دکتر علیرضا کیامنش
دانشیار دانشگاه تربیت معلم

نظریّهٔ تعمیم‌پذیری در اندازه‌گیری آموزشی

به منظور تعیین تعداد ۱) سؤالهای لازم در آزمون علوم برای اندازه‌گیری توان حقیقی یا نمره حقیقی فرد درحد قابل قبول و ۲) تعداد تصحیح کنندهٔ لازم برای رسیدن به پایایی مناسب بین نظر تصحیح کننده‌ها، پاسخنامه‌های ۱۲۰ نفر دانش‌آموزکلاس پنجم ابتدایی در آزمون علوم از ۱۵ دبستان (هر دبستان ۸ نفر) به صورت تصادفی جمع‌آوری[1] و باتوجه 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 بانتیجهٔ آزمایش را با ذکر دلیل به صورت تشریحی در پاسخنامهٔ خود بیان کند.

گزارش پژوهش درقالب معرفّی نظریّهٔ تعمیم‌پذیری ارائه شده است. بدین‌معنی که ابتدا ضمن مقایسه مختصر نظریّهٔ کلاسیک و نظریه تعمیم‌پذیری ، مفاهیم و اصطلاحات مورد استفاده در نظریّه تعمیم‌پذیری موردبحث قرار گرفته است، سپس از طریق تحلیل واریانس "Anova" مؤلفّه‌های واریانس مربوط به هر یک از منابع تغییرپذیری درطرحهای اندازه‌گیری برآورد، و ضریب تعمیم‌پذیری نمره هر فرد در یک سؤال به مجموعه مرجع 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 . پاسخنامه‌های دانش آموزان قبل از تصحیح درمدارس ، جمع‌آوری و در پژوهش مورد استفاده قرارگرفته

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[1] و روایی[2] را ضروری می‌کند. اگر عملکرد آزمون، توان یا نظر پاسخ دهندگان را در موضوع موردنظر به صورت دقیق اندازه‌گیری نکند، یعنی :(۱)عملکرد فرد در آزمون با عملکرد حقیقی وی تفاوت داشته باشد یا(۲)عملکرد مشاهده شده معّرف عملکرد حقیقی باشد ولی آنچه اندازه گرفته شده منعکس کنندهٔ هدف موردنظر در اندازه‌گیری نباشد، استفاده از نتایج آزمون بی‌حاصل و گاه مضّر خواهد بود.

خصوصیات مختلف یک آزمون بااستفاده از تئوریها یا نظریّه های روان سنجی[3] مورد مطالعه قرار می‌گیرد. دو نظریّه عمده در روان‌سنجی عبارتند از: نظریّه نمونه‌گیری تصادفی[4] و نظریّه سؤال - پاسخ[5]. در نظریّه نمونه‌گیری تصادفی «به رابطهٔ میان عملکرد مشاهده شده و عملکرد حقیقی همانند تعمیم از یک نمونه به یک جامعه توجّه می‌شود. درحالیکه نظریّه سؤال - پاسخ برآن است که اگر چگونگی ارتباط هر سؤال آزمون با توانایی فرد مورد مطالعه درک شود، می‌توان نمرهٔ حقیقی فرد را به صورت مستقیم برآورد کرد» (۱). نظریّه نمونه‌گیری تصادفی خود شامل دو نظریّه نسبتاً متفاوت است: نظریّه کلاسیک[6] یا پایایی کلاسیک و نظریّه تعمیم‌پذیری[7].

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

---

1 . Reliability

2 . Validity

3 . Psychometric

4 . Random sampling theory

5 . Item- response theory

6 . Classical theory

نظریّهٔ کلاسیک

نظریّه کلاسیک نمرهٔ مشاهده شده از هر فرد در هر آزمون را ترکیبی از دو نمره حقیقی[1] فرد و خطای تصادفی[2] فرض می‌کند. تفکیک و برآورد سهم هر یک از این دو جز به صورت مستقیم امکان‌پذیر نیست، ولی با فرض تکرار عمل اندازه‌گیری در دفعات متّعدد و مشاهده یا جمع‌آوری نمره‌های متّعدد از یک فرد، مقدار مورد انتظار[3] نمرهٔ مشاهده شده (معّدل نمره‌های مشاهده شده در دفعات متعّدد) برابر نمره حقیقی فرد و مقدار مورد انتظار خطا صفر خواهد شد. در این نظریّه ، پایایی عبارت از میزان ارتباط بین نمرهٔ مشاهده شده و نمرهٔ حقیقی یا نسبت واریانس نمرهٔ واقعی به واریانس نمرهٔ مشاهده شده است. این نسبت بیانگر« همبستگی میان نمره‌های حقیقی و تمام نمره‌های مشاهده شده ممکن از تکرار دفعات متعدد آزمون است» (۲).

مفروضهٔ نظری بی‌نهایت تکرار آزمون ، در عمل به مفروضهٔ دو آزمون هم ارزیاهمتا[4] تبدیل می‌شود. اگر بتوان مفروضهٔ هم ارز بودن ، یعنی مساوی بودن نمرهٔ حقیقی هرفرد در دو آزمون و مساوی بودن واریانس خطای دو آزمون را برقرار کرد، از طریق اجرای دو آزمون هم ارز 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 هم ارز به مجذور همبستگی بین نمرهٔ مشاهده شده و نمرهٔ حقیقی تبدیل می‌شود» (۳).

ضریب محاسبه شده در نظریّه کلاسیک ضمن آنکه برمفروضهٔ هم ارز بودن آزمونها متکّی است، در هر بار اجرای آزمون تمام منابع مؤثّر 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 اندازه‌گیری‌بررسی می‌کند و واریانس به سؤالهای آزمون و اثرمتقابل میان‌سؤال و اجرای آزمون(تفاوت درپاسخگویی به یک سؤال در دو زمان) رانادیده‌می‌گیرد، درحالی که

---

1 . True score

2 . Random error

3 . Expected value

ضریب پایایی محاسبه شده از طریق اجرای دو آزمون هم ارز در دو زمان بافاصلهٔ نسبتاً کوتاه، فقط هم ارز بودن سؤالهای دو آزمون را به عنوان خطای اندازه‌گیری مورد توجه قرار می‌دهد و واریانس‌های دیگر را نادیده می‌گیرد. ضریب پایایی محاسبه شدهٔ سؤالهای آزمون از طریق روشهای تقسیم فقط دو یا چند جزء (هر سؤال یک جزء) را منبع خطای اندازه‌گیری می‌داند. بدین ترتیب در نظریّه کلاسیک، نمره‌های خطا برای محاسبهٔ ضریب پایایی باتوجه به موقعیّت خاص یا شیوهٔ خاص به کارگرفته شده تغییر کرده، درنتیجه مقدار ضریب پایایی نیز تغییر خواهد کرد. تفاوت در مقدار ضریب پایایی و همبستگی منفی میان ضریب پایایی و خطای معیار اندازه‌گیری ، به محاسبهٔ مقادیر مختلف خطای معیار اندازه‌گیری و درنتیجه تشکیل فاصله‌های متفاوت اطمینان برای نمرهٔ هر فرد آزمودنی منجر خواهد شد. از این رو، می‌توان گفت که نظریّه کلاسیک بانادیده گرفتن موقعّیت یا شرایط اجرای آزمون، خطای اندازه‌گیری را« از طریق یک مفهوم مبهم که به آن خطای تصادفی گفته می‌شود» (۴) تعریف می‌کند. باتوجه‌به‌محدودیت‌های‌موردبحث‌می‌توان‌به‌این‌نتیجه‌گیری‌اشاره‌کردکه«نظریّه‌کلاسیک باوجودتأثیرنسبتاًطولانی‌برحوزهٔاندازه‌گیریهای‌آموزشی‌وروانی‌وکاربرد عملی‌آن، فاقدانعطاف‌پذیری‌لازم برای حّل‌همهٔ‌مسائل‌مربوط به‌پایایی درحوزهٔاندازه‌گیری است(۵)

مفاهیم زیر بنایی نظریّهٔ تعمیم‌پذیری

رویه‌ها[1] و مؤلفّه‌های واریانس

نظریّه‌کلاسیک بدون توجه به زمینهٔ اجرای آزمون (زمان و شرایط اجرای آزمون، ثبات نظر تصحیح کنندگان یا ثبات نظر یک تصحیح کننده در دفعات مختلف تصحیح) ، نمرهٔ حقیقی را خصیصه‌ای ثابت فرض کرده برای هر فرد فقط یک نمرهٔ حقیقی محاسبه می‌کند. در نظریّه تعمیم‌پذیری تأثیر عوامل مختلف بر اجرای آزمون و نمره مشاهده شده مورد توجه قرارمی‌گیرد. درنتیجه‌برای هرفرد برحسب شرایط درنظرگرفته شده در آزمون، نمره‌های متفاوتی محاسبه می‌شود. در این نظریّه هر نمرهٔ مشاهده شده معّرف

یک نمونه از تمام نمره‌های ممکن یا مجموعهٔ مرجع نمره[1] فرض می‌شود که به وسیلهٔ یک یا چند واریانس بیان می‌شود. برای بیان ساده‌تر و درک سریع‌تر، مطالب مربوط به مفاهیم نظریّه تعمیم‌پذیری و محاسبات آماری در قالب چند مثال ارائه می‌شوند.

زمانی که تهیه کنندهٔ آزمون تصمیم می‌گیرد با چند سؤال (تشریحی یا عینی)، پیشرفت درسی دانش آموزان 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[2] یا سطوح[3] اندازه‌گیری که قصد داریم عملکرد آزمودنی‌ها را در این سطوح مورد سنجش قرار دهیم. بنابراین، برای تهیه کنندهٔ موردبحث سؤالهای آزمون رویه اندازه‌گیری و هر سؤال نیز یک سطح مورد قبول از رویهٔ اندازه‌گیری به حساب می‌آید. از این رویه تمام سطوح قابل قبول برای تهیه کنندهٔ آزمون، مجموعه مرجع سطوح قابل قبول[4] گفته می‌شود. بدین ترتیب نمرهٔ مشاهده شدهٔ سؤالهای آزمون فرد، یک نمونهٔ تصادفی از تمام سؤالهای ممکن یا مجموعهٔ مرجع سؤالهای قابل قبول فرض می‌شود و نظریّه تعمیم‌پذیری ، میزان اتکّاءپذیری[5] یا تعمیم پذیری نمرهٔ حاصل از یک سؤال نمونه را مجموعهٔ مرجع سؤالهای قابل قبول مشخص می‌کند.

اگر تهیه کنندهٔ آزمون ، علاوه‌بر سؤالهای آزمون از دو یا چند نفر تصحیح کننده برای نمره‌گذاری آزمون استفاده کند، تصحیح کننده یک رویه و هر فرد تصحیح کننده یک سطح از رویهٔ موردبحث را تشکیل می‌دهد. نمرهٔ داده شده توسط این افراد می‌تواند یک نمونهٔ تصادفی از تمام نمره‌هایی باشد که توسط مجموعهٔ مرجع تصحیح کنندگان قابل قبول به دست می‌آید. تغییر پذیری در نمرهٔ تصحیح کنندگان، بیانگر خطای اندازه‌گیری است که از عدم پیروی از معیار یکسان در نمره‌گذاری حاصل می‌گردد. در نظریّه کلاسیک، این منبع خطاپذیری را فقط بانادیده گرفتن منبع تغییر پذیری سؤالهای آزمون می‌توان محاسبه کرد. نظریّه تعمیم‌پذیری با تکیه برمیزان تغییرپذیری نمونه‌گیری مربوط

---

1 . Universw - score

2 . Conditions

3 . Levels

4 . Universe of admissible conditions

به سؤالها، تصحیح کننده‌ها (یا هر چند رویهٔ دیگر) و ترکیب‌های مختلف آنها با یکدیگر، میزان خطای اندازه‌گیری را در قالب مؤلّفه‌های 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 مؤلّفه را به صورت جداگانه برآورد می‌کند. درچنین حالتی مجموعهٔ مرجع نمره‌های قابل قبول[1] برای تهیهٔ کنندهٔ آزمون تمام سطوح رویه‌های مورد استفاده درآزمون را در برمی‌گیرد، و نمرهٔ مشاهده شدهٔ هرفرد به تمام سطوح ممکن یا قابل قبول تعمیم داده می‌شود.

باتوجه به دو نمونهٔ موردبحث، نمرهٔ 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 نمرهٔ مشاهده شده باتوجه به رویه‌های مورد استفاده در آزمون معرّف عملکرد فرد در همان رویه‌هاست. لذا اگر فقط یک رویه در آزمون به کار گرفته شود. طرح آزمون را می‌توان طرح یک رویه‌ای[2] نامید. در این طرح افراد آزمودنی هدف اندازه‌گیری بوده و تشخیص تفاوتهای فردی میان آنها تفاوت حقیقی یا واریانس نمرهٔ حقیقی بین آنها را به دست می‌دهد. تغییر پذیری نمونه‌گیری از سؤالها و اثر متقابل بین سؤال و فرد، هر دو عامل بالقوهٔ خطای اندازه‌گیری بوده که از مجموع واریانس‌های این دو عامل و واریانس نمرهٔ حقیقی ، واریانس نمرهٔ مشاهده شده محاسبه می‌شود. در این طرح نمرهٔ کسب شده به مجموعهٔ مرجع نمره‌های مورد قبول در رویهٔ سؤال (نمرهٔ 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 شمارهٔ الف ارائه شده است. در نمودار ۱ / الف ، p نشان دهندهٔ فرد آزمودنی یا هدف اندازه‌گیری، i نشان دهندهٔ سؤال و e نشان دهندهٔ خطاهای غیر سیستماتیک در اندازه‌گیری است. نمرهٔ مشاهده شدهٔ هر فرد (xpi) در هر سؤال ترکیبی است از:

$xpi =$	u	میانگین‌کل
$+(u_p = u)$	اثرفرد	σ^2_p یا نمره واقعی
$+(u_i - u)$	اثرسؤال	σ^2_i واریانس‌خطا
$+(x_{pi} - u_p - u_i + u)$	باقیمانده یاخطا	$\sigma^2_{pi,e}$

1 . Universe of admissible scores

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

$$\sigma^2(X_{pi})=\sigma^2_p+\sigma^2_i+\sigma^2_{pi,e}$$

اگر دو رویه به صورت همزمان در آزمون به کار گرفته شود، طرح آزمون را می‌توان طرحی دو رویه‌ای[1] دانست. در این طرح نیز افراد آزمودنی هدف اندازه‌گیری بوده و تفاوت فردی بین آنها واریانس نمرهٔ حقیقی محسوب می‌شود. تغییرپذیری نمونه‌گیری در رویه‌های طرح و اثرهای متقابل به وجود آمده همگی عامل‌های بالقّوهٔ خطای اندازه‌گیری به شمار می‌روند. درچنین طرحی، نمرهٔ کسب شده توّسط هر فرد به مجموعهٔ مرجع نمره های قابل قبول در همان دو رویه تعمیم داده می‌شود. در نمودار ۱/ب واریانس‌های مختلف حاصل از طرح دو رویه‌ای که تمام سطح‌های دو رویه و هدف آزمون (افراد) با هم متقاطع[2] هستند، ارائه شده است.

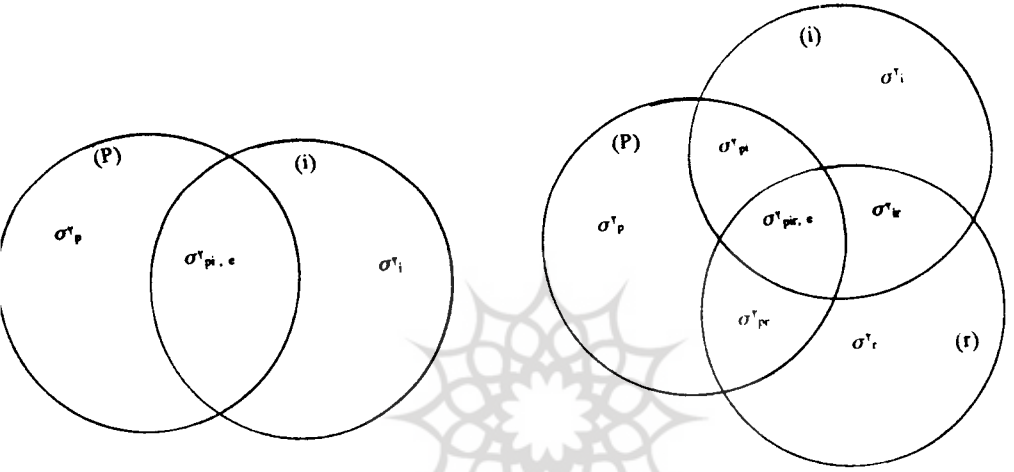
در نمودار ۱/ ب ، P آزمودنی، i سؤال (یا یکی از رویه‌ها)، r تصحیح کننده (یا رویهٔ دیگر) و سایر علامتها اثرهای متقابل محاسبه شده را نشان می‌دهند. در این طرح نمرهٔ مشاهده شدهٔ هر فرد در هر یک از خانه‌های جدول (برخوردیک سطح از یک رویه با یک سطح از رویهٔ دیگر برای هر فرد) یا X_{pir} برابر است با:

$X_{pir} = u$	میانگین		
$+(u_p-u)$	اثرفرد	واریانس واقعی	σ^2_p
$+(u_i-u)$	اثرسؤال	واریانس سؤال	σ^2_p
$+(u_r-u)$	اثر تصحیح‌کننده	واریانس تصحیح کننده	σ^2_r
$+(u_{pr}-u_p-u_r+u)$	اثر متقابل p,r		σ^2_{pr}
$+(u_{pi}-u_p-ui+u)$	اثرمتقابل i,p		σ^2_{pi}
$+(u_{ir}-u_i-u_r+u)$	اثرمتقابل r,i		σ^2_{ir}
$+(X_{pir}-u_{pi}-u_{pr}-u_{ri}+u_p+u_i+u_r-u)$	باقیمانده‌یاخطا		$\sigma^2_{pir,e}$

واریانس خطا

---

1 . Two facets design



پژوهشگاه علوم انسانی و مطالعات فرهنگی

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

نظریّه تعمیم‌پذیری با استفاده از تحلیل واریانس مقدار هر یک از مؤلّفه‌های واریانس را محاسبه می‌کند. واریانس نمرهٔ مشاهده شده برابر است با مجموع مقدار تمام مؤلّفه‌های واریانس:

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

طرح‌های متقاطع و طرح‌های آشیانه‌ای [۱]

در طرح یک رویه‌ای ارائه شده در نمودار ۱/ الف هرفرد آزمودنی (p) به تمام

سؤالهای آزمون (i) پاسخ می‌دهد. درچنین حالتی ارتباط بین p و i متقاطع طرح با Pxi مشخص می‌شود. در نمودار ۱ / ب هر فرد آزمودنی به تمام سؤالها پاسخ داده و هر فرد تصحیح کننده، به هر سئوال هر فرد نمره می‌دهد. در این حالت نیز ارتباط بین r,i,p متقاطع و طرح با pxixr مشخّص 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 مجموعهٔ 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[1]، برخلاف طرح یک رویه‌ای متقاطع که چهار مؤلّفهٔ واریانس را می‌توان برآورد نمود، فقط دو مؤلّفهٔ واریانس را می‌توان برآورد کرد (نمودار شمارهٔ۲).

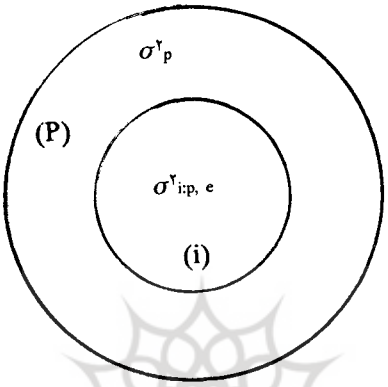
در این طرح سؤال مربوط به تغییرپذیری سؤالهای آزمون بدون پاسخ می‌ماند. نمرهٔ هر فرد آزمودنی X_{pi} برابر است با:

$X_{pi}=$	u	میانگین	
σ^2_p واریانس واقعی	$+(u_p-u)$	اثروفرد	
$\sigma^2_{i:pi,e}$ واریانس خطا	$+(X_{pi}-u_p)$	خطا	

و واریانس نمرهٔ مشاهده شده برابر است با مجموع دو مؤلّفهٔ برآورد شده:

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

از مقایسهٔ نمودارهای شماره ۱ / الف و ۲ نتیجه گرفته می‌شود که تفاوت عمدهٔ آماری بین این دو طرح در این است که اثر مستقیم مؤلّفهٔ واریانس مربوط به رویهٔ آشیانه‌ای را جدا از مؤلّفهٔ واریانس اثر متقابل نمی‌توان بررسی کرد. به عبارت دیگر در طرح آشیانه‌ای اثر مستقیم i همراه با اثر متقابل pi و سایر خطاهای نامشخّص برآورد می‌شود. بحث درمورد طرح‌های آشیانه‌ای دو رویه‌ای یا بیشتر به علّت گستردگی ، از حوصلهٔ این مقاله خارج است.



نمودار شمارهٔ ۲ . طرح یک رویه‌ای آشیانه‌ای

رویه‌های ثابت[۱]، تصادفی و مجموعهٔ مرجع نمره

اگر سطوح رویهٔ موردبررسی در اندازه‌گیری به صورت تصادفی از میان مجموعه مرجع سطوح ممکن آن رویه انتخاب شده باشند، رویه را تصادفی می‌نامند. درچنین حالتی سطوح انتخاب شده برای مطالعه درهربار اجرای آزمون متفاوت خواهند بود. مثلاً هر بار تعداد خاصّی از سؤالهای آزمون یا هر بار تعداد خاصّی از افراد تصحیح کننده از میان سطوح قابل قبول برای سؤال یا تصحیح کننده 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 رویهٔ ثابت در اندازه‌گیری استفاده می‌شود. با استفاده از رویهٔ ثابت "هدف ارزشیابی را می‌توان با پایایی و روایی بیشتری اندازه‌گیری کرد» (۶) باوجود این، "تعمیم‌پذیری یک نظریّه اندازه‌گیری با رویهٔ تصادفی

است و در هر طرح اندازه‌گیری حداقل یکی از رویه‌ها باید تصادفی باشد.» (۷) در بیشتر مطالعات تعمیم‌پذیری افراد یا آزمودنی‌ها هدف اندازه‌گیری[1] محسوب می‌شوند. ولی گاه از عوامل دیگر نیز می‌توان به عنوان هدف اندازه‌گیری استفاده کرد. (۸) مثلاً در مطالعات ارزشیابی معمولاً کلاسها هدف اندازه‌گیری و دانش آموزان و سایر رویه‌ها (دانش آموزان در این وضعیت رویه محسوب می‌شوند)، مجموعهٔ مرجع تعمیم[2] را تشکیل می‌دهند. در هر صورت مجموعه مرجع نمره عبارت است از میانگین اعضای هدف اندازه‌گیری در تمام شرایط موجود در مجموعهٔ مرجع تعمیم. واریانس حاصل از نمره‌های جامعه[3] هدف اندازه‌گیری (آزمودنی یا کلاس) یعنی واریانس مجموعه مرجع نمرهٔ مشابه واریانس نمرهٔ حقیقی در نظریّه کلاسیک است. (۹) قصد هر اندازه‌گیری فراهم آوردن شرایطی است که مقدار واریانس نمرهٔ حقیقی (واریانس مجموعهٔ مرجع نمره) زیاد و مقادیر سایر مؤلّفه‌های 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. بدین‌ترتیب ، باتوجه به رویه‌ها و سطوح به کار گرفته شده در اندازه‌گیری، میزان نمرهٔ حقیقی تعیین می‌شود و تغییر در تعداد هر یک از رویه‌ها یا سطوح رویه‌ها به تغییر در مجموعهٔ مرجع نمره و مجموعه مرجع تعمیم منجر خواهد شد. به عبارت دیگر میزان واریانس مجموعهٔ مرجع نمره در نظریّه تعمیم‌پذیری باتوجه به مجموعه مرجع تعمیم تغییر می‌کند. برهمین روال تعداد و میزان مؤلّفه‌های واریانس خطا و سرانجام مقدار ضریب تعمیم‌پذیری[4] برای سازه[5] مورد نظر فرق خواهد کرد.

مطالعات تعمیم‌پذیری و مطالعات تصمیم‌گیری[6]

پس از برآورد مقادیر مؤلّفه‌های واریانس در مطالعهٔ تعمیم‌پذیری و مشاهدهٔ مقدار

---

1 . Object of measurement

2 . Universe of generalization

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

4 . Generaliza bility Cefficient

5 . Construct

واریانس‌های مربوط به مؤلّفهٔ می‌توان با تغییر در تعداد سطوح رویه یا رویه‌هایی 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[1] در نظریّه کلاسیک تا حدودی معادل دانست.

محاسبهٔ مؤلّفه‌های واریانس و ضرایب تعمیم‌پذیری

نظریّه تعمیم‌پذیری را می‌توان در چارچوب تحلیل واریانس "ANOVA" بررسی نمود.[2] در دنبالهٔ مقاله با استفاده از داده‌های جمع‌آوری شده از نمرهٔ ۱۲۰ دانش آموز، چگونگی محاسبهٔ مقادیر مؤلّفه‌های واریانس ضریب تعمیم‌پذیری و همچنین مطالعات تصمیم‌گیری (مطالعات D)، در قالب دو طرح اندازه‌گیری برای پاسخگویی به سؤالهای مطرح شده در پژوهش، موردبحث قرار می‌گیرد.

طرح اندازه‌گیری یک رویه‌ای

به منظور پاسخگویی به سؤال پژوهش : «تعداد سؤالهای موردنیاز برای رسیدن به ضریب تعمیم‌پذیری مناسب » نتایج حاصل از نمرهٔ دانش آموزان ابتدا در قالب طرح اندازه‌گیری یک رویه‌ای (آزمودنی P ، سئوال i) موردتجزیه وتحلیل قرار گرفته است. به عبارت دیگر در طرح PXi فقط نمره‌های حاصل از تصحیح یک فرد نمره‌گذار موردتوجه و بررسی قرار گرفته است. در این طرح چون تمام افراد به تمام سؤالهای آزمون پاسخ داده‌اند، طرح اندازه‌گیری از نوع یک رویه‌ای متقاطع است. با استفاده از تحلیل واریانس یک عاملی، اطّلاعات مربوط به مجموع مجذورات (SS)، درجات آزادی (df) و میانگین مجذورات (MS) برای هر یک از منابع تغییر پذیری محاسبه می‌شود. جدول شماره ۱ چارچوب لازم برای این اطلاعات را نشان می‌دهد. آخرین ستون جدول برخلاف جدول خلاصهٔ اطلاعات "ANOVA"که به آزمون F اختصاص دارد، به مقدار نظری مجموع

---

1 . Spearman - brown prophecy formula

مجذورات اختصاص داده شده است. مقدار نظری مجموع مجذورات مربوط به هر عامل تغییرپذیری با استفاده از مقادیر واریانس‌های گروه نمونه موردبررسی (نه جامعه) برآورد می‌شود.

جدول شماره ۱. اطلاعات ANOVA برای طرح اندازه‌گیری (Pxi)

منبع تغییرپذیری	ss	df	MS	E(MS)
آزمودنی(p)	SSp	n_p-1	$\frac{SSP}{np-1}$	$E(MS_p)=\sigma^2_{pi,e}+n_i\sigma^2_p$
سؤال (i)	SSi	n_i-1	$\frac{SSi}{n_i-1}$	$E(MS_i)=\sigma^2_{pi,e}+n_p\sigma^2 i$
اثرمتقابل pi و باقیمانده(pi,e)	$SS_{pi,e}$	$(n_p-1)(n_i-1)$	$\frac{SS_{pi,e}}{(n_p-1)(n_i-1)}$	$E(MS_{pi,e})=\sigma^2_{pi,e}$

با استفاده از نرم افزارهای کامپیوتری نظیر BMDP-8V ، اطلاعات مربوط به تمام ستونهای جدول شماره ۱ را می‌توان مستقیماً دریافت نمود.[۱] درصورت عدم دسترسی به نرم افزارهای کامپیوتری، اطلاعات مربوط به سه ستون سمت چپ جدول شماره ۱ از طریق محاسبات آماری "ANOVA" و اطلاعات مربوط به آخرین ستون جدول با استفاده از نتایج به دست آمده از "ANOVA" و فرمول‌های داده شده در جدول شمارهٔ ۱ محاسبه می‌شود. برای محاسبهٔ مقدار واریانس هر یک از منابع تغییرپذیری، از پایین جدول یعنی $M_{spi,e}$ شروع می‌کنیم. مقدار واریانس نمونهٔ اثر متقابل آزمودنی سؤال و خطاهای غیر سیستماتیک $(\sigma^2_{pi,e})$ برابر است با $M_{spi,e}$ ، پس :

$$\hat{\sigma}^2_{pi,e} = MS_{pi,e} = ۰/۶۱۳$$

با جا به جایی در فرمول $E(MS_i)$، مقدار $\hat{\sigma}^2_i$ برابر است با:

---

۱. نرم افزارهای SYSTAT ، SAS ، SPSS نیز اطلاعات مربوط به تحلیل واریانس را می‌توان

$$\hat{\sigma}^2_i = \frac{(MS_i - \hat{\sigma}^2_{pi,e})}{n_p} = \frac{(۱۱/۶۹۴ - ۰/۶۱۳)}{۱۲۰} = ۰/۰۹۲$$

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

$$\hat{\sigma}^2_p = \frac{(MS_p - \hat{\sigma}^2 pi,e)}{n_i} = \frac{(۲/۲۸ - ۰/۶۱۳)}{۵} = ۰/۳۳۳$$

چون مقدار واریانس از مقدار بارم سؤال تأثیر می‌پذیرد، برای تفسیر مؤلّفه‌های واریانس این مقادیر به فراوانی نسبی یا درصدی تبدیل می‌شوند. باتوّجه به جدول شمارهٔ 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 $\frac{۱}{۴}$ واریانس برآورد شده برای آزمودنی هاست (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 موفّق‌تر بوده‌اند.

جدول شمارهٔ ۲ . داده‌های حاصل از ANOVA و مقدار مؤلّفه‌های واریانس طرح اندازه‌گیری (Pxi)

درصدواریانس	مقدارمؤلّفه واریانس	MS	df	ss	منبع تغییرپذیری
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 مؤلّفه‌های واریانس را می‌توان به دو صورت تفسیر کرد: تفسیر هنجاری یا نسبی[1] و تفسیر ملاکی یا مطلق[2]. تفسیرهای نسبی به رتبه بندی و مقایسهٔ افراد در آزمون توجه دارند. درچنین حالتی، واژهٔ خطا باتوجه به مؤلّفه‌هایی که بروضعّیت 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 مجموعهٔ 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 مؤلّفهٔ واریانس سؤال (σ^2_i) 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[3] آزمون برابر است با:

$$\sigma^2_{\text{نسبی}} = \frac{\sigma^2_{pi,e}}{n'i} = \frac{۰/۶۳۱}{۱} = ۰/۶۱۳$$

و مقدار واریانس در تصمیم‌های مطلق برابر است با:

$$\sigma^2_{\text{مطلق}} = \frac{\sigma^2 i}{n'i} + \frac{\sigma^2_{pi,e}}{n'i} = \frac{۰/۰۹۲}{۱} + \frac{۰/۶۳۱}{۱} = ۰/۷۰۵$$

مطالعه تعمیم‌پذیری (مطالعه G) ضمن آنکه به مؤلفه‌های واریانس و تفسیر این مؤلفه‌ها توجه دارد، همانند نظریّه 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

n' مقدار D مطالعات ... باتوجه به طرح اندازه‌گیری و تعداد سطوح رویه

ضریب تعمیم‌پذیری تقریباً برابر است با n مقدار نظری مجذور همبستگی بین نمره مشاهده شده مجموعهٔ 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^2_p) و واریانس خطای اندازه‌گیری که بر وضعیت نسبی فرد اثر می‌گذارند. در طرح مورد بحث تنها $\sigma^2_{pi,e}$ بروضعیت نسبی فرد اثر می‌گذارد، درنتیجه ضریب تعمیم‌پذیری در طرح یک وجهی متقاطع برای تصمیم‌گیری نسبی برابر است با:

$$FP^2_{\text{نسبی}} = \frac{\sigma^2_p}{(\sigma^2_p + \sigma^2_{\text{نسبی}})} = \frac{۰/۳۳۳}{(۰/۳۳۳ + ۰/۶۱۳)} = ۰/۳۵۲$$

و در تصمیم‌گیری مطلق برابر است با:

$$\hat{\phi} = \frac{\sigma^2_p}{(\sigma^2_p + \sigma^2_{\text{مطلق}})} = \frac{۰/۳۳۳}{(۰/۳۳۳ + ۰/۷۰۵)} = ۰/۳۲$$

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

کوچک بودن مقدار ضرایب تعمیم‌پذیری (حداکثر یک و حداقل صفر) عدم دقّت در تعمیم نمرهٔ فرد در یک سؤال به نمرهٔ وی در مجموعهٔ مرجع سؤالها را نشان می‌دهد. به عبارت دیگر نمرهٔ مشاهده شدهٔ فرد در یک سؤال را نمی‌توان نمرهٔ حقیقی وی قلمداد کرد. برای مشخص کردن تعداد سؤالهای لازم و رسیدن به ضریب تعمیم‌پذیری مطلوب (فرمول اسپیرمن - براون در نظریّه کلاسیک همین نقش را ایفا می‌کند)، مقدار مؤلّفه‌های واریانس حاصل از مطالعهٔ 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 سؤالها و اثر متقابل کاهش می‌یابد. درنتیجه واریانس

سؤال داشته باشد (n'i= ۵) ضرایب تعمیم‌پذیری نسبی و مطلق به ترتیب ۰/۷۳۱ و ۰/۷۰۴ و اگر ۱۰ سؤال داشته باشد (n'i=۱۰) این ضرایب به ترتیب ۰/۸۴۵ و ۰/۸۲۶ خواهند شد. اگر هزینه، وقت، و نیروی انسانی اجازه دهد با افزایش تعداد سؤالها می‌توان ضرایب مناسب‌تری به دست آورد. باتوجه به جدول شماره ۳ برای رسیدن به ضریب تعمیم‌پذیری بالاتراز۰/۸۵، حداقل به ۱۵ سؤال ازنوع سؤالهای مورداستفاده درآزمون علوم نیازداریم.

طرح اندازه‌گیری دو رویه‌ای

برای پاسخگویی به سؤالهای پژوهش: (۱) تعداد سؤال مورد نیاز (۲) تعداد تصحیح کنندهٔ موردنیاز، نتایج حاصل از نمرهٔ دانش آموزان در قالب طرح اندازه‌گیری دو رویه‌ای 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، رابطهٔ بین هدف اندازه‌گیری (p) و رویه‌های طرح (r,i) یک رابطهٔ متقاطع (P×i×r) است.

نمودار ۱ / ب). در طرح دو رویه‌ای متقاطع ۷ مؤلّفهٔ واریانس به صورت همزمان برآورد می‌شود.[۱] سه مؤلّفه برای اثرهای مستقیم (r,i,p)، سه مولّفه برای اثرهای متقابل (ri,pr,pi) و یک مؤلّفه برای اثر متقابل سه عامل همراه با واریانس خطاهای غیر سیستماتیک در اندازه‌گیری (pir, e).

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

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

$$\hat{\sigma}^2_{pir,e} = MS_{pir,e} = ۰/۱۴۸$$

جدول شماره ۳. نتایج حاصل از مطالعهٔ G و مطالعات D در طرح اندازه‌گیری (P×i)

منبع تغییرپذیری	$\hat{\sigma}^2$	مطالعهٔ G	مطالعات D			
		$n_i=1$	۵	۸	۱۰	۱۵
آزمودنی	$\hat{\sigma}^2_p$	۰/۳۳۳	۰/۳۳۳	۰/۳۳۳	۰/۳۳۳	۰/۳۳۳
سئوال	$\hat{\sigma}^2_i$	۰/۰۹۲	۰/۰۱۸	۰/۰۱۱	۰/۰۰۹	۰/۰۰۶
اثرمتقابل وباقیمانده	$\hat{\sigma}^2_{pi,e}$	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 اندازه‌گیری (Pxixr)

منبع تغییرپذیری	SS	df	MS	E(MS)
(P) آزمودنی			$\frac{SSp}{dfp}$	$E(MS_p)=\sigma^2_{pri,e}+n_i\sigma^2_{pr}+n_r\sigma^2_{pi}+n_in_r\sigma^2_p$
(i) سؤال			$\frac{SS_i}{df_i}$	$E(ms_i)=\sigma^2_{pri,e}+n_r\sigma^2_{pi}+n_p\sigma^2_ir+n_pn_r\sigma^2_i$
(r) تصحیح کننده			$\frac{SS_r}{dfr}$	$E(MS_r)=\sigma^2_{pri,e}+n_i\sigma^2_{pr}+n_p\sigma^2_{ir}+n_pn_i\sigma^2_r$
Pi			$\frac{SS_{pi}}{df_{pi}}$	$E(MS_{pi})=\sigma^2_{pri,e}+n_r\sigma^2_{pi}$
pr			$\frac{SS_{pr}}{df_{pr}}$	$E(MS_{pi})=\sigma^2_{pri,e}+n_i\sigma^2_{pr}$
ir			$\frac{SS_{ir}}{df_{ir}}$	$E(MS_{pi})=\sigma^2_{pri,e}+n_p\sigma^2_{ir}$
Pir,e			$\frac{SS_{pir,e}}{dfpir,e}$	$E(MS_{pir,e})=\sigma^2_{pir,e}$

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

$$\hat{\sigma}^2_{pi} = \frac{(Ms_{pi} - \hat{\sigma}^2 pir,e)}{r} = \frac{(۱/۶۲۸ - ۰/۱۴۸}{۳}) = ۰/۴۹۳$$

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 شمارهٔ ۴ برآورد می‌شود. مثلا مقدار $\hat{\sigma}^2 i$ برابر است با:

$$\hat{\sigma}^2_i = \frac{(MS_i - \hat{\sigma}^2_{pir,e} - n_p \hat{\sigma}^2_{ir} - n_r \hat{\sigma}^2_{pi})}{n_r n_p} = \frac{[۳۴/۹۱۳ - ۰/۱۴۸ - ۱۲۰(۰/۰۰۱) - ۳(۰/۴۹۳)]}{(۳ \times ۱۲۰)} = ۰/۰۹۲$$

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

منبع تغییرپذیری	درجهٔ‌آزادی	MS	مؤلّفه‌واریانس	درصدواریانس
P	۱۱۹	۶/۰۹۵	۰/۲۹۷	۲۸/۷
i	۴	۳۴/۹۱۳	۰/۰۹۲	۸/۹
r	۲	۲/۳۴۹	۰/۰۰۳	۰/۳
pi	۴۷۶	۱/۶۲۸	۰/۴۹۳	۴۷/۷
pr	۲۳۸	۰/۱۴۹	۰/۰۰۰	۰/۰۰۰
ir	۸	۰/۳۵۹	۰/۰۰۱	۰/۰۰۰
pir,e	۹۵۲	/۱۴۰		

مؤلّفهٔ واریانس مربوط به هدف اندازه‌گیری یا واریانس حقیقی ۲۸/۷ درصد از واریانس نمرهٔ مشاهده شده را به خود اختصاص داده است. این واریانس تفاوت فردی میان آزمودنی‌ها در درس علوم را نشان می‌دهد. مؤلّفهٔ واریانس سؤالهای آزمون ۸/۹ 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. مؤلّفه اثر متقابل آزمودنی‌ها و سؤالها (۴۷/۷ درصد) نشان می‌دهد که وضعیّت 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 مؤلّفه واریانس تصحیح کننده ها (۰/۲ درصد) دلیل آن است که میانگین نمرهٔ افراد آزمودنی در سؤالهای آزمون توسط سه نفر تصحیح کننده تقریباً برابر است. صفر بودن مقدار مؤلفه‌های واریانس $\hat{\sigma}^2_{pr}$ و $\hat{\sigma}^2_{ir}$ نشان می‌دهد که نظر تصحیح کننده‌ها در نمره‌گذاری افراد مختلف آزمودنی و همچنین سؤالهای مختلف آزمون یکسان است. باتوجه به تشریحی بودن سؤالهای آزمون علوم، این نتیجه جالب توجه است. مقدار مؤلّفه واریانس $\hat{\sigma}^2_{pir,e}$ ترکیبی است از اثر متقابل p ، r و i و خطاهای غیر منظّم 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 دو رویه‌ای متقاطع واریانس‌های هر یک از قسمت های قرار گرفته شده در داخل دایره مربوط به هدف اندازه‌گیری (نمودار ۱ / ب) بر رتبه یا وضعیّت نسبی فرد اثر می‌گذارد. بنابراین مقدار واریانس در تصمیم‌های نسبی برای یک تصحیح کننده در یک سؤال برابر است با:

$$\hat{\sigma}^2_{\text{نسبی}} = \frac{\hat{\sigma}^2_{pr}}{n'_r} + \frac{\hat{\sigma}^2_{pi}}{n'_i} + \frac{\hat{\sigma}^2_{pir,e}}{(n'_r n'_i)} = \frac{۰/۰۰۰}{۱} + \frac{۰/۴۹۳}{۱} + \frac{۰/۱۴۸}{(۱\times۱)} = ۰/۶۴۱$$

تمام مؤلّفه‌های واریانس، باستثنای مؤلّفه واریانس مربوط به هدف اندازه‌گیری (σ^2_p) بر وضعیت مطلق فرد اثر می‌گذارند. درنتیجه مقدار واریانس در تصمیم‌های مطلق برابر است با:

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

$$= \frac{۰/۰۰۳}{۱} + \frac{۰/۰۹۲}{۱} + \frac{۰/۴۹۳}{۱} + \frac{۰/۰۰۰}{۱} + \frac{۰/۰۰۱}{۱} + \frac{۰/۱۴۸}{(۱\times۱)} = ۰/۷۳۷$$

ضریب تعیمم‌پذیری ، مقدار اتّکاپذیری تعمیم نمرهٔ مشاهده شدهٔ یک فرد در یک سؤال توسط یک نمره‌گذار به نمره فرد در مجموعه مرجع مشاهده‌ها (سؤالها و تصحیح کننده ها) را نشان می‌دهد. ضریب تعمیم‌پذیری برای تصمیم‌گیری نسبی برابر است با واریانس نمره حقیقی (σ^2_p) به واریانس نمرهٔ مشاهده شده (واریانس نمره حقیقی + واریانس نسبی):

$$EP^2 = \frac{\hat{\sigma}^2 p}{\hat{\sigma}^2 p + \sigma^2_{\text{نسبی}}} = \frac{۰/۲۹۷}{۰/۲۹۷ + ۰/۶۴۱} = ۰/۳۱۷$$

و برای تصمیم‌گیری مطلق برابر است با:

$$\phi = \frac{\sigma^2_p}{\sigma^2_p + \sigma^2_{\text{مطلق}}} = \frac{۰/۲۹۷}{۰/۲۹۷ + ۰/۷۳۷} = ۰/۲۸۷$$

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

منبع تغییر		مطالعه G			مطالعات D				
	$\hat{\sigma}^2$	$n'_i = ۱$	۵	۱۰	۵	۱۰	۵	۱۰	۱۵
پذیری		$n'_r = ۱$	۱	۱	۲	۲	۳	۳	۱
P	$\hat{\sigma}^2 p$	۰/۲۹۷	۰/۲۹۷	۰/۲۹۷	۰/۲۹۷	۰/۲۹۷	۰/۲۹۷	۰/۲۹۷	۰/۲۹۷
i	$\hat{\sigma}^2_i$	۰/۰۹	۰/۰۱۸	۰/۰۰۹	۰/۰۱۸	۰/۰۰۹	۰/۰۱۸	۰/۰۰۹	۰/۰۰۶
r	$\hat{\sigma}^2_r$	۰/۰۰۳	۰/۰۰۳	۰/۰۰۳	۰/۰۰۲	۰/۰۰۲	۰/۰۰۱	۰/۰۰۱	۰/۰۰۱
Pi	$\hat{\sigma}^2_{pi}$	۰/۴۹۳	۰/۰۹۹	۰/۰۴۹	۰/۰۹۹	۰/۰۴۹	۰/۰۹۹	۰/۰۴۹	۰/۰۳۳
pr	$\hat{\sigma}^2_{pr}$	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰
ir	$\hat{\sigma}^2_{ir}$	۰/۰۰۱	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰	۰/۰۰۰
Pir,e	$\hat{\sigma}^2_{Pir,e}$	۰/۱۴۸	۰/۰۳۰	۰/۰۱۵	۰/۰۱۵	۰/۰۰۷	۰/۰۱	۰/۰۰۵	۰/۰۱۰
$\hat{\sigma}^2_{\text{نسبی}}$		۰/۶۳۶	۰/۱۲۹	۰/۰۶۴	۰/۱۱۴	۰/۰۵۶	۰/۱۰۹	۰/۰۵۴	۰/۰۴۹
$\hat{\sigma}^2_{\text{مطلق}}$		۰/۷۳۲	۰/۱۵	۰/۰۷۷	۰/۱۳۴	۰/۰۶۵	۰/۱۲۸	۰/۰۶۴	۰/۰۵۰
$\hat{\rho}^2$		۰/۳۱۷	۰/۶۹۷	۰/۸۲۳	۰/۷۲۴	۰/۸۴۱	۰/۷۳۱	۰/۸۴۶	۰/۸۶
ϕ		۰/۲۸۸	۰/۶۶۴	۰/۸۰۰	۰/۶۹	۰/۸۲	۰/۷۰۹	۰/۸۲۳	۰/۸۵۳

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

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

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

صورت فراهم آوردن یک دستورالعمل مناسب نمره‌گذاری و آموزش مؤثّر- هر چند کوتاه مدت - به تصحیح کنندگان [۱]، اوراق تشریحی را نیز با پایایی و دقت نسبتاً زیاد می‌توان نمره‌گذاری نمود. در این مطالعه برای رسیدن به ضریب تعمیم‌پذیری برای تصمیم‌گیریهای نسبی بیشتر از ۰/۸۴ به ۱۰ سئوال و ۲ تصحیح کننده و یا ۱۵ سؤال و یک تصحیح کننده نیاز هست. مقایسهٔ 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 طرح‌مناسب‌برای‌اندازه‌گیری توان حقیقی دانش آموزان در درس علوم را مشخص سازد.

نتیجه گیری

نظریّه تعمیم‌پذیری با بهره‌گیری از طرح‌های مختلف تحلیل واریانس، ضمن برآورد مؤلّفه‌های مختلف واریانس و مشخص کردن سهم هر یک از این مؤلّفه‌ها در نمرهٔ مشاهده شده، مفهوم پایایی در قالب نظریّه کلاسیک را گسترش داده است. برآورد مؤلفه‌های واریانس نمره مشاهدهٔ شده فرد در یک سؤال و یک سطح از سطوح هر یک از رویه‌های مورد استفاده در طرح اندازه‌گیری ، به تهیه کنندهٔ آزمون امکان می‌دهد تا با افزایش تعداد سطوح رویه‌ای که مقدار مؤلّفهٔ واریانس آن و یا مقدار مؤلّفه واریانس اثرهای متقابل آن زیاد است، مقدار یا مقادیر واریانس خطا راکاهش داده و از این طریق ضریب تعیمم‌پذیری نمرهٔ مشاهده شده در یک سؤال نمونه به نمره مجموعهٔ مرجع سؤالهای ممکن را افزایش دهد.

در این پژوهش با استفاده از نظریّه تعمیم‌پذیری، نتایج حاصل از نمرهٔ دانش آموزان پایهٔ پنجم ابتدایی در درس علوم توّسط دو طرح مختلف اندازه‌گیری موردبررسی قرارگرفت. یافته‌های حاصل از هر دو طرح اندازه‌گیری نشان داد که خطای اندازه‌گیری تا حدودی از تغییر پذیری سؤالهای آزمون تأثیر می‌پذیرد. در طرح دو رویه‌ای سهم تغییرپذیری تصحیح کننده و اثرهای متقابل آن در مقایسه با سهم تغییرپذیری سؤالهای آزمون ناچیز و کم اهمیت بود. باتوجه به یافته‌های حاصل از هر دو طرح اندازه‌گیری ،

---

۱ . در این پژوهش، سه نفر تصحیح کننده در یک جلسه سه ساعتی ضمن تبادل نظر درمورد کلاسهای تصحیح سؤالها به صورت آزمایشی پاسخنامه ۵ نفر را در ۵ سؤال به صورت مستقّل تصحیح و ضمن مقایسهٔ نمره های

نمرهٔ مشاهده شدهٔ فرد در یک آزمون ۱۵ سؤال را می‌توان با اتکاپذیری بیش از ۰/۸۴ به مجموعه مرجع نمره ممکن تعمیم داد. از یافته‌های حاصل از این پژوهش می‌توان در تهیه آزمونهای پیشرفت تحصیلی علوم (یا سؤالهای مشابه سؤالهای آزمون موردبحث 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.