۲/۹۰	۱۰/۷۰	تاریخ و جغرافیا
۱۰/۷۰	۴/۵۰	۱۵/۲۰	۱۲/۰۰	علوم طبیعی
۶/۳۰	۷/۹۰	۱۴/۲۰	۱۱/۵۰	زبان خارجه

زُرژ نوازه^۷، روان‌شناس دیگر فرانسوی نیز، انحرافات را شبیه آنچه پیرون و همکاران او مشاهده کرده‌اند، در مورد همان رشته‌ها و همان مواد کتبی و در سه سال متوالی به دست آورده است. قبل از پرداختن به نتایج او نتایج پیرون را بررسی می‌کنیم.

نگاه گذرا به اعداد جدول شماره ۲ و با در نظر گرفتن انحراف بالاترین و پایین‌ترین نمرات، جای تردید باقی نمی‌ماند که ارزشیابها در ارزشیابی اوراق امتحانی تفاوت‌های فردی غیر قابل گذشت دارند. البته در مورد قبول یا عدم قبول تفاوت‌های موجود بین نمرات ارزشیابها، از استدلال آماری کمک می‌گیرند. بدین ترتیب که در اطراف میانگین کل حدود اطمینان^۸ را حساب می‌کنند. یعنی حدودی را به دست می‌آورند که میانگینهای ارزشیابها قانوناً می‌تواند در آن محدوده نوسان داشته باشد. اگر میانگینهای ارزشیابها خارج از حدود اطمینان باشد نتیجه می‌گیرند که ارزشیابها اوراق را با شیوه‌های کاملاً متفاوت ارزشیابی می‌کنند.

نوازه^۹ در یکی از امتحانات سال چهارم دبیرستانها، حدود اطمینان را در سطوح ۵٪ و ۱٪ برای میانگینهای ارزشیابها محاسبه کرده است که ما به ذکر چند نمونه از آنها اکتفا می‌کنیم. در مورد درس فلسفه رشته ریاضیات میانگینهای نمرات ۱۹٪ ارزشیابها خارج از حدود اطمینان قرار گرفته است. در صورتی که نسبت قابل قبول ۵٪ بوده است. در مورد نسبت قابل قبول ۱٪ هم ۱۲٪ آنها خارج از حدود اطمینان بوده است. این محاسبات نشان داده که در همه موارد درصد ارزشیابهایی که خارج از حدود اطمینان قرار گرفته‌اند بسیار زیاد بوده و این درصد مخصوصاً در بعضی مواد کاملاً چشمگیرتر از مواد دیگر بوده است. مثلاً در امتحان فیزیک، در سطح اطمینان ۵٪، ۷۷٪ ارزشیابها خارج از حدود اطمینان قرار گرفته‌اند.

۴ - اختلافات ارزشیابها در موقعیتهای ساختگی

نتایج تحقیقات فوق و دهها تحقیق دیگر جای تردید باقی نمی‌گذارد که معلمان مدارس، استادان دانشگاهها و کلیه کسانی که در باره نوشته‌های دیگران قضاوت می‌کنند تفاوت‌های فردی بسیار زیادی را نشان می‌دهند. این تفاوتها در حدی است که سرنوشت صاحبان نوشته‌ها را دگرگون می‌سازد. نتایج به دست آمده بعضی روان‌شناسان را به این نتیجه‌گیری واداشته است که: برای پیش‌بینی نمره یک دانش‌آموز در یک امتحان، به جای شناختن خود او، بهتر است تصحیح کننده ورقه او را شناخت. تفسیر مطلب به این صورت است که معلمانی وجود دارند که همیشه جانب اغماص را پیش می‌گیرند و اکثر داوطلبان را قبول اعلام می‌کنند. در مقابل، معلمان دیگری وجود دارند که بسیار سخت گیرند و درصد کمتری را قبول می‌کنند. حال اگر شما این ارزشیابها را از قسبل بشناسید می‌توانید پیش‌بینی کنید که فلان دانش‌آموز قبول خواهد شد یا نه.

در تحقیقاتی که از آنها نام بردیم، ارزشیابهای مختلف اوراق مختلفی را ارزشیابی کرده‌اند. درست است که در این تحقیقات ارزشیابها بطور تصادفی انتخاب شده‌اند و اوراق نیز بطور تصادفی بین آنها توزیع گردیده است، نتایج به دست آمده نشان نمی‌دهد که چقدر از اختلافات موجود در نمرات ناشی از تفاوت‌های فردی ارزشیابها و چقدر از آنها ناشی از تفاوت‌های موجود بین خود اوراق است. برای رفع این ابهام روان‌شناسان سعی کرده‌اند شرایطی را فراهم آورند که در آنها ارزشیابهای

مختلف اوراق یکسانی را تصحیح کنند. یعنی سعی کرده‌اند ارزشیابها را در شرایط یکسانی قرار دهند تا اثر تفاوت‌های فردی خود آنها روشن شود.

روشی که در این خصوص به کار می‌رود اصطلاحاً روش چند تصحیحی^{۱۱} نامیده می‌شود. در این روش، ارزشیابهای متعدد اوراق معینی را جدا از یکدیگر و در یک زمان تصحیح می‌کنند، یا یک ارزشیاب معین اوراق معینی را در زمانهای متعدد ارزشیابی می‌کند. بنابراین با این روش، هم می‌توان نشان داد که ارزشیابها با یکدیگر تفاوت دارند و هم می‌توان ثابت کرد که آنها در زمانهای مختلف به صورتهای مختلف قضاوت می‌کنند.

برای اثبات این که ارزشیابها در تصحیح اوراق داوطلبان تفاوت‌های فردی قابل توجهی دارند بررسی کوچکی را در دانشکده روان‌شناسی دانشگاه علامه طباطبائی انجام داده‌ایم که در زیر خلاصه‌ای از آن را درج می‌کنیم:

از یکی از دانشجویان که معلم تاریخ و جغرافیای کلاس پنجم ابتدایی بود خواستیم که کلیه اوراق امتحانی کلاس خود را در اختیار ما بگذارد. از میان اوراق، پنج ورقه بطور تصادفی انتخاب و هر یک در شش نسخه تکثیر گردید. اوراق تکثیر شده به همراه بارم سؤالات در اختیار شش معلم تاریخ و جغرافیای کلاس پنجم ابتدایی قرار گرفت تا مورد ارزشیابی قرار گیرند. ارزشیابها همه مرد بودند و هیچ اطلاعی از یکدیگر نداشتند. (جدول شماره ۳). نمرات خام^{۱۲} هر ارزشیاب، میانگین نمرات آنها، بالاترین و پایین‌ترین نمره و نیز انحراف بین آنها را نشان می‌دهد. همچنین جدول مزبور نمرات داده شده به یک ورقه، میانگین آنها، بالاترین، پایین‌ترین و نیز انحراف آنها را نشان می‌دهد.

در تحقیقاتی که از روش چند تصحیحی استفاده می‌شود، معمولاً هدف این است که نمرات ارزشیابها برحسب میانگینها، پراکندگیها و طبقه‌بندی‌ها مقایسه شود. میانگین نمرات هر ارزشیاب وضعیت او را در مقیاس سخت‌گیری - سهل‌گیری نشان می‌دهد. پراکندگی نمرات (که معمولاً با واریانس نشان داده می‌شود)، شیوه ارزشیاب در بهره‌گیری از مقیاس نمره‌گذاری را تعیین می‌کند: آیا ارزشیاب از همه اعداد مقیاس بهره می‌گیرد (از صفر تا ۲۰) یا از قسمتی از آن (مثلاً بین ۱۰ و ۱۷)؟ بدیهی است که این دو ویژگی از یکدیگر مستقل هستند. ممکن است دو ارزشیاب از نظر میانگین نمراتی که داده‌اند اختلاف نداشته باشند اما از نظر پراکندگی نمرات کاملاً تفاوت نشان دهند (مثلاً عدد ۱۲ ممکن است میانگین دو عدد ۱۰ و ۱۴، ۱۶ و ۶، ۱۸ و ۱۰... باشد). حال ۶ ارزشیاب شرکت کننده در تصحیح پنج ورقه را از لحاظ میانگین پراکندگی و طبقه‌بندی نمراتی که به اوراق داده‌اند مقایسه می‌کنیم و تفاوت‌های فردی آنها را در هر یک از سه مورد روشن می‌سازیم.

الف - اختلافات ارزشیابها از نظر میانگینها: جدول ۳ میانگین کل نمراتی را که هر یک از ۶ ارزشیاب به ۵ ورقه داده است نشان می‌دهد (۱۶/۲۰، ۱۲/۴۵، ۱۵/۱۵، ۱۴/۴۰، ۱۶/۳۵ و ۱۴/۴۰). این اعداد میزان سخت‌گیری و سهل‌گیری ارزشیابها را مشخص می‌کنند. از مقایسه ۶ میانگین معلوم می‌شود که ارزشیاب ۲ با میانگین ۱۲/۴۵ سخت‌گیرترین و ارزشیاب ۵ با میانگین ۱۶/۳۵ سهل‌گیرترین آنها هستند. انحراف میانگینهای این دو نفر ۳/۹۰ یعنی نزدیک به ۴ است. ارزشیاب ۲ در همه موارد کمتر از دیگران نمره داده است. بنابراین در یک حوزه امتحانی اگر این دو نفر جزء تصحیح کنندگان اوراق باشند، اوراقی که توسط نفر دوم ارزشیابی شوند رو بهمرفته نمراتی کمتر از

جدول ۳ - نمرات خام، میانگین نمرات، بالاترین نمره، پایین ترین نمره و نیز بالاترین انحراف
 بین نمراتی که یک ارزشیاب داده یا نمراتی که یک ورقه دریافت کرده است.

انحراف	پایین ترین نمره	بالاترین نمره	میانگین	ارزشیابها						اوراق	
				۶	۵	۴	۳	۲	۱		
۲/۲۵	۱۲/۲۵	۱۵/۵۰	۱۳/۸۳	۱۴/۵۰	۱۵/۰۰	۱۳/۵۰	۱۲/۲۵	۱۲/۲۵	۱۲/۲۵	۱۵/۵۰	ورقه ۱
۲/۵۰	۱۳/۵۰	۱۷/۰۰	۱۵/۷۰	۱۴/۷۵	۱۶/۵۰	۱۶/۰۰	۱۶/۵۰	۱۳/۵۰	۱۳/۵۰	۱۷/۰۰	ورقه ۲
۲/۵۰	۱۴/۰۰	۱۷/۵۰	۱۶/۱۶	۱۶/۲۵	۱۶/۷۵	۱۶/۰۰	۱۶/۵۰	۱۶/۰۰	۱۴/۰۰	۱۷/۵۰	ورقه ۳
۵/۰۰	۱۱/۲۵	۱۶/۲۵	۱۳/۷۰	۱۳/۲۵	۱۶/۲۵	۱۳/۰۰	۱۲/۰۰	۱۲/۰۰	۱۱/۲۵	۱۵/۵۰	ورقه ۴
۶/۲۵	۱۱/۲۵	۱۷/۵۰	۱۴/۷۰	۱۳/۲۵	۱۷/۲۵	۱۳/۵۰	۱۷/۵۰	۱۷/۵۰	۱۱/۲۵	۱۵/۵۰	ورقه ۵
۳/۹۰	۱۲/۲۵	۱۶/۳۵	—	۱۴/۴۰	۱۶/۳۵	۱۴/۴۰	۱۵/۱۵	۱۲/۲۵	۱۲/۲۵	۱۶/۲۰	میانگین
—	—	—	—	۱۶/۲۵	۱۷/۲۵	۱۶/۰۰	۱۷/۵۰	۱۴/۰۰	۱۴/۰۰	۱۷/۵۰	بالاترین نمره
—	—	—	—	۱۳/۲۵	۱۵/۰۰	۱۳/۰۰	۱۲/۲۵	۱۲/۲۵	۱۱/۲۵	۱۵/۵۰	پایین ترین نمره
—	—	—	—	۳/۰۰	۲/۲۵	۳/۰۰	۵/۲۵	۲/۷۵	۲/۷۵	۲/۰۰	انحراف

اوراق دیگر کسب می‌کنند. اوراقی هم که توسط نفر پنجم ارزشیابی شوند رویهمرفته وضع بهتری نسبت به بقیه پیدا خواهند کرد. این که ورقه یک داوطلب معین به دست کدامیک از این دو نفر یا دیگران برسد صرفاً به شانس بستگی دارد.

ب - اختلافات ارزشیابها از نظر پراکندگی نمرات: گفتیم که میانگین نمرات یک ارزشیاب می‌تواند نشانه‌ای از میزان سخت‌گیری او باشد اما این میانگین مستقل از پراکندگی نمرات است. مثلاً می‌بینیم که ارزشیاب ۱، با این که از نظر میانگین نمرات بالاتر از ارزشیاب ۳ است، دامنهٔ تغییر کمتری را در نمره گذاری نشان می‌دهد (۲ در مقابل ۵/۲۵، ردیف آخر). وقتی دامنه تغییر نمرات همه ارزشیابها را در نظر می‌گیریم (۲، ۲/۷۵، ۵/۲۵، ۳، ۲/۲۵ و ۳) به این نتیجه می‌رسیم که ارزشیابها از نظر شیوه استفاده از مقیاس نمره گذاری نیز تفاوت‌های فردی نشان می‌دهند. وقتی ستون آخر جدول، یعنی انحراف بین بالاترین و پایین‌ترین نمرهٔ یک ورقه را در نظر می‌گیریم می‌بینیم که بالاترین انحراف در مورد ورقه ۵ دیده می‌شود. یعنی ارزشیابهای ۲ و ۵ ورقهٔ پنجم را با اختلاف ۶/۲۵ نمره داده‌اند. ورقهٔ پنجم با اختلاف ۵ نمره، ورقه‌های دوم و سوم با اختلاف ۳/۵۰ و ورقهٔ اول با اختلاف ۲/۲۵ نمره گذاری شده است.

ج - اختلافات ارزشیابها از نظر طبقه‌بندی اوراق: ممکن است دو ارزشیاب از نظر میانگین نمرات و دامنه پراکندگی آنها با یکدیگر هیچ تفاوتی نداشته باشند، اما به اوراق یکسان نمرات متفاوتی را بدهند. همچنین ممکن است دو ارزشیاب اوراق را به طور یکسان طبقه‌بندی کنند، اما از لحاظ میانگین و دامنهٔ تغییر نمرات تفاوت داشته باشند. جدول شماره ۴، که از جدول شماره ۳ استخراج شده است، به روشنی نشان می‌دهد که ارزشیابها از لحاظ طبقه‌بندی ۵ ورقه نیز اختلاف دارند.

جدول شماره ۴ - طبقه‌بندی ۵ ورقه توسط ۶ ارزشیاب

ارزشیابها						اوراق
۶	۵	۴	۳	۲	۱	
رتبهٔ سوم	رتبهٔ پنجم	رتبهٔ چهارم	رتبهٔ پنجم	رتبهٔ سوم	رتبهٔ چهارم	ورقه ۱
رتبهٔ دوم	رتبهٔ سوم	رتبهٔ دوم	رتبهٔ سوم	رتبهٔ دوم	رتبهٔ دوم	ورقه ۲
رتبهٔ اول	رتبهٔ دوم	رتبهٔ اول	رتبهٔ دوم	رتبهٔ اول	رتبهٔ اول	ورقه ۳
رتبهٔ پنجم	رتبهٔ چهارم	رتبهٔ پنجم	رتبهٔ چهارم	رتبهٔ پنجم	رتبهٔ پنجم	ورقه ۴
رتبهٔ چهارم	رتبهٔ اول	رتبهٔ سوم	رتبهٔ اول	رتبهٔ چهارم	رتبهٔ سوم	ورقه ۵

همانطور که جدول شماره ۴ نشان می‌دهد، در طبقه‌بندی اوراق توسط ارزشیابها نیز تفاوت‌های چشمگیری دیده می‌شود. مثلاً از ۶ ارزشیاب، دو نفر ورقهٔ اول را در رتبهٔ سوم، دو نفر در رتبهٔ چهارم و دو نفر در رتبهٔ پنجم قرار داده‌اند. در مورد ورقهٔ دوم، ۴ نفر از ارزشیابها آن را در رتبهٔ دوم و دو نفر دیگر در رتبهٔ سوم قرار داده‌اند که بترتیب ۶۷٪ و ۳۳٪ ارزشیابها را تشکیل می‌دهند. در مورد اوراق ۳ و ۴ نیز درصد توافق بین ارزشیابها ۶۷٪ است. اگر به رتبه‌بندی ورقهٔ پنجم خوب توجه کنیم می‌بینیم

که $\frac{1}{3}$ ارزشیابها آن را بهترین ورقه، $\frac{1}{3}$ دیگر آن را به عنوان ورقه متوسط و $\frac{1}{3}$ بقیه به عنوان یک ورقه تقریباً ضعیف در بین پنج ورقه به حساب آورده‌اند.

با توجه به آنچه در بندهای الف، ب و ج گفته شد، ارزشیابها می‌توانند به صورتهای مختلف تفاوت داشته باشند. از نظر میانگین نمراتی که می‌دهند، پراکندگی و طبقه‌بندی اوراق تفاوتی را نشان می‌دهند که نمی‌توان آنها را به تصادف نسبت داد. همانطور که در صفحات قبل اشاره کردیم، این تفاوتها در تصحیح همه اوراق تشریحی به چشم می‌خورد. اثر این اختلافات زمانی غم‌انگیز خواهد بود که بخواهند تنها بر اساس نمرات آنها افراد را برای استخدام یا ورود به دانشگاه انتخاب کنند، یا بورس تحصیلی در اختیار بعضی از آزمون دهندگان قرار دهند.

در بررسی دیگری، یکی از اوراق ترجمه انگلیسی دانشجویان را تکثیر کردیم و آنها را در اختیار ۱۸ معلم قرار دادیم تا به طور مستقل تصحیح کنند. نتایج به دست آمده به صورتی است که در جدول شماره ۵ می‌بینیم:

جدول شماره ۵ - تصحیح یک ورقه ترجمه انگلیسی توسط ۱۸ ارزشیاب

ارزشیابها																	
۱۸	۱۷	۱۶	۱۵	۱۴	۱۳	۱۲	۱۱	۱۰	۹	۸	۷	۶	۵	۴	۳	۲	۱
$\frac{11}{100}$	$\frac{10}{25}$	$\frac{16}{25}$	$\frac{15}{50}$	$\frac{16}{75}$	$\frac{11}{100}$	$\frac{13}{50}$	$\frac{8}{75}$	$\frac{17}{75}$	$\frac{13}{100}$	$\frac{13}{100}$	$\frac{13}{100}$	$\frac{14}{100}$	$\frac{14}{50}$	$\frac{14}{75}$	$\frac{15}{25}$	$\frac{12}{75}$	$\frac{10}{25}$

با توجه به جدول شماره ۵ معلوم می‌شود که ارزشیاب ۱۰ سهل‌گیرترین و ارزشیاب ۱۱ سخت‌گیرترین ارزشیابها هستند. این دو نفر بترتیب نمرات $\frac{17}{75}$ و $\frac{8}{75}$ داده‌اند که در واقع ۹ نمره اختلاف دارند. اگر معیار قبولی را ۱۰ در نظر بگیریم، ورقه مورد تصحیح از نظر ارزشیاب ۱۱ مردود و از نظر سایر ارزشیابها قبول خواهد بود. ممکن است تفاوت نمرات ارزشیابهای ۱، ۱۳، ۱۷ و ۱۱ را ناشی از تصادف دانست، اما تفاوت نمرات این ارزشیابها با نمرات ارزشیابهای ۳، ۱۰، ۱۴، ۱۵ و ۱۶ بدون تردید ناشی از تصادف نخواهد بود. بنابراین جای تردید باقی نمی‌ماند که ارزشیابها در تصحیح اوراق داوطلبان تفاوتی فردی غیر قابل اغماض دارند و این تفاوتها مستقل از محتوای ورقه مورد ارزشیابی است. بدین معنا که چون محتوای ورقه برای همه ارزشیابها یکسان بوده، بنابراین می‌توان گفت که اثر آن خنثی شده است. تنها عاملی که در تصحیح اوراق فرق می‌کرد خود ارزشیابهاست. ارزشیابها برحسب میزان تجربه، سن، میزان تحصیلات، شخصیت و عوامل متعدد دیگری می‌توانند نگرشهای متفاوتی در تصحیح اوراق داشته باشند.

جای تردید وجود ندارد که میزان تحصیلات معلمان در شیوه ارزشیابی آنها مؤثر خواهد بود. مثلاً دیپلمه‌ها متفاوت از فوق دیپلمه‌ها و فوق دیپلمه‌ها متفاوت از لیسانس‌ها ارزشیابی خواهند کرد. سن نیز می‌تواند عامل دیگری باشد. بدین صورت که معلمان جوان متفاوت از معلمان میانسال یا معلمان مسن نمره می‌دهند. البته جنس نیز می‌تواند عامل مهم دیگری به حساب آید. مثلاً اعتقاد اکثریت

بر این است که خانمها بیشتر از آقایان نمره می‌دهند. حتی این امر ممکن است در مورد دو جنس مختلف متفاوت باشد. بدین معنا که ارزشیابهای زن به اوراق دختران متفاوت از اوراق پسران نمره بدهند و ارزشیابهای مرد نیز به همین صورت. شخصیت نیز عاملی است که می‌تواند وارد عمل شود و تفاوت‌های فردی ارزشیابها را به وجود آورد. در صفحات آینده، به اجمال در مورد عوامل فوق‌الذکر بحث خواهیم کرد.

۵ - علل وجود اختلاف در ارزشیابی

در صفحات قبل، مخصوصاً در بررسیهایی که در آنها دسته‌مینی از اوراق توسط تعدادمینی از ارزشیابها تصحیح شده بود یا ورقه‌مینی را تعداد زیادی از افراد ارزشیابی کرده بودند، دیدیم که اختلافات فاحشی بین ارزشیابها وجود دارد. این اختلافات در مورد میانگینها، درصد قبولی اوراق، پراکندگی نمرات و حتی رتبه‌بندی داوطلبان دیده می‌شد. به این ترتیب جای تردید باقی نمی‌ماند که ارزشیابهای مختلف اوراق یکسانی را به صورتهای مختلف ارزشیابی می‌کنند. از طرف دیگر می‌توان ثابت کرد که یک ارزشیاب معین در زمانهای مختلف نمرات مختلفی را به یک ورقه‌معین می‌دهد. برای اثبات این قضیه می‌توان اوراق یکسانی را در زمانهای مختلف در اختیار ارزشیابهای معین قرار داد و نمرات آنها را در زمانهای مختلف مقایسه کرد. قبلاً از انجام این قسمت صرف‌نظر می‌کنیم و تنها به تشریح علل اختلافات می‌پردازیم.

الف - اختلاف در شخصیت ارزشیابها: همانطور که انسانها از نظر قد، وزن، چهره، مقاومت بدنی، هوش، استعداد و سایر جنبه‌ها اختلاف دارند از نظر جنبه‌های شخصیتی نیز متفاوت هستند. به همان ترتیب که افراد برونگرا، درونگرا، خونسرد، تندخو، سلطه‌جو، سلطه‌پذیر، خوش بین و بدبین وجود دارند، افراد سخت‌گیر و سهل‌گیر هم وجود دارند. معلمی داریم که بندرت بالاتر از ۱۷ و ۱۸ نمره می‌دهد. در مقابل معلم دیگری داریم که معمولاً کمتر از ۱۵ نمره نمی‌دهد. استادی داریم که کمتر نمره ۱۰ می‌دهد، در عوض استاد دیگری داریم که معمولاً کمتر از الف یا ب نمره نمی‌دهد. خانمهایی را می‌توان یافت که وقتی برای هشت نفر غذا می‌پزند چهار نفر هم سیر نمی‌شوند، در عوض خانمهای دیگری یافت می‌شوند که وقتی برای چهار نفر غذا می‌پزند نه تنها هشت نفر هم نمی‌توانند تمام کنند بلکه مقداری از آن را بیرون می‌ریزند. بعضی از افراد افزاینده‌اند بعضی دیگر کاهنده. بعضیها جانب افراط را می‌گیرند بعضی دیگر جانب تفریط را.

بنابراین اولین علت اختلاف در ارزشیابی اوراق، وجود تفاوت‌های فردی در شخصیت ارزشیابهاست که می‌توان آنها را در مقیاس سهل‌گیری، سخت‌گیری طبقه‌بندی کرد، همچنان که می‌توانیم افراد را در مقیاس درونگرایی - برونگرایی طبقه‌بندی کنیم. تیپ سخت‌گیر انعطاف‌پذیری کمتری از خود نشان می‌دهد، کمتر گذشت می‌کند و معمولاً جنبه‌های منفی را ملاک نمره‌گذاری قرار می‌دهد. تیپ سهل‌گیر جانب اغماض را در نظر می‌گیرد، گذشت بیشتری دارد و به دنبال نکات مثبت می‌گردد تا آنها را بهانه قرار دهد و به صاحب ورقه نمره بدهد. بدیهی است که هر دو تیپ در دو انتهای منحنی طبیعی قرار دارند و تعداد آنها بسیار کم است. تعداد ارزشیابهایی که در حد وسط این دو گروه جای می‌گیرند، به مراتب بیشتر از تیبهای دیگر است. در تصحیح اوراق، اگر یک ورقه متوسط تصادفاً

به دست یک ارزشیاب سهل گیر بیفتند نه تنها قبول خواهد شد، بلکه بیشتر از حد، نمره خواهد گرفت. اگر همان ورقه تصادفاً به دست یک ارزشیاب سخت گیر برسد نه تنها نمره کمتر از حد، دریافت خواهد کرد بلکه به احتمال زیاد مردود خواهد شد و بالاخره اگر توسط یک ارزشیاب متعادل ارزشیابی شود نمره‌ای نزدیک به نمره حقیقی خود دریافت خواهد کرد.

ب - اختلاف در موقعیت ارزشیابی: اختلاف در ارزشیابی اوراق امتحانی مسلماً از طریق متغیرهای شخصیت تعیین می‌شود، اما این اختلاف مخصوصاً توسط متغیرهای موقعیت نیز توجیه می‌گردد. به عبارت دیگر، عامل دیگری که موجب می‌شود تا دو ارزشیاب مختلف در مقابل یک کار معین با روشهای متفاوت عمل کنند و به نتایج متفاوتی برسند این است که آنها در شرایط متفاوتی قرار می‌گیرند و در نتیجه اطلاعات متفاوتی را بررسی می‌کنند.

بنابراین، براساس آنچه در سطور بالا گفته شد، ارزشیابی یعنی پاسخگویی به یک موقعیت. طبق معمول پاسخ را با R نشان می‌دهیم. این پاسخ از یک طرف تابع^{۱۲} شخصیت ارزشیاب و از طرف دیگر تابع اطلاعاتی است که او به هنگام تصحیح اوراق، یا حتی قبل از شروع به تصحیح دریافت می‌کند. اگر متغیر اول، یعنی شخصیت ارزشیاب را با P و متغیر دوم، یعنی موقعیت ارزشیابی را با S نشان دهیم، ارزشیابی تابعی خواهد بود از کنش متقابل بین شخصیت ارزشیاب و موقعیت ارزشیابی:

$$R=f(P \leftrightarrow S)$$

وجود اختلاف در شخصیت ارزشیابها واقعیتی است که با گذشته آنها ارتباط پیدا می‌کند. یعنی ارزشیابها بنا به تجربه‌های قبلی خود و احتمالاً بنا به آمادگیهای ذاتی خود یاد گرفته‌اند که سهل گیر یا سخت گیر یا در فاصله‌ای از این دو آنها قرار گیرند. البته جای تردید وجود ندارد که می‌توان ارزشیابها را آموزش داد تا تفاوتهای فردی فاحش خود را تا اندازه‌ای کاهش دهند و به سوی تعادل پیش بروند^{۱۳}. برای نیل به این هدف کافی است کلاس خاصی برای تصحیح کنندگان اوراق امتحانی تشکیل داد و طی آن، آنها را با شیوه‌های جدید طرح سوالات، نمره‌گذاری، تبدیل نمرات مختلف به یک مقیاس معین و... آشنا کرد.

همانطور که در بالا اشاره کردیم، تفاوت در شخصیت ارزشیابها واقعیتی است که با گذشته آنها پیوند می‌خورد و امکان تعدیل این اختلافات تا اندازه‌ای وجود دارد. حال ببینیم عوامل چگونه دست به دست می‌دهند تا دو ارزشیاب، یک ورقه معین با اوراق معینی را به دو صورت متفاوت ارزشیابی کنند.

۶ - انتظارات ارزشیاب: اثرات همسان‌سازی^{۱۴} و تلاش برای ایجاد هماهنگی شناختی^{۱۵} بعید است تصحیح کننده‌ای بخواهد اوراقی را تصحیح کند اما در مورد صاحبان آنها اطلاعاتی نداشته باشد. تصحیح کننده نه تنها در مورد صاحبان اوراق اطلاع دارد بلکه می‌داند که آنها در چه شرایطی کار کرده‌اند. اطلاعات ارزشیاب در مورد صاحبان اوراق می‌تواند به دو صورت باشد: عمومی و خصوصی. اطلاعات عمومی آنهاست که کل صاحبان اوراق و کل شرایط امتحان را در بر می‌گیرد. مثلاً ارزشیاب می‌داند که اوراق مورد ارزشیابی به دانش‌آموزان کلاس پنجم ابتدایی، سوم راهنمایی، چهارم دبیرستان، سال اول دانشگاه یا کسان دیگری تعلق دارد. همچنین می‌داند که اوراق در یک ربع ساعت، نیم ساعت، دو ساعت یا بیشتر نوشته شده است. اطلاعات خصوصی تک تک

داوطلبان را در بر می‌گیرد. مثلاً ارزشیاب می‌داند ورقه‌ای که در زیر دست دارد از آن دانش‌آموزی است که در ثلث اول شاگرد اول شده است یا دانش‌آموزی است که با خود او یا با مدیر مدرسه یا با یکی از همکاران او نسبتی دارد. می‌توان اطلاعات خصوصی را به آسانی از بین برد و در نتیجه آثار آنها را خنثی کرد. بدین ترتیب که اوراق را بدون اسم در اختیار ارزشیابها قرار داد. اما از بین بردن اطلاعات عمومی نه امکان‌پذیر است و احتمالاً نه به صلاح.

بنابراین هر ارزشیاب، قبل از شروع به ارزشیابی، اطلاعاتی در مورد صاحبان اوراق دارد که ما آنها را اطلاعات اولیه می‌نامیم. سؤال این است که آیا اطلاعات اولیه می‌تواند ارزشیابها را تحت تأثیر قرار دهد، ارزشیابی را از مسیر اصلی خارج سازد و نهایتاً موجب اختلاف در امر ارزشیابی شود؟ اگر پاسخ مثبت است، این انحراف چگونه صورت می‌گیرد و در چه جهتی سیر می‌کند؟ برای پاسخگویی به این سؤالات، نتایج آزمایشهایی را ارائه می‌دهیم که در این خصوص انجام گرفته است.

الف - اثر اطلاعاتی که به موقعیت و عملکردهای قبلی دانش‌آموزان مربوط می‌شود. آیا ارزشیابها در مورد دانش‌آموزانی که می‌شناسند و دانش‌آموزانی که نمی‌شناسند به صورتهای متفاوت قضاوت می‌کنند؟ آزمایشهایی که در این خصوص انجام گرفته نشان می‌دهد که اطلاعات اولیه در مورد صاحبان اوراق، ارزشیابی تکالیف آنها را تحت تأثیر قرار می‌دهد.

یکی از این آزمایشها، که بر روی ارزشیابی ترجمه‌های یک متن انگلیسی، در سطح دیپلم انجام گرفته، دارای دو مرحله بود. در مرحله اول از ارزشیابها خواسته‌اند که هر یک ۱۲ ورقه را تصحیح کند. اوراق را طوری انتخاب کرده بودند که نمرات مختلفی را کسب کنند. بدین معنا که سطح آنها بسیار متفاوت انتخاب شده بود. در مرحله دوم، که دو ماه بعد انجام گرفته و ارزشیابها قبلاً اطلاعی از آنها نداشتند، از ارزشیابها خواسته‌اند که ۶ ورقه دیگر را ارزشیابی کنند. ۶ ورقه اخیر از متن دیگری انتخاب شده بودند و تقریباً سطح یکسانی داشتند. روی هر یک از اوراق نمره‌ای وجود داشت و چنین وانمود شده بود که این نمره توسط صاحب ورقه در امتحان قبلی گرفته شده است. البته نمراتی که در روی این شش ورقه نوشته شده بود از میان نمراتی انتخاب گردیده بود که ارزشیابها همان نمرات را در مرحله اول، یعنی در تصحیح ۱۲ ورقه اول، داده بودند. محققان در این کار دو سری نمره انتخاب کرده، آنها را روی اوراق نوشته بودند: نمرات پایین و نمرات بالا. جدول ۶ نشان می‌دهد که هر وقت به ارزشیابها گفته‌اند که صاحبان اوراق در امتحان قبلی نمره بیشتری اخذ کرده‌اند، آنها نیز نمره بیشتری به اوراق داده‌اند و برعکس.

جدول شماره ۶ - میانگینهای نمرات بر حسب نمرات ساختگی قبلی

نمرات قبلی	
زیاد	کم
۱۲	۱۰

با توجه به جدول شماره ۶ معلوم می‌شود که قضاوت ارزشیابها در مورد اوراق، بر حسب اطلاع از موقعیت صاحبان آنها، انجام گرفته است. وقتی به ارزشیابها گفته‌اند که نمره قبلی صاحبان اوراق کم بوده است (مثلاً بین ۷ الی ۱۰) با میانگین ۱۰ و زمانی که گفته‌اند نمره قبلی آنها زیاد بوده است (مثلاً بین ۱۵ و ۱۸) با میانگین ۱۲ ارزشیابی کرده‌اند.

معنای تفاوت میانگینها این است که ارزشیابیهای بعدی تحت تأثیر ارزشیابیهای قبلی قرار می‌گیرد. به عبارت دیگر، ارزشیابیهای بعدی به ارزشیابیهای قبلی وابستگی پیدا می‌کنند و این وابستگی از طریق همسان سازی ارزشیابیهای بعدی با ارزشیابیهای قبلی نمایان می‌شود. به اعتقاد تعداد زیادی از مریبان، یکی از عللی که موجب می‌شود تا شاگرد اول‌ها در اغلب اوقات شاگرد اول شوند این است که معلمان آنها را به عنوان شاگردان ساعی می‌شناسند. این شناخت موجب می‌شود که ارزشیابها اوراق دانش‌آموزان را با دید مثبت تصحیح می‌کنند و بدین وسیله قضاوت‌های بعدی را با قضاوت‌های قبلی خود همسان می‌سازند. البته علت دیگری هم وجود دارد که شاگردان ممتاز در اکثر موارد به همان صورت ممتاز جلوه می‌کنند و آن عبارت است از ثبات نسبی استعدادهای که اینجا جای تشریح آن نیست.

حال به تشریح نتایج آزمایشی می‌پردازیم که خود ما در دانشکده روان‌شناسی دانشگاه علامه طباطبائی انجام داده‌ایم. از استاد زبان انگلیسی دانشکده خواهش کردیم که اوراق زبان انگلیسی یکی از کلاسهای خود را در اختیار ما بگذارد. از میان کل اوراق، ده ورقه به طور تصادفی انتخاب شد. اوراق انتخاب شده در ۱۰ نسخه تکثیر و ترجمه صحیح متن انگلیسی نیز ضمیمه آنها گردید. اوراق اسم نداشت و از ۱ تا ۱۰ شماره گذاری شده بود. پنج سری آنها در اختیار پنج دبیر انگلیسی گذاشته شد تا به طور مستقل تصحیح کنند. به آنها گفته شد که اوراق به دانش‌آموزان سال چهارم تجربی تعلق دارد. پنج سری دیگر در اختیار پنج تصحیح‌کننده دیگر قرار گرفت تا در همان شرایط نمره گذاری کنند. به گروه اخیر گفته شد که اوراق به دانش‌آموزان سال دوم رشته زبان انگلیسی دانشگاه تعلق دارد. نتایج به دست آمده نشان داد که گروه اول (اوراقی که به دانش‌آموزان سال چهارم تجربی نسبت داده شده بود) نمرات بیشتری دریافت کرده است. به سخن دیگر، در نمرات دو گروه ارزشیاب، تفاوت‌های چشمگیری وجود داشت.

تفسیر مطلب به این صورت است که اطلاع از موقعیت صاحبان اوراق در ارزشیابها انتظار به وجود می‌آورد و این انتظار ارزشیابی آنها را تحت تأثیر قرار می‌دهد. در این مورد خاص، بدیهی است که انتظار تصحیح‌کنندگان از دانشجویان رشته زبان بمراتب بیشتر از دانش‌آموزان سال چهارم تجربی است. بنابراین کار دانش‌آموزان چهارم تجربی در نظر ارزشیابها بهتر و چشمگیرتر جلوه می‌کند و در نتیجه نمرات بهتری را به دست می‌آورد.

همان اوراق با همان شرایط دوباره در ده نسخه تکثیر شد. پنج سری در اختیار یک گروه از دبیران زبان دبیرستانها و پنج سری دیگر در اختیار گروه دیگری از دبیران زبان، که همه به صورت تصادفی انتخاب شده بودند، گذاشته شد. به هر دو گروه گفته شد که این اوراق به دانش‌آموزان سال چهارم ریاضی تعلق دارد و در امتحان نلت آخر نوشته شده است. به گروه اول تصحیح‌کنندگان گفته شد که نمرات صاحبان اوراق در دو نلت اول و دوم بین ۷ و ۱۰ بوده است. به گروه دوم از تصحیح

کنندگان گفته شد که نمرات صاحبان این اوراق در دو ثلث اول و دوم بین ۱۷ و ۲۰ بوده است. ارزشیابها به طور مستقل و جدا از یکدیگر کار کردند. نتایج به دست آمده در جدول شماره ۷ نشان داده شده است:

جدول شماره ۷ - میانگینهای نمرات بر حسب نمرات ساختگی قبلی

نمرات قبلی	
بین ۱۷ و ۲۰	بین ۷ و ۱۰
۱۴	۱۱

همانطور که جدول ۷ نشان می‌دهد، زمانی که اوراق به دانش‌آموزان قوی نسبت داده شده با میانگین ۱۴ و زمانی که به دانش‌آموزان ضعیف نسبت داده شده با میانگین ۱۱ نمره گذاری شده است. معنای میانگینهای ۱۱ و ۱۴ این است که اطلاع از موقعیت و عملکردهای قبلی داوطلبان موجب می‌شود که ارزشیابها در ارزشیابی خود تفاوت‌های فردی نشان دهند.

تفسیری که از تفاوت دو میانگین ۱۱ و ۱۴ ارائه می‌شود این است که وقتی به تصحیح کنندگان گفته می‌شود که نمرات دو ثلث اول و دوم صاحبان اوراق بین ۱۷ و ۲۰ یا بین ۷ و ۱۰ بوده است انتظاری در آنها به وجود می‌آید. یعنی تصحیح کنندگان پیش خود می‌گویند که باید نمرات آخر سال هم تقریباً نزدیک به نمرات دو ثلث قبلی باشند. فرض کنید یکی از ارزشیابها، که اوراق به اصطلاح دانش‌آموزان قوی را تصحیح می‌کند، احساس می‌کند که یکی از اوراق مثلاً ارزش نمره ۱۸ را ندارد بلکه در نظر او نهایتاً سزاوار نمره ۱۰ است. در اینجا نوعی ناهماهنگی شناختی به ارزشیاب دست می‌دهد که بنوبه خود ناراحت کننده است و باید به نحوی حل شود. ارزشیاب با کاهش دادن نمره‌ای که قبلاً انتظار آن را داشت و افزایش دادن نمره‌ای که ورقه حایز آن است، ناهماهنگی شناختی خود را از بین می‌برد. به عبارت دیگر ارزشیاب فاصله دو نمره را کاهش می‌دهد و به صاحب ورقه مثلاً نمره ۱۴ می‌دهد که به اصطلاح نه سیخ بسوزد و نه کباب.

برای تأیید این که اطلاع از موقعیت و عملکردهای قبلی دانش‌آموزان می‌تواند قضاوت‌های بعدی معلمان را تحت تأثیر قرار دهد، کافی است به نتایج امتحانات نهایی چند مدرسه مراجعه کنیم خواهیم دید که در خیلی از موارد، کسانی که در مدرسه شاگرد اول بوده‌اند در نهایی شاگرد اول شده‌اند، چون ارزشیابها آنها را نمی‌شناختند.

عوامل دیگری نیز وجود دارد که می‌تواند ارزشیابها را تحت تأثیر قرار دهد و موجب تفاوت قضاوت‌های آنها درباره صاحبان اوراق گردد، مثلاً اطلاع از طبقه اجتماعی - اقتصادی دانش‌آموزان یکی از آنهاست. اگر شما اوراق معینی را یک بار به دانش‌آموزان شمال شهر تهران و بار دیگر به دانش‌آموزان جنوب آن نسبت دهید و از دو گروه معلم بخواهید تا به آنها نمره بدهند خواهید دید که به دو صورت متفاوت نمره می‌دهند.

اجملاً این که ارزشیابها وقتی در موقعیتهای متفاوت قرار می‌گیرند به صورتهای مختلف

قضایوت می‌کنند. بنابراین می‌توان گفت که تفاوت موقعیتها می‌تواند یکی از علل وجود اختلاف ارزشیایها در امر ارزشیایی باشد.



زیرنویسها:

1 – Evaluation

2 – Mesvrement

3 – Critere

4 – Evaluation scolaire

5 – piéron, H. Examens et Docimologie. P.U.F. 1969.

۶ – مقیاس نمره‌گذاری بین صفر و ۲۰ بوده است.

۷ – زرژ نوازه وزان بل کاورنی، روان‌شناسی ارزشیایی تحصیلی، ترجمه حمزه گنجی، تهران انتشارات اطلاعات، ۱۳۶۴، صفحه ۲۹.

8 – Limites de Confiance.

۹ – مأخذ شماره ۷، صفحه ۳۱.

10 – Multicorrection

11 – Notes brutes.

12 – Fonction.

۱۳ – روان‌شناسی ارزشیایی تحصیلی، اثر زرژ نوازه وزان بل کاورنی، ترجمه حمزه گنجی، تهران، انتشارات اطلاعات، ۱۳۶۴، فصل دوم.

14 – Assimilation

15 – Consonance Cognitive.

۱۶ – در اینجا لازم می‌دانیم از همکاری صمیمانه خانم یاسمن مندی، استاد زبان انگلیسی دانشکده، تشکر کنیم.

منابع:

1 – BONBOIR, A., La Docimologie, Paris, P.U.F., 1972.

2 – NONNIOL. J.-J., Les divergences de notation tenant aux effets d'ordre de Correction, Cahiers de Psychologie, 1965.

3 – CAVERNI, J. –P., Influence exercée sur la notation Par Le type d'échelle utilisée Par le notatair, Aix-enprovence, université de Provence, Laboratoire de P.E. 1975.

4 – FABRE.J. – M., et NOIZET, G., Confiance attachée aux réponses à des questions à Choix multiple, Journal de Psychologie normale et Pathologie, 1977 a.

5 – NOIZET, G. et CAVERNI, J., P. Psychologie de l'évaluation scolaire, P.V.F., 1978.

6 – PIÉRON, H., Examens et Docimologie, Paris, P.U.F., 1969.

7 – POITOU, J., – P., La dissonance cognitive, Paris, Armand Colin, 1974.

8 – REUCHLIN, M., Expérimentation hors du Laboratoire, Psychologie française, 1976.