

نظریهٔ تعمیم‌پذیری در اندازه‌گیری آموزشی

به منظور تعیین تعداد (۱) سؤالهای لازم در آزمون علوم برای اندازه‌گیری توان حقیقی یا نمره حقیقی فرد در حد قابل قبول (۲) تعداد تصحیح‌کننده لازم برای رسیدن به پایایی مناسب بین نظر تصحیح‌کننده‌ها، پاس‌نامه‌های ۱۲۰ نفر دانش‌آموز کلاس پنجم ابتدایی در آزمون علوم از ۱۵ دبستان (هر دبستان ۸ نفر) به صورت تصادفی جمع‌آوری^۱ و با توجه به دستورالعمل تعیین شده برای نمره‌گذاری، توسط ۳ نفر به صورت مستقل تصحیح و نمره داده شد. در این آزمون برای اندازه‌گیری توانایی سطوح بالای یادگیری (فهمیدن کاربرد و تجزیه و تحلیل) برای هر فرد شرایط آزمایشگاهی فراهم می‌آید و از وی خواسته می‌شود که آزمایشهای مربوط به پنج فعالیت مختلف (سؤال) را انجام داده، چگونگی انجام دادن آزمایش همراه با نتیجهٔ آزمایش را با ذکر دلیل به صورت تشریحی در پاس‌نامه خود بیان کند. گزارش پژوهش در قالب معرفی نظریهٔ تعمیم‌پذیری ارائه شده است. بدین معنی که ابتدا ضمن مقایسه مختصر نظریهٔ کلاسیک و نظریه تعمیم‌پذیری، مفاهیم و اصطلاحات مورد استفاده در نظریه تعمیم‌پذیری مورد بحث قرار گرفته است، سپس از طریق تحلیل واریانس "Anova" مؤلفه‌های واریانس مربوط به هر یک از منابع تغییرپذیری در طرحهای اندازه‌گیری برآورد، و ضریب تعمیم‌پذیری نمره هر فرد در یک سؤال به مجموعه نمره‌های ممکن محاسبه شده است. در پایان با استفاده از مطالعات تصمیم‌گیری، پاسخ سؤالهای مطرح شده در پژوهش در قالب دو طرح اندازه‌گیری یک رویه‌ای و دو رویه‌ای فراهم آمده است.

۱. پاس‌نامه‌های دانش‌آموزان قبل از تصحیح در مدارس، جمع‌آوری و در پژوهش مورد استفاده قرار گرفته

مقدمه

استفاده از آزمونهای روانی و آموزشی و تأثیر نتایج حاصل از این آزمونها در تصمیم‌گیریهای مختلف، اطلاع دقیق از شاخصهای مختلف این آزمونها از شاخصهای پایایی^۱ و روایی^۲ را ضروری می‌کند. اگر عملکرد آزمون، توان یا نظر پاسخ دهندگان را در موضوع مورد نظر به صورت دقیق اندازه‌گیری نکند، یعنی: (۱) عملکرد فرد در آزمون با عملکرد حقیقی وی تفاوت داشته باشد یا (۲) عملکرد مشاهده شده معرّف عملکرد حقیقی باشد ولی آنچه اندازه گرفته شده منعکس کننده هدف مورد نظر در اندازه‌گیری نباشد، استفاده از نتایج آزمون بی‌حاصل و گاه مضّر خواهد بود.

خصوصیات مختلف یک آزمون با استفاده از تئوریها یا نظریه‌های روان‌سنجی^۳ مورد مطالعه قرار می‌گیرد. دو نظریه عمده در روان‌سنجی عبارتند از: نظریه نمونه‌گیری تصادفی^۴ و نظریه سؤال-پاسخ^۵. در نظریه نمونه‌گیری تصادفی «به رابطه میان عملکرد مشاهده شده و عملکرد حقیقی همانند تعمیم از یک نمونه به یک جامعه توجه می‌شود. درحالی‌که نظریه سؤال-پاسخ بر آن است که اگر چگونگی ارتباط هر سؤال آزمون با توانایی فرد مورد مطالعه درک شود، می‌توان نمره حقیقی فرد را به صورت مستقیم برآورد کرد» (۱). نظریه نمونه‌گیری تصادفی خود شامل دو نظریه نسبتاً متفاوت است: نظریه کلاسیک^۶ یا پایایی کلاسیک و نظریه تعمیم‌پذیری^۷.

1 . Reliability

2 . Validity

3 . Psychometric

4 . Random sampling theory

5 . Item- response theory

6 . Classical theory

نظریه کلاسیک

نظریه کلاسیک نمره مشاهده شده از هر فرد در هر آزمون را ترکیبی از دو نمره حقیقی^۱ فرد و خطای تصادفی^۲ فرض می‌کند. تفکیک و برآورد سهم هر یک از این دو جزء به صورت مستقیم امکان‌پذیر نیست، ولی با فرض تکرار عمل اندازه‌گیری در دفعات متعدد و مشاهده یا جمع‌آوری نمره‌های متعدد از یک فرد، مقدار مورد انتظار^۳ نمره مشاهده شده (معدل نمره‌های مشاهده شده در دفعات متعدد) برابر نمره حقیقی فرد و مقدار مورد انتظار خطا صفر خواهد شد. در این نظریه، پایایی عبارت از میزان ارتباط بین نمره مشاهده شده و نمره حقیقی یا نسبت واریانس نمره واقعی به واریانس نمره مشاهده شده است. این نسبت بیانگر «همبستگی میان نمره‌های حقیقی و تمام نمره‌های مشاهده شده ممکن از تکرار دفعات متعدد آزمون است» (۲).

مفروضه نظری بی‌نهایت تکرار آزمون، در عمل به مفروضه دو آزمون هم‌ارزیاهمتا^۴ تبدیل می‌شود. اگر بتوان مفروضه هم‌ارز بودن، یعنی مساوی بودن نمره حقیقی هر فرد در دو آزمون و مساوی بودن واریانس خطای دو آزمون را برقرار کرد، از طریق اجرای دو آزمون هم‌ارز در دو زمان متفاوت، می‌توان ضریب همبستگی بین نتایج دو آزمون را محاسبه نمود. درچنین شرایطی «ضریب همبستگی بین نمره‌های مشاهده شده در دو آزمون هم‌ارز به مجذور همبستگی بین نمره مشاهده شده و نمره حقیقی تبدیل می‌شود» (۳).

ضریب محاسبه شده در نظریه کلاسیک ضمن آنکه بر مفروضه هم‌ارز بودن آزمونها متکی است، در هر بار اجرای آزمون تمام منابع مؤثر در خطای اندازه‌گیری را با عنوان خطای تصادفی محاسبه و بررسی می‌کند. مثلاً در تکرار یک آزمون در دو زمان متفاوت، فقط واریانس حاصل از تفاوت در زمان اجرای آزمون را به عنوان خطای اندازه‌گیری بررسی می‌کند و واریانس به سؤالیهای آزمون و اثر متقابل میان سؤال و اجرای آزمون (تفاوت در پاسخگویی به یک سؤال در دو زمان) را نادیده می‌گیرد، درحالی‌که

1. True score
2. Random error
3. Expected value

ضریب پایایی محاسبه شده از طریق اجرای دو آزمون هم ارز در دو زمان بافاصله نسبتاً کوتاه، فقط هم ارز بودن سؤالهای دو آزمون را به عنوان خطای اندازه گیری مورد توجه قرار می دهد و واریانس های دیگر را نادیده می گیرد. ضریب پایایی محاسبه شده سؤالهای آزمون از طریق روشهای تقسیم فقط دو یا چند جزء (هر سؤال یک جزء) را منبع خطای اندازه گیری می داند. بدین ترتیب در نظریه کلاسیک، نمره های خطا برای محاسبه ضریب پایایی با توجه به موقعیت خاص یا شیوه خاص به کار گرفته شده تغییر کرده، در نتیجه مقدار ضریب پایایی نیز تغییر خواهد کرد. تفاوت در مقدار ضریب پایایی و همبستگی منفی میان ضریب پایایی و خطای معیار اندازه گیری، به محاسبه مقادیر مختلف خطای معیار اندازه گیری و در نتیجه تشکیل فاصله های متفاوت اطمینان برای نمره هر فرد آزمودنی منجر خواهد شد. از این رو، می توان گفت که نظریه کلاسیک بانادیده گرفتن موقعیت یا شرایط اجرای آزمون، خطای اندازه گیری را «از طریق یک مفهوم مبهم که به آن خطای تصادفی گفته می شود» (۴) تعریف می کند. با توجه به محدودیت های مورد بحث می توان به این نتیجه گیری اشاره کرد که «نظریه کلاسیک با وجود تأثیر نسبتاً طولانی بر حوزه اندازه گیری های آموزشی و روانی و کاربرد عملی آن، فاقد انعطاف پذیری لازم برای حل همه مسائل مربوط به پایایی در حوزه اندازه گیری است» (۵)

مفاهیم زیر بنایی نظریه تعمیم پذیری

رویه ها و مؤلفه های واریانس

نظریه کلاسیک بدون توجه به زمینه اجرای آزمون (زمان و شرایط اجرای آزمون، ثبات نظر تصحیح کنندگان یا ثبات نظر یک تصحیح کننده در دفعات مختلف تصحیح)، نمره حقیقی را خصیصه ای ثابت فرض کرده برای هر فرد فقط یک نمره حقیقی محاسبه می کند. در نظریه تعمیم پذیری تأثیر عوامل مختلف بر اجرای آزمون و نمره مشاهده شده مورد توجه قرار می گیرد. در نتیجه برای هر فرد بر حسب شرایط در نظر گرفته شده در آزمون، نمره های متفاوتی محاسبه می شود. در این نظریه هر نمره مشاهده شده معرف

یک نمونه از تمام نمره‌های ممکن یا مجموعه‌ی مرجع نمره^۱ فرض می‌شود که به وسیله‌ی یک یا چند واریانس بیان می‌شود. برای بیان ساده‌تر و درک سریع‌تر، مطالب مربوط به مفاهیم نظریه‌تعمیم‌پذیری و محاسبات آماری در قالب چند مثال ارائه می‌شوند.

زمانی که تهیه‌کننده‌ی آزمون تصمیم می‌گیرد با چند سؤال (تشریحی یا عینی)، پیشرفت درسی دانش‌آموزان را زمینه‌ای خاص اندازه‌گیری کند، یکی از رویه‌ها یا عوامل مورد علاقه خود را در اندازه‌گیری مشخص کرده است. رویه عبارت است از مجموعه‌ای از حالتها^۲ یا سطوح^۳ اندازه‌گیری که قصد داریم عملکرد آزمودنی‌ها را در این سطوح مورد سنجش قرار دهیم. بنابراین، برای تهیه‌کننده‌ی موردبحث سؤالاتی از آزمون رویه اندازه‌گیری و هر سؤال نیز یک سطح مورد قبول از رویه‌ی اندازه‌گیری به حساب می‌آید. از این رویه تمام سطوح قابل قبول برای تهیه‌کننده‌ی آزمون، مجموعه مرجع سطوح قابل قبول^۴ گفته می‌شود. بدین ترتیب نمره‌ی مشاهده شده سؤالاتی از آزمون فرد، یک نمونه تصادفی از تمام سؤالاتی ممکن یا مجموعه مرجع سؤالاتی قابل قبول فرض می‌شود و نظریه‌تعمیم‌پذیری، میزان اتکاء‌پذیری^۵ یا تعمیم‌پذیری نمره حاصل از یک سؤال نمونه را مجموعه مرجع سؤالاتی قابل قبول مشخص می‌کند.

اگر تهیه‌کننده‌ی آزمون، علاوه بر سؤالاتی از دو یا چند نفر تصحیح‌کننده برای نمره‌گذاری آزمون استفاده کند، تصحیح‌کننده یک رویه و هر فرد تصحیح‌کننده یک سطح از رویه موردبحث را تشکیل می‌دهد. نمره داده شده توسط این افراد می‌تواند یک نمونه تصادفی از تمام نمره‌هایی باشد که توسط مجموعه مرجع تصحیح‌کنندگان قابل قبول به دست می‌آید. تغییرپذیری در نمره تصحیح‌کنندگان، بیانگر خطای اندازه‌گیری است که از عدم پیروی از معیار یکسان در نمره‌گذاری حاصل می‌گردد. در نظریه کلاسیک، این منبع خطاپذیری را فقط بانادیده گرفتن منبع تغییرپذیری سؤالاتی از آزمون می‌توان محاسبه کرد. نظریه تعمیم‌پذیری با تکیه بر میزان تغییرپذیری نمونه‌گیری مربوط

1 . Universw - score

2 . Conditions

3 . Levels

4 . Universe of admissible conditions

به سؤاها، تصحیح کننده‌ها (یا هر چند رویه دیگر) و ترکیب‌های مختلف آنها با یکدیگر، میزان خطای اندازه‌گیری را در قالب مؤلفه‌های واریانس برآورد می‌کند. به عبارت دیگر، تمام منابع بالقوه خطا (رویه‌ها) و ترکیب‌های مختلف (اثر متقابل) را در قالب یک مطالعه بررسی کرده و مقدار واریانس هر مؤلفه را به صورت جداگانه برآورد می‌کند. در چنین حالتی مجموعه مرجع نمره‌های قابل قبول^۱ برای تهیه کننده آزمون تمام سطوح رویه‌های مورد استفاده در آزمون را در بر می‌گیرد، و نمره مشاهده شده هر فرد به تمام سطوح ممکن یا قابل قبول تعمیم داده می‌شود.

باتوجه به دو نمونه مورد بحث، نمره مشاهده شده در یک آزمون از رویه‌های مختلف مورد استفاده در آزمون نظیر سؤال، تعداد دفعات اجرای آزمون، تعداد تصحیح کننده و... تأثیر می‌پذیرد و نمره مشاهده شده باتوجه به رویه‌های مورد استفاده در آزمون معرف عملکرد فرد در همان رویه‌هاست. لذا اگر فقط یک رویه در آزمون به کار گرفته شود. طرح آزمون را می‌توان طرح یک رویه‌ای^۲ نامید. در این طرح افراد آزمودنی هدف اندازه‌گیری بوده و تشخیص تفاوت‌های فردی میان آنها تفاوت حقیقی یا واریانس نمره حقیقی بین آنها را به دست می‌دهد. تغییر پذیری نمونه‌گیری از سؤاها و اثر متقابل بین سؤال و فرد، هر دو عامل بالقوه خطای اندازه‌گیری بوده که از مجموع واریانس‌های این دو عامل واریانس نمره حقیقی، واریانس نمره مشاهده شده محاسبه می‌شود. در این طرح نمره کسب شده به مجموعه مرجع نمره‌های مورد قبول در رویه سؤال (نمره ممکن در تمام سؤالهای ممکن) تعمیم داده می‌شود. واریانس‌های مختلف حاصل از طرحی یک رویه‌ای که تمام افراد آزمودنی به تمام سؤالهای آزمون پاسخ می‌دهند در نمودار شماره الف ارائه شده است. در نمودار ۱ / الف، p نشان دهنده فرد آزمودنی یا هدف اندازه‌گیری، i نشان دهنده سؤال و e نشان دهنده خطاهای غیر سیستماتیک در اندازه‌گیری است. نمره مشاهده شده هر فرد (x_{pi}) در هر سؤال ترکیبی است از:

$x_{pi} = u$	میانگین کل	
$+ (u_p = u)$	اثر فرد	σ^2_p یا نمره واقعی
$+ (u_i - u)$	اثر سؤال	σ^2_i واریانس خطا
$+ (x_{pi} - u_p - u_i + u)$	باقیمانده یا خطا	$\sigma^2_{pi.e}$

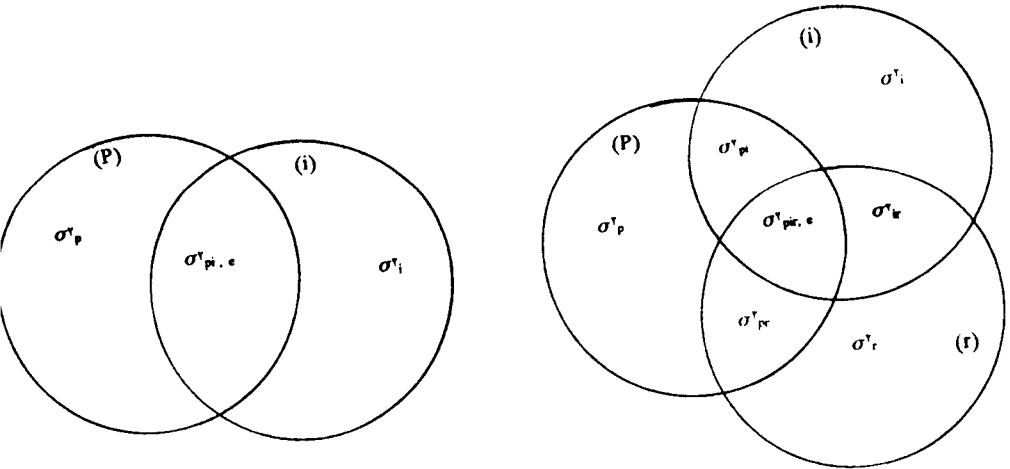
به استثنای میانگین، سایر اجزاء رابطه بالا - برای تمام افراد آزمودنی - هر کدام توزیعی دارند با میانگین صفر و واریانس σ^2 . نظریه تعمیم‌پذیری با استفاده از تحلیل واریانس "ANOVA"، مقدار واریانس هر یک از این توزیع‌ها را مشخص می‌سازد. واریانس نمره مشاهده شده از ترکیب واریانس‌های محاسبه شده برآورد می‌شود:

$$\sigma^2(X_{pi}) = \sigma_p^2 + \sigma_i^2 + \sigma_{pi,e}^2$$

اگر دو رویه به صورت همزمان در آزمون به کار گرفته شود، طرح آزمون را می‌توان طرحی دو رویه‌ای^۱ دانست. در این طرح نیز افراد آزمودنی هدف اندازه‌گیری بوده و تفاوت فردی بین آنها واریانس نمره حقیقی محسوب می‌شود. تغییرپذیری نمونه‌گیری در رویه‌های طرح و اثرهای متقابل به وجود آمده همگی عامل‌های بالقوه خطای اندازه‌گیری به شمار می‌روند. درچنین طرحی، نمره کسب شده توسط هر فرد به مجموعه مرجع نمره‌های قابل قبول در همان دو رویه تعمیم داده می‌شود. در نمودار a/b واریانس‌های مختلف حاصل از طرح دو رویه‌ای که تمام سطح‌های دو رویه و هدف آزمون (افراد) با هم متقاطع^۲ هستند، ارائه شده است.

در نمودار a/b ، P آزمودنی، i سؤال (یا یکی از رویه‌ها)، r تصحیح‌کننده (یا رویه دیگر) و سایر علامتها اثرهای متقابل محاسبه شده را نشان می‌دهند. در این طرح نمره مشاهده شده هر فرد در هر یک از خانه‌های جدول (برخوردیک سطح از یک رویه با یک سطح از رویه دیگر برای هر فرد) یا X_{pir} برابر است با:

$X_{pir} = u$	میانگین		
$+(u_p - u)$	اثر فرد	واریانس واقعی	σ_p^2
$+(u_i - u)$	اثر سؤال	واریانس سؤال	σ_i^2
$+(u_r - u)$	اثر تصحیح‌کننده	واریانس تصحیح‌کننده	σ_r^2
$+(u_{pr} - u_p - u_r + u)$	اثر متقابل p,r		σ_{pr}^2 واریانس
$+(u_{pi} - u_p - u_i + u)$	اثر متقابل i,p		σ_{pi}^2 خطا
$+(u_{ir} - u_i - u_r + u)$	اثر متقابل r,i		σ_{ir}^2
$+(X_{pir} - u_p - u_i - u_r + u)$	باقیمانده یا خطا		$\sigma_{pir,e}^2$



نمودار شماره ۱. طرح یک رویه‌ای الف، طرح دو رویه‌ای ب

نظریه تعمیم‌پذیری با استفاده از تحلیل واریانس مقدار هر یک از مؤلفه‌های واریانس را محاسبه می‌کند. واریانس نمره مشاهده شده برابر است با مجموع مقدار تمام مؤلفه‌های واریانس:

$$\sigma^2(X_{pic}) = \sigma^2_p + \sigma^2_i + \sigma^2_r + \sigma^2_{pi} + \sigma^2_{pr} + \sigma^2_{ir} + \sigma^2_{pic.e}$$

طرح‌های متقاطع و طرح‌های آشیانه‌ای^۱
در طرح یک رویه‌ای ارائه شده در نمودار ۱ / الف هر فرد آزمودنی (p) به تمام

سؤالهای آزمون (i) پاسخ می‌دهد. درچنین حالتی ارتباط بین p و i متقاطع طرح با P_{xi} مشخص می‌شود. در نمودار ۱ / ب هر فرد آزمودنی به تمام سؤالها پاسخ داده و هر فرد تصحیح کننده، به هر سؤال هر فرد نمره می‌دهد. در این حالت نیز ارتباط بین r, i, p متقاطع و طرح با $pxixr$ مشخص می‌شود. در بعضی شرایط، مثلاً "ارزشیابی برنامه‌های آموزشی، گاه به دلیل حجم زیاد هدفهای آموزشی، می‌توان با نمونه‌گیری تصادفی چند گروه نمونه یا پاره آزمون از مجموعه مرجع سؤالها انتخاب و عملکرد هر گروه از آزمودنی را فقط در یک پاره آزمون مورد اندازه‌گیری قرار داد. درچنین حالتی ارتباط بین آزمودنی‌ها (p) و سؤالها (i) ارتباطی آشیانه‌ای است و با $i:p$ نشان داده می‌شود. در این مثال سؤالها در درون آزمودنی‌ها لانه دارند. در طرح یک رویه‌ای آشیانه‌ای^۱، برخلاف طرح یک رویه‌ای متقاطع که چهار مؤلفه واریانس را می‌توان برآورد نمود، فقط دو مؤلفه واریانس را می‌توان برآورد کرد (نمودار شماره ۲).

در این طرح سؤال مربوط به تغییرپذیری سؤالهای آزمون بدون پاسخ می‌ماند. نمره هر فرد آزمودنی X_{pi} برابر است با:

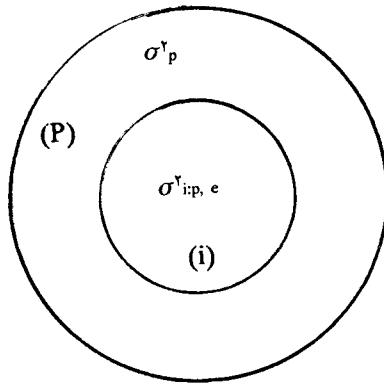
$$X_{pi} = u + (u_p - u) + (X_{pi} - u_p)$$

میانگین	
σ_p^2 واریانس واقعی	اثر وفرد
$\sigma_{i:pi,e}^2$ واریانس خطا	خطا

واریانس نمره مشاهده شده برابر است با مجموع دو مؤلفه برآورد شده:

$$\sigma^2(X_{pi}) = \sigma_p^2 + \sigma_{i:pi,e}^2$$

از مقایسه نمودارهای شماره ۱ / الف و ۲ نتیجه گرفته می‌شود که تفاوت عمده آماری بین این دو طرح در این است که اثر مستقیم مؤلفه واریانس مربوط به رویه آشیانه‌ای را جدا از مؤلفه واریانس اثر متقابل نمی‌توان بررسی کرد. به عبارت دیگر در طرح آشیانه‌ای اثر مستقیم i همراه با اثر متقابل pi و سایر خطاهای نامشخص برآورد می‌شود. بحث در مورد طرح‌های آشیانه‌ای دو رویه‌ای یا بیشتر به علت گستردگی، از حوصله این مقاله خارج است.



نمودار شماره ۲. طرح یک رویه‌ای آشیانه‌ای

رویه‌های ثابت^۱، تصادفی و مجموعه مرجع نمره

اگر سطوح رویه مورد بررسی در اندازه‌گیری به صورت تصادفی از میان مجموعه مرجع سطوح ممکن آن رویه انتخاب شده باشند، رویه را تصادفی می‌نامند. در چنین حالتی سطوح انتخاب شده برای مطالعه در هر بار اجرای آزمون متفاوت خواهند بود. مثلاً هر بار تعداد خاصی از سؤالهای آزمون یا هر بار تعداد خاصی از افراد تصحیح‌کننده از میان سطوح قابل قبول برای سؤال یا تصحیح‌کننده انتخاب و عملکرد آنها به مجموعه مرجع سطوح مورد قبول در همان رویه‌ها تعمیم داده می‌شود.

وقتی سطوح رویه یا رویه‌های مورد نظر محدود هستند، یا قصد تعمیم یافته‌ها به سطوح دیگر رویه‌های مورد مطالعه را نداریم، از رویه ثابت در اندازه‌گیری استفاده می‌شود. با استفاده از رویه ثابت "هدف ارزشیابی را می‌توان با پایایی و روایی بیشتری اندازه‌گیری کرد" (۶) با وجود این، "تعمیم‌پذیری یک نظریه اندازه‌گیری با رویه تصادفی

است و در هر طرح اندازه‌گیری حداقل یکی از رویه‌ها باید تصادفی باشد.» (۷) در بیشتر مطالعات تعمیم‌پذیری افراد یا آزمودنی‌ها هدف اندازه‌گیری^۱ محسوب می‌شوند. ولی گاه از عوامل دیگر نیز می‌توان به عنوان هدف اندازه‌گیری استفاده کرد. (۸) مثلاً در مطالعات ارزشیابی معمولاً^۲ کلاسها هدف اندازه‌گیری و دانش آموزان و سایر رویه‌ها (دانش آموزان در این وضعیت رویه محسوب می‌شوند)، مجموعه مرجع تعمیم^۳ را تشکیل می‌دهند. در هر صورت مجموعه مرجع نمره عبارت است از میانگین اعضای هدف اندازه‌گیری در تمام شرایط موجود در مجموعه مرجع تعمیم. واریانس حاصل از نمره‌های جامعه^۴ هدف اندازه‌گیری (آزمودنی یا کلاس) یعنی واریانس مجموعه مرجع نمره مشابه واریانس نمره حقیقی در نظریه کلاسیک است. (۹) قصد هر اندازه‌گیری فراهم آوردن شرایطی است که مقدار واریانس نمره حقیقی (واریانس مجموعه مرجع نمره) زیاد و مقادیر سایر مؤلفه‌های واریانس (واریانس‌های خطا) کم باشد. این هدف با کم و زیاد شدن رویه‌ها و سطح‌های هر یک از رویه‌های به کار گرفته شده در طرح اندازه‌گیری حاصل می‌گردد. بدین ترتیب، با توجه به رویه‌ها و سطوح به کار گرفته شده در اندازه‌گیری، میزان نمره حقیقی تعیین می‌شود و تغییر در تعداد هر یک از رویه‌ها یا سطوح رویه‌ها به تغییر در مجموعه مرجع نمره و مجموعه مرجع تعمیم منجر خواهد شد. به عبارت دیگر میزان واریانس مجموعه مرجع نمره در نظریه تعمیم‌پذیری با توجه به مجموعه مرجع تعمیم تغییر می‌کند. بر همین روال تعداد و میزان مؤلفه‌های واریانس خطا و سرانجام مقدار ضریب تعمیم‌پذیری^۴ برای سازه^۵ مورد نظر فرق خواهد کرد.

مطالعات تعمیم‌پذیری و مطالعات تصمیم‌گیری^۶

پس از برآورد مقادیر مؤلفه‌های واریانس در مطالعه تعمیم‌پذیری و مشاهده مقدار

1 . Object of measurement

2 . Universe of generalization

۳. در نظریه تعمیم از واژه universe مجموعه مرجع برای رویه و از واژه population (جامعه) برای هدف اندازه‌گیری استفاده می‌شود.

4 . Generalizability Coefficient

5 . Construct

واریانس‌های مربوط به مؤلفه می‌توان با تغییر در تعداد سطوح روبه‌رو یا روبه‌هایی که سهم بیشتری در خطای اندازه‌گیری دارند، میزان این خطاها را کاهش داده و ضریب تعمیم‌پذیری نتایج را افزایش داد. به مطالعات انجام شده جهت تعیین بهترین طرح اندازه‌گیری برای رسیدن به ضریب تعمیم‌پذیری مطلوب، مطالعات تصمیم‌گیری گفته می‌شود. مطالعات تصمیم‌گیری را می‌توان با فرمول اسپیرمن - براون^۱ در نظریه کلاسیک تا حدودی معادل دانست.

محاسبه مؤلفه‌های واریانس و ضرایب تعمیم‌پذیری

نظریه تعمیم‌پذیری را می‌توان در چارچوب تحلیل واریانس "ANOVA" بررسی نمود.^۲ در دنباله مقاله با استفاده از داده‌های جمع‌آوری شده از نمره ۱۲۰ دانش‌آموز، چگونگی محاسبه مقادیر مؤلفه‌های واریانس ضریب تعمیم‌پذیری و همچنین مطالعات تصمیم‌گیری (مطالعات D)، در قالب دو طرح اندازه‌گیری برای پاسخگویی به سؤالهای مطرح شده در پژوهش، مورد بحث قرار می‌گیرد.

طرح اندازه‌گیری یک روبه‌ای

به منظور پاسخگویی به سؤال پژوهش: «تعداد سؤالهای مورد نیاز برای رسیدن به ضریب تعمیم‌پذیری مناسب» نتایج حاصل از نمره دانش‌آموزان ابتدا در قالب طرح اندازه‌گیری یک روبه‌ای (آزمودنی P، سؤال i) مورد تجزیه و تحلیل قرار گرفته است. به عبارت دیگر در طرح PX_i فقط نمره‌های حاصل از تصحیح یک فرد نمره‌گذار مورد توجه و بررسی قرار گرفته است. در این طرح چون تمام افراد به تمام سؤالهای آزمون پاسخ داده‌اند، طرح اندازه‌گیری از نوع یک روبه‌ای متقاطع است. با استفاده از تحلیل واریانس یک عاملی، اطلاعات مربوط به مجموع مجذورات (SS)، درجات آزادی (df) و میانگین مجذورات (MS) برای هر یک از منابع تغییرپذیری محاسبه می‌شود. جدول شماره ۱ چارچوب لازم برای این اطلاعات را نشان می‌دهد. آخرین ستون جدول برخلاف جدول خلاصه اطلاعات "ANOVA" که به آزمون F اختصاص دارد، به مقدار نظری مجموع

علوم تربیتی

$$\hat{\sigma}_{\pi}^2 = \frac{(MS_{\pi} - \hat{\sigma}_{\pi i, e}^2)}{n_p} = \frac{(11/694 - 0/613)}{120} = 0/092$$

و با جا به جایی در فرمول $E(MSP)$ ، مقدار $\hat{\sigma}_{\pi}^2$ برابر است با:

$$\hat{\sigma}_{\pi}^2 = \frac{(MS_p - \hat{\sigma}_{\pi i, e}^2)}{n_i} = \frac{(2/28 - 0/613)}{5} = 0/333$$

چون مقدار واریانس از مقدار بارم سؤال تأثیر می‌پذیرد، برای تفسیر مؤلفه‌های واریانس این مقادیر به فراوانی نسبی یا درصدی تبدیل می‌شوند. با توجه به جدول شماره ۲ واریانس بین آزمودنی‌ها ۳۲ درصد کل واریانس نمره را به خود اختصاص داده است. واریانس بین آزمودنی‌ها - واریانس مطلوب - تفاوت بین دانش آموزان در درس علوم را نشان می‌دهد. واریانس برآورد شده برای سؤالها تقریباً $\frac{1}{4}$ واریانس برآورد شده برای آزمودنی‌هاست (۸/۸۶ درصد). کوچک بودن این درصد بیانگر آن است که سؤالهای آزمون از نظر سطح دشواری تفاوت زیادی با یکدیگر ندارند. آخرین منبع تغییر پذیری (Pi, e) نشان می‌دهد که عملکرد دانش آموزان در پاسخگویی به سؤالهای مختلف متفاوت است. به عبارت دیگر وضعیت پاسخگویی فرد از یک سؤال به سؤال دیگر فرق می‌کند. بعضی از دانش آموزان در پاسخگویی به یک یا چند سؤال و بعضی دیگر در پاسخگویی به چند سؤال دیگر موفق‌تر بوده‌اند.

جدول شماره ۲. داده‌های حاصل از ANOVA و مقدار مؤلفه‌های واریانس طرح اندازه‌گیری (Pxi)

درصد واریانس	مقدار مؤلفه واریانس	MS	df	SS	منبع تغییر پذیری
۳۲	۰/۳۳۳	۲/۲۸	۱۱۹	۲۷۱/۳۶	آزمودنی
۸/۸۶	۰/۰۹۲	۱۱/۶۹۴	۴	۴۶/۷۷	سؤال
۵۸/۹	۰/۶۱۳	۰/۶۱۳	۴۷۶	۲۹۱/۶۲	اثر متقابل و باقیمانده

مقادیر مؤلفه‌های واریانس را می‌توان به دو صورت تفسیر کرد: تفسیر هنجاری یا نسبی^۱ و تفسیر ملاکی یا مطلق^۲. تفسیرهای نسبی به رتبه بندی و مقایسه افراد در آزمون توجه دارند. درچنین حالتی، واژه خطا با توجه به مؤلفه‌هایی که بروضعیت نسبی فرد تأثیر می‌گذارد مشخص می‌شود. در طرح مورد بحث تنها واریانس $\sigma^2_{pi,e}$ یا قسمت قرار گرفته شده در داخل دایره مربوط به هدف اندازه‌گیری (نمودار ۱ / الف) به رتبه یا وضعیت نسبی فرد اثر می‌گذارد.

اگر سطح دشواری سؤالهای مجموعه مرجع سؤال با هم تفاوت داشته باشد، انتخاب سؤالهای ساده یا سؤالهای مشکل برای آزمون، در وضعیت مطلق فرد تأثیر می‌گذارد. لذا مقدار خطا در تصمیم‌گیریهای مطلق از ترکیب دو مؤلفه واریانس سؤال (σ^2_{pi}) و اثر متقابل ($\sigma^2_{pi,e}$) یا تمام قسمتهای نمودار ۱ / الف به استثنای بخش مربوط به σ^2_p به دست می‌آید. مقدار واریانس در تصمیم‌های نسبی برای یک فرد در یک سؤال^۳ آزمون برابر است با:

$$\sigma^2_{نسبی} = \frac{\sigma^2_{pi,e}}{n'i} = \frac{0/631}{1} = 0/613$$

و مقدار واریانس در تصمیم‌های مطلق برابر است با:

$$\sigma^2_{مطلق} = \frac{\sigma^2_{pi}}{n'i} + \frac{\sigma^2_{pi,e}}{n'i} = \frac{0/092}{1} + \frac{0/631}{1} = 0/705$$

مطالعه‌تعمیم‌پذیری (مطالعه G) ضمن آنکه به مؤلفه‌های واریانس و تفسیر این مؤلفه‌ها توجه دارد، همانند نظریه کلاسیک یک ضریب پایایی که به آن ضریب تعمیم‌پذیری گفته می‌شود فراهم می‌آورد. ضریب تعمیم‌پذیری عبارت است از "نسبت واریانس مجموع مرجع نمره به مقدار نظری واریانس نمره مشاهده شده" (۱۰). مقدار

1 . Relative

2 . Absolute

۳. در مطالعات D مقدار n' با توجه به طرح اندازه‌گیری و تعداد سطح رویه

ضریب تعمیم‌پذیری تقریباً برابر است با n مقدار نظری مجذور همبستگی بین نمره مشاهده شده مجموعه مرجع نمره که با علامت EP^2 نشان داده می‌شود. (۱۱).

مقدار نظری واریانس نمره مشاهده شده در آزمونهای ملاک نسبی برابر است با مجموع واریانس مجموع مرجع نمره (σ_p^2) و واریانس خطای اندازه‌گیری که بر وضعیت نسبی فرد اثر می‌گذارند. در طرح مورد بحث تنها σ_{pic}^2 بروضعیت نسبی فرد اثر می‌گذارد، در نتیجه ضریب تعمیم‌پذیری در طرح یک وجهی متقاطع برای تصمیم‌گیری نسبی برابر است با:

$$FP^2_{\text{نسبی}} = \frac{\sigma_p^2}{(\sigma_p^2 + \sigma^2_{\text{نسبی}})} = \frac{0/333}{(0/333 + 0/613)} = 0/352$$

و در تصمیم‌گیری مطلق برابر است با:

$$\hat{\phi} = \frac{\sigma_p^2}{(\sigma_p^2 + \sigma^2_{\text{مطلق}})} = \frac{0/333}{(0/333 + 0/705)} = 0/32$$

واریانس‌های برآورد شده در مطالعه G میزان خطا در تعمیم نمره حاصل از عملکرد فرد در یک سؤال مجموعه مرجع نمره و ضرایب محاسبه شده مقدار اتکاء پذیری تعمیم نمره مشاهده شده فرد در یک سؤال به نمره وی در مجموعه مرجع سؤالا را نشان می‌دهد.

کوچک بودن مقدار ضرایب تعمیم‌پذیری (حداکثر یک و حداقل صفر) عدم دقت در تعمیم نمره فرد در یک سؤال به نمره وی در مجموعه مرجع سؤالا را نشان می‌دهد. به عبارت دیگر نمره مشاهده شده فرد در یک سؤال را نمی‌توان نمره حقیقی وی قلمداد کرد. برای مشخص کردن تعداد سؤالهای لازم و رسیدن به ضریب تعمیم‌پذیری مطلوب (فرمول اسپیرمن - براون در نظریه کلاسیک همین نقش را ایفا می‌کند)، مقدار مؤلفه‌های واریانس حاصل از مطالعه G را در مطالعات D مورد استفاده قرار می‌دهیم. در مطالعات D با تغییر در تعداد سطوح رویه (در این پژوهش افزایش تعداد سؤالهای آزمون)، واریانس مربوط به تفاوت فردی یا واریانس حقیقی بدون تغییر باقی می‌ماند، ولی واریانس مربوط به تفاوت میان سؤالا و اثر متقابل کاهش می‌یابد. در نتیجه واریانس

سؤال داشته باشد ($n'i=5$) ضرایب تعمیم پذیری نسبی و مطلق به ترتیب $0/731$ و $0/704$ و اگر ۱۰ سؤال داشته باشد ($n'i=10$) این ضرایب به ترتیب $0/845$ و $0/826$ خواهند شد. اگر هزینه، وقت، و نیروی انسانی اجازه دهد با افزایش تعداد سؤالها می توان ضرایب مناسب تری به دست آورد. باتوجه به جدول شماره ۳ برای رسیدن به ضریب تعمیم پذیری بالاتر از $0/85$ ، حداقل به ۱۵ سؤال از نوع سؤالهای مورد استفاده در آزمون علوم نیاز داریم.

طرح اندازه گیری دو رویه ای

برای پاسخگویی به سؤالهای پژوهش: (۱) تعداد سؤال مورد نیاز (۲) تعداد تصحیح کننده مورد نیاز، نتایج حاصل از نمره دانش آموزان در قالب طرح اندازه گیری دو رویه ای مورد تجزیه و تحلیل قرار گرفت. در این طرح، چون تمام افراد آزمودنی به تمام سؤالهای آزمون پاسخ داده و هر فرد تصحیح کننده نیز تمام سؤالهای هر یک از افراد آزمودنی را تصحیح کرده است، رابطه بین هدف اندازه گیری (P) و رویه های طرح (r,i) یک رابطه متقاطع ($P \times i \times r$) است.

نمودار ۱/ب). در طرح دو رویه ای متقاطع ۷ مؤلفه واریانس به صورت همزمان برآورد می شود. ۱ سه مؤلفه برای اثرهای مستقیم (r,i,p)، سه مؤلفه برای اثرهای متقابل (ri,pr,pi) و یک مؤلفه برای اثر متقابل سه عامل همراه با واریانس خطاهای غیر سیستماتیک در اندازه گیری (pir, e).

تحلیل واریانس "ANOVA" اطلاعات مربوط به مجموع مجذورات. درجات آزادی و میانگین مجذورات هر یک از هفت منبع تغییر پذیری را فراهم می آورد (جدول شماره ۴). در این جدول نیز همانند جدول شماره ۱ مقدار مورد انتظار مجموع مجذورات، جای F را گرفته است.

در جدول شماره ۵، اطلاعات حاصل از تحلیل واریانس همراه با مقدار واریانس و فراوانی درصدی هر یک از مؤلفه های واریانس ارائه شده است. مقدار واریانس خطا $\hat{\sigma}_{pir,e}^2$ برابر است با $MS_{pir,e}$ یا:

$$\hat{\sigma}_{pir,e}^2 = MS_{pir,e} = 0/148$$

جدول شماره ۳. نتایج حاصل از مطالعه G و مطالعات D در طرح اندازه گیری (Pxi)

منبع تغییرپذیری	$\hat{\sigma}^2$	مطالعه G		مطالعات D		
		$n_i=1$	۵	۸	۱۰	۱۵
آزمودنی	$\hat{\sigma}_p^2$	۰/۳۳۳	۰/۳۳۳	۰/۳۳۳	۰/۳۳۳	۰/۳۳۳
سؤال	$\hat{\sigma}_{i}^2$	۰/۰۹۲	۰/۰۱۸	۰/۰۱۱	۰/۰۰۹	۰/۰۰۶
اثرمتقابل وباقیمانده	$\hat{\sigma}_{pi,e}^2$	۰/۶۱۳	۰/۱۲۳	۰/۰۷۶	۰/۰۶۱	۰/۰۴۱
نسبی	$\hat{\sigma}^2$	۰/۶۱۳	۰/۱۲۳	۰/۰۷۶	۰/۰۶۱	۰/۰۴۱
مطلق	$\hat{\sigma}^2$	۰/۷۰۵	۰/۱۴۱	۰/۰۸۷	۰/۰۷۰	۰/۰۴۷
\hat{P}^2		۰/۳۵۲	۰/۷۳۰	۰/۸۱۴	۰/۸۴۵	۰/۸۹۰
$\hat{\phi}$		۰/۳۲۰	۰/۷۰۲	۰/۷۹۲	۰/۸۲۶	۰/۸۷۶

جدول شماره ۴: اطلاعات ANOVA برای طرح اندازه گیری (Pxixr)

منبع تغییرپذیری	SS	df	MS	E(MS)
(P) آزمودنی			$\frac{SS_p}{df_p}$	$E(MS_p) = \sigma_{pi,e}^2 + n_i \sigma_{pr}^2 + n_r \sigma_{pi}^2 + n_i n_r \sigma_p^2$
(i) سؤال			$\frac{SS_i}{df_i}$	$E(ms_i) = \sigma_{pi,e}^2 + n_r \sigma_{pi}^2 + n_p \sigma_{ir}^2 + n_p n_r \sigma_i^2$
(r) تصحیح کننده			$\frac{SS_r}{df_r}$	$E(MS_r) = \sigma_{pi,e}^2 + n_i \sigma_{pr}^2 + n_p \sigma_{ir}^2 + n_p n_i \sigma_r^2$
Pi			$\frac{SS_{pi}}{df_{pi}}$	$E(MS_{pi}) = \sigma_{pi,e}^2 + n_r \sigma_{pi}^2$
pr			$\frac{SS_{pr}}{df_{pr}}$	$E(MS_{pi}) = \sigma_{pi,e}^2 + n_i \sigma_{pr}^2$
ir			$\frac{SS_{ir}}{df_{ir}}$	$E(MS_{pi}) = \sigma_{pi,e}^2 + n_p \sigma_{ir}^2$
Pir,e			$\frac{SS_{pir,e}}{df_{pir,e}}$	$E(MS_{pir,e}) = \sigma_{pir,e}^2$

پس از برآورد واریانس خطا (باقیمانده) واریانس‌های مربوط به هر یک از سه اثر متقابل با جا به جایی در فرمول‌های ارائه شده در جدول ۴ برآورد می‌شود. مثلاً "مقدار $\hat{\sigma}_{pi}^2$ برابر است با:

$$\hat{\sigma}_{pi}^2 = \frac{(MS_{pi} - \hat{\sigma}_{pir,e}^2)}{r} = \frac{(1/628 - 0/148)}{3} = 0/493$$

و پس از برآورد ۴ واریانس بالا واریانس‌های مربوط به هر یک از سه اثر مستقیم با جا به جایی در فرمول‌های ارائه شده در جدول شماره ۴ برآورد می‌شود. مثلاً مقدار $\hat{\sigma}_{i}^2$ برابر است با:

$$\hat{\sigma}_{i}^2 = \frac{(MS_i - \hat{\sigma}_{pir,e}^2 - n_p \hat{\sigma}_{ir}^2 - n_r \hat{\sigma}_{pi}^2)}{n_r n_p} = \frac{[34/913 - 0/148 - 120(0/001) - 3(0/493)]}{(3 \times 120)} = 0/092$$

جدول شماره ۵: داده‌های حاصل از ANOVA و مقدار مؤلفه‌های واریانس در طرح اندازه‌گیری (PXIXR)

درصد واریانس	مؤلفه واریانس	MS	درجه آزادی	منبع تغییرپذیری
۲۸/۷	۰/۲۹۷	۶/۰۹۵	۱۱۹	P
۸/۹	۰/۰۹۲	۳۴/۹۱۳	۴	i
۰/۳	۰/۰۰۳	۲/۳۴۹	۲	r
۴۷/۷	۰/۴۹۳	۱/۶۲۸	۴۷۶	pi
۰/۰۰۰	۰/۰۰۰	۰/۱۴۹	۲۳۸	pr
۰/۰۰۰	۰/۰۰۱	۰/۳۵۹	۸	ir
		۱/۴۰	۹۸۲	pir.e

مؤلفه واریانس مربوط به هدف اندازه‌گیری یا واریانس حقیقی $28/7$ درصد از واریانس نمره مشاهده شده را به خود اختصاص داده است. این واریانس تفاوت فردی میان آزمودنی‌ها در درس علوم را نشان می‌دهد. مؤلفه واریانس سؤالهای آزمون $8/9$ درصد از واریانس را به خود اختصاص داده است. کوچک بودن این واریانس بیانگر آن است که سؤالهای آزمون در نظر سطح دشواری با یکدیگر تفاوت چندانی ندارند. مؤلفه اثر متقابل آزمودنی‌ها و سؤالها ($47/7$ درصد) نشان می‌دهد که وضعیت نسبی افراد از یک سؤال به سؤال دیگر فرق می‌کند. پاسخ صحیح یک فرد به یک سؤال دلیل پاسخ صحیح وی به سؤال دیگر نیست. ناچیز بودن مقدار مؤلفه واریانس تصحیح‌کننده‌ها ($0/2$ درصد) دلیل آن است که میانگین نمره افراد آزمودنی در سؤالهای آزمون توسط سه نفر تصحیح‌کننده تقریباً برابر است. صفر بودن مقدار مؤلفه‌های واریانس $\hat{\sigma}_{ir}^2$ و $\hat{\sigma}_{pr}^2$ نشان می‌دهد که نظر تصحیح‌کننده‌ها در نمره‌گذاری افراد مختلف آزمودنی و همچنین سؤالهای مختلف آزمون یکسان است. با توجه به تشریحی بودن سؤالهای آزمون علوم، این نتیجه جالب توجه است. مقدار مؤلفه واریانس $\hat{\sigma}_{pir,e}^2$ ترکیبی است از اثر متقابل r ، p و i و خطاهای غیر منظم و نامشخص در اندازه‌گیری. چون از هر ترکیب p ، i و r فقط یک مشاهده وجود دارد، نمی‌توان سهم هر یک از این دو عامل را جداگانه مشخص کرد. در طرح دورویه‌ای متقاطع واریانس‌های هر یک از قسمت‌های قرار گرفته شده در داخل دایره مربوط به هدف اندازه‌گیری (نمودار ۱ / ب) بر رتبه یا وضعیت نسبی فرد اثر می‌گذارد. بنابراین مقدار واریانس در تصمیم‌های نسبی برای یک تصحیح‌کننده در یک سؤال برابر است با:

$$\hat{\sigma}_{\text{نسبی}}^2 = \frac{\hat{\sigma}_{pr}^2}{n'_r} + \frac{\hat{\sigma}_{pi}^2}{n'_i} + \frac{\hat{\sigma}_{pir,e}^2}{(n'_r n'_i)} = \frac{0/000}{1} + \frac{0/493}{1} + \frac{0/148}{(1 \times 1)} = 0/641$$

تمام مؤلفه‌های واریانس، بااستثنای مؤلفه واریانس مربوط به هدف اندازه‌گیری (σ_p) بر وضعیت مطلق فرد اثر می‌گذارند. در نتیجه مقدار واریانس در تصمیم‌های مطلق برابر است با:

$$\sigma_{\text{مطلق}}^2 = \frac{\sigma_r^2}{n'_r} + \frac{\sigma_i^2}{n'_i} + \frac{\sigma_{pi}^2}{n'_i} + \frac{\sigma_{pr}^2}{n'_r} + \frac{\sigma_{ri}^2}{n'_r n'_i} + \frac{\sigma_{pir,e}^2}{(n'_r n'_i)}$$

نظریه تعمیم پذیری

$$= \frac{0/003}{1} + \frac{0/092}{1} + \frac{0/493}{1} + \frac{0/000}{1} + \frac{0/001}{1} + \frac{0/148}{(1 \times 1)} = 0/737$$

ضریب تعمیم پذیری، مقدار اتکا پذیری تعمیم نمره مشاهده شده یک فرد در یک سؤال توسط یک نمره گذار به نمره فرد در مجموعه مرجع مشاهده ها (سؤالها و تصحیح کننده ها) را نشان می دهد. ضریب تعمیم پذیری برای تصمیم گیری نسبی برابر است با واریانس نمره حقیقی (σ_p^2) به واریانس نمره مشاهده شده (واریانس نمره حقیقی + واریانس نسبی):

$$EP^2 = \frac{\hat{\sigma}_p}{\hat{\sigma}_p + \sigma^2_{نسب}} = \frac{0/297}{0/297 + 0/641} = 0/317$$

و برای تصمیم گیری مطلق برابر است با:

$$\phi = \frac{\sigma_p^2}{\sigma_p^2 + \sigma^2_{مطلق}} = \frac{0/297}{0/297 + 0/737} = 0/287$$

جدول شماره ۶ نتایج حاصل از مطالعه G و مطالعات D در طرح اندازه گیری (Pxixr)

منبع تغییر پذیری	$\hat{\sigma}^2$	مطالعه G			مطالعات D				
		$n_i = 1$	۵	۱۰	۵	۱۰	۵	۱۰	۱۵
		$n'_i = 1$	۱	۱	۲	۲	۳	۳	۱
P	$\hat{\sigma}^2_p$	0/297	0/297	0/297	0/297	0/297	0/297	0/297	0/297
i	$\hat{\sigma}^2_i$	0/09	0/018	0/09	0/018	0/09	0/018	0/09	0/06
r	$\hat{\sigma}^2_r$	0/003	0/003	0/003	0/002	0/002	0/001	0/001	0/001
P _i	$\hat{\sigma}^2_{P_i}$	0/493	0/99	0/49	0/99	0/49	0/99	0/49	0/33
P _r	$\hat{\sigma}^2_{P_r}$	0/000	0/000	0/000	0/000	0/000	0/000	0/000	0/000
r _i	$\hat{\sigma}^2_{r_i}$	0/001	0/000	0/000	0/000	0/000	0/000	0/000	0/000
Pir _{i,c}	$\hat{\sigma}^2_{Pir,c}$	0/148	0/30	0/15	0/15	0/07	0/1	0/05	0/10
$\hat{\sigma}^2_{نسب}$		0/636	0/129	0/64	0/114	0/056	0/109	0/054	0/49
$\hat{\sigma}^2_{مطلق}$		0/737	0/15	0/077	0/134	0/065	0/128	0/064	0/50
\hat{p}^2		0/317	0/697	0/823	0/724	0/841	0/731	0/846	0/86
ϕ		0/288	0/664	0/800	0/69	0/82	0/709	0/823	0/853

علوم تربیتی

کوچک بودن مقدار ضرایب تعمیم‌پذیری عدم دقت در تعمیم نمره مشاهده شده فرد در یک سؤال توسط یک تصحیح‌کننده به نمره وی در مجموعه مرجع مشاهده‌ها را نشان می‌دهد. برای مشخص کردن تعداد سؤالها و تعداد تصحیح‌کننده لازم برای رسیدن به ضریب مناسب مقدار مؤلفه‌های واریانس حاصل از مکالمه G را در مطالعات D مورد استفاده قرار می‌دهیم. برای این منظور تعداد سؤال و تصحیح‌کننده را هر بار تغییر داده و مقدار مؤلفه‌های واریانس و ضرایب تعمیم‌پذیری را محاسبه می‌کنیم. جدول شماره ۶ نتایج حاصل از مطالعه G و شش مطالعه مختلف D را نشان می‌دهد. باتوجه به اطلاعات جدول شماره ۶ با افزایش تعداد سؤالها به علت زیاد بودن مقدار مؤلفه σ^2_{pi} و تاحدودی σ^2_{ϵ} میزان خطاپذیری این دو مؤلفه کاهش یافته، و در نتیجه، ضریب تعمیم‌پذیری افزایش می‌یابد. ضریب تعمیم‌پذیری برای تصمیم‌گیری نسبی حاصل از ۱، ۵، و ۱۰ سؤال توسط یک تصحیح‌کننده (به ترتیب ۰/۳۱۷، ۰/۶۹۷ و ۰/۸۲۳) نشان می‌دهد که با افزایش تعداد سؤالهای آزمون می‌توان نمره حقیقی فرد را با دقت بیشتری اندازه‌گیری کرد.

ضریب تعمیم‌پذیری نسبی برای ۵ و ۱۰ سؤال توسط دو تصحیح‌کننده به ترتیب ۰/۷۲۴ و ۰/۸۴۱ و توسط سه تصحیح‌کننده به ترتیب ۰/۷۳۱ و ۰/۸۴۶ برآورد شده است. مقایسه ضرایب طرح‌های ۵ سؤالی با ۱، ۲ و ۳ نفر تصحیح‌کننده (به ترتیب ۰/۶۹۷، ۰/۷۲۴ و ۰/۷۳۱) با طرح‌های ۱۰ سؤالی با ۱، ۲ و ۳ نفر تصحیح‌کننده (به ترتیب ۰/۸۲۳، ۰/۸۴۱ و ۰/۸۴۶) نشان می‌دهد که با در نظر گرفتن هزینه و وقت لازم برای دو یا سه بار نمره‌گذاری، افزایش ضریب تعمیم‌پذیری ناچیز است. تفاوت ضریب حاصل از یک و دو تصحیح‌کننده حدود ۰/۰۲ و تفاوت ضریب حاصل از ۲ و ۳ تصحیح‌کننده کمتر از ۰/۰۱ است. تغییر پذیری ناچیز در مؤلفه‌های واریانس مربوط به تصحیح‌کننده و اثرهای متقابل تصحیح‌کننده با سایر عوامل (P و i) را می‌توان دلیل پایا یا دقیق بودن نمره‌گذاری توسط تصحیح‌کنندگان دانست.

باتوجه به تشریحی بودن سؤالهای آزمون علوم، می‌توان نتیجه گرفت که در

صورت فراهم آوردن یک دستورالعمل مناسب نمره‌گذاری و آموزش مؤثر - هر چند کوتاه مدت - به تصحیح‌کنندگان^۱، اوراق تشریحی را نیز با پایایی و دقت نسبتاً زیاد می‌توان نمره‌گذاری نمود. در این مطالعه برای رسیدن به ضریب تعمیم‌پذیری برای تصمیم‌گیرهای نسبی بیشتر از ۰/۸۴ به ۱۰ سؤال و ۲ تصحیح‌کننده و یا ۱۵ سؤال و یک تصحیح‌کننده نیاز هست. مقایسه هزینه و وقت لازم برای تصحیح ۱۰ سؤال توسط ۲ نفر یا اجرای ۱۵ سؤال و تصحیح توسط یک نفر به تصمیم‌گیرنده امکان می‌دهد تا طرح مناسب برای اندازه‌گیری توان حقیقی دانش آموزان در درس علوم را مشخص سازد.

نتیجه‌گیری

نظریه تعمیم‌پذیری با بهره‌گیری از طرح‌های مختلف تحلیل واریانس، ضمن برآورد مؤلفه‌های مختلف واریانس و مشخص کردن سهم هر یک از این مؤلفه‌ها در نمره مشاهده شده، مفهوم پایایی در قالب نظریه کلاسیک را گسترش داده است. برآورد مؤلفه‌های واریانس نمره مشاهده شده فرد در یک سؤال و یک سطح از سطوح هر یک از رویه‌های مورد استفاده در طرح اندازه‌گیری، به تهیه‌کننده آزمون امکان می‌دهد تا با افزایش تعداد سطوح رویه‌ای که مقدار مؤلفه واریانس آن و یا مقدار مؤلفه واریانس اثرهای متقابل آن زیاد است، مقدار یا مقادیر واریانس خطا را کاهش داده و از این طریق ضریب تعمیم‌پذیری نمره مشاهده شده در یک سؤال نمونه به نمره مجموعه مرجع سؤالهای ممکن را افزایش دهد.

در این پژوهش با استفاده از نظریه تعمیم‌پذیری، نتایج حاصل از نمره دانش آموزان پایه پنجم ابتدایی در درس علوم توسط دو طرح مختلف اندازه‌گیری مورد بررسی قرار گرفت. یافته‌های حاصل از هر دو طرح اندازه‌گیری نشان داد که خطای اندازه‌گیری تا حدودی از تغییرپذیری سؤالهای آزمون تأثیر می‌پذیرد. در طرح دو رویه‌ای سهم تغییرپذیری تصحیح‌کننده و اثرهای متقابل آن در مقایسه با سهم تغییرپذیری سؤالهای آزمون ناچیز و کم اهمیت بود. با توجه به یافته‌های حاصل از هر دو طرح اندازه‌گیری،

۱. در این پژوهش، سه نفر تصحیح‌کننده در یک جلسه سه ساعته ضمن تبادل نظر در مورد کلاسهای تصحیح سؤالها به صورت آزمایشی پاسخنامه ۵ نفر را در ۵ سؤال به صورت مستقل تصحیح و ضمن مقایسه نمره‌های

نمره مشاهده شده فرد در یک آزمون ۱۵ سؤال را می توان با اتکاپذیری بیش از ۰/۸۴ به مجموعه مرجع نمره ممکن تعمیم داد. از یافته های حاصل از این پژوهش می توان در تهیه آزمونهای پیشرفت تحصیلی علوم (یا سؤالهای مشابه سؤالهای آزمون مورد بحث در پژوهش) برای حجم زیاد آزمودنی بهره گرفت.

مآخذ

1. Suen, Hoi.H. *principles of Test Theories*. Lawrence Erlbaum Associates, publishers, 1990, P.8.
2. Crocker, L., and Algina, J. *Introduction to classical and Modern test. Theory*. Holt, Rinehart and winston, Inc, 1986, P. 115.
3. Suen, P. 31.
4. Ibid, P. 41.
5. Crocker, P. 159.
6. Suen, p.47.
7. Shavelson. R. J., Webb. N.M. and Rowley, G.L. "Generalizability theory", *American psychologist* (44), 1989, p. 66.
8. Kane, M.T., and Brennan, R.L. "The generalizability of class means". *Review of Educational Research*, (47), 1977.
9. Brennan, R.L. *Elements of Generalizability Iowa City, IA: American College Testing program*, 1983.
10. Cronbach, L.J., Gleser, G.C., Nanda, H., and Rajaratnam, N. *The dependability of Behavioral Measurement: Theory of Generalizability of Scores and profiles*. New York: John wiley, 1972, p.17.
11. Cronbach, p.82.