

مقایسه پیشرفت تحصیلی دانشجویان دبیری دانشگاه شهید چمران در آزمونهای چند-گزینه‌ای و صحیح - غلط در درس سنجش و اندازه‌گیری با کنترل غیبت‌های آنها

دکتر حسین سپاسی*

چکیده

امروزه آموزگاران، دبیران، و اساتید دانشگاه از یک روش واحد برای اندازه‌گیری آموخته‌های فراگیران خود استفاده نمی‌کنند. برخی سؤالات آزمون انشایی را بر سؤالات آزمونهای عینی بویژه سؤالات چهارگزینه‌ای و یا صحیح - غلط ترجیح می‌دهند. با این که سؤالات چهارگزینه‌ای از شهرت بسیاری برخوردارند، برخی از روانسنان معتقدند که تفاوتی بین این سؤالات و سؤالات صحیح - غلط وجود ندارد. در این تحقیق پیشرفت دانشجویان به وسیله دو نوع سؤال (چهار-گزینه‌ای و صحیح غلط) مورد مقایسه قرار گرفت. ۳۰۴ دانشجو به طور تصادفی به دو گروه مساوی ۱۵۲ نفری تقسیم شدند. در هر گروه ۱۰۰ پسر و ۵۲ دختر قرار داشت. از دو فرم آزمون یکی چهارگزینه‌ای و دیگر صحیح-غلط که هر دو یک محتوا را اندازه‌گیری می‌کردند به عنوان ابزارهای تحقیق استفاده شد. از سه روش متفاوت برای نمره‌گذاری سؤالات دو آزمون استفاده گردید. نتایج تحلیل یافته‌ها نشان داد که وقتی سؤالات دو آزمون به روش سنتی (روش اول) نمره‌گذاری شد و یا این که برای برابر کردن شانس کسری معادل $\frac{0}{1}$ به نمرات اضافه شد (روش سوم) اختلاف معنی‌داری بین نمره‌های ترم دانشجویان در آزمونهای چند-گزینه‌ای و صحیح-غلط مشاهده نگردید. ولی زمانی که به پاسخهای غلط نمره منفی تعلق گرفت (روش دوم)، بین الگوهای پاسخ دانشجویان پسر و دختر تفاوت معنی‌داری مشاهده گردید. در کل پسران عملکرد بهتری از دختران نشان دادند. با این حال، ترتیب رتبه نمره‌ها در هر سه روش نمره‌گذاری یکسان باقی ماند.

مقدمه

بررسی ابعاد یک حیطه یا یک سازه روانی در فرآیند رفتار بسیار گسترده است و عوامل مختلفی می‌توانند در آن حیطه رفتاری خاص تأثیر بگذارند. بنابراین اندازه‌گیری کامل آن حیطه رفتاری بسیار دشوار است. باید تمهیدی اندیشید تا از آن طریق بتوان نمونه‌ای از حیطه رفتاری مورد نظر را تحت شرایط مطلوبی استخراج کرد. طرح سئوالات آزمون یک روش استاندارد برای حصول به این هدف و به دست آوردن نمونه‌ای از حیطه رفتار مورد نظر به شمار می‌رود. به عنوان مثال، می‌توان به سئوالات معلم در یک کلاس درس اشاره کرد. چون وی قادر به اندازه‌گیری تمام محتوای حیطه مورد نظر نیست اقدام به طرح سئوالات خاص می‌نماید، به این امید که بتواند نمونه این مجموعه کوچک سئوالات را به محتوای حیطه گسترده‌تری از سئوالات که احتمال پرسیدن آنها ممکن است وجود داشته باشد تعمیم بدهد.

روانسنجان در این نکته که باید برای اندازه‌گیری توانایی آزمودنیها ابزاری ساخت که به وسیله آن بتوان برخی الگوهای رفتاری را بررسی، پیش‌بینی و یا کنترل کرد اتفاق نظر دارند. ایبل^۱ (۱۹۹۱) می‌گوید اگر بخواهیم تلاش و کوشش یادگیری را در شاگردان

برانگیزیم، اگر می‌خواهیم مدارس سودبخش داشته باشیم و اگر می‌خواهیم به طور منصفانه با افراد بر اساس پیشرفتشان برخورد کنیم، به آزمونهای خوب بیشتری نیاز داریم. آزمونهایی که عملکرد آزمودنیها را اندازه می‌گیرند به انشایی و عینی تقسیم می‌شوند. در خصوص این تصمیم که از کدام یک از این آزمونها برای اندازه‌گیری عملکرد آزمودنیها استفاده کنیم بین روانسنجان توافق نظر وجود ندارد. برخی عقیده دارند که آزمونهای انشایی نشانه بهتری از پیشرفت واقعی آزمودنیها فراهم می‌آورند (استانلی^۲، ۱۹۶۴؛ امان و گلوک^۳، ۱۹۷۱). آنها اضافه می‌کنند به جای این که به دانش‌آموزان پاسخهای حاضر و آماده‌ای داده شود، آنها باید قادر باشند تا بر دانش و معلومات خود تسلط کافی پیدا کنند. این کار آنها را قادر خواهد ساخت تا حقایق و اصول را بهم ارتباط دهند و آنها را به صورت منطقی سازماندهی نمایند و آنگاه در مورد مقایسه و تشریح این افکار و عقاید به صورت نوشتاری به قضاوت بنشینند. افزون بر این، برخی معتقدند از پاسخهایی که آزمودنیها به سئوالات آزمونهای انشایی می‌دهند می‌توان به ماهیت و کیفیت فسرآیندهای ذهنی آنها پی برد

1- Ebel

2- Stanly

3- Ahmann and Glock

درس و تأثیری که تعداد غیبت‌های دانشجویان ممکن است بر نمره‌های پیشرفت آنها بگذارد فکر بسیاری از مدیران و مسئولین آموزشی مؤسسات آموزش عالی را به خود جلب کرده است. در این تحقیق سعی خواهد شد پیشرفت تحصیلی دانشجویان در آزمونهای صحیح-غلط و چند-گزینه‌ای با کنترل و بدون کنترل متغیر غیبت و به تفکیک جنسیت نیز مورد مقایسه قرار گیرد. بررسی شاخصهای روان سنجی سؤالات (ضریب دشواری، ضریب تمیز، ضریب پایایی، و جز اینها) این دو نوع آزمون اهداف دیگر این تحقیق را تشکیل می‌دهند.

بازنگری پیشینه تحقیق

سوابق آزمون کردن به ۲۳۰۰ سال قبل از میلاد مسیح برمی‌گردد. در آن زمان در چین از آزمونهای کتبی برای گزینش افراد لایق برای احراز تصدی امور مملکتی استفاده می‌کردند (اپیل، ۱۹۹۱). مطالعه آثار تربیتی بزرگان کشورمان حاکی از آن است که آنها برای ارزشیابی از اندوخته‌های علمی شاگردان

(کافمن،^۱ ۱۹۷۱؛ چیس،^۲ ۱۹۷۹؛ هیوز، کیلینگ و تاکر^۳، ۱۹۸۳).

روانسنجان جدید به دلیل محدودیتهایی که آزمونهای انشایی دارند (محدودیت در نمونه‌گیری کامل از محتوی درس، عدم ثبات و دقت اندازه‌گیری، ذهنیت در نمره‌گذاری، و جز اینها) استفاده از آزمونهای عینی بویژه آزمونهای صحیح-غلط و یا چند-گزینه‌ای را پیشنهاد کرده‌اند. (هوغان، کوئل‌مالز و کاپل^۴، ۱۹۸۰)، چو^۵ (۱۹۹۱)، اودل^۶ (۱۹۸۱). نتیجه تحقیقات اودل (۱۹۷۵) در رابطه با محدودیت آزمون انشایی نشان می‌دهد که طول پاسخی که یک فراگیر به یک سؤال انشایی می‌دهد با نمره‌ای که دریافت می‌کند ارتباط دارد. پاسخهای طولانی‌تر رتبه‌های بالاتری کسب می‌کنند.

تحقیق حاضر در پی آزمون این فرضیه است که اختلاف معنی‌داری بین آزمونهای صحیح-غلط و چند-گزینه‌ای به لحاظ اندازه‌گیری میزان پیشرفت آموذنی‌های مورد تحقیق وجود ندارد. افزون بر این، در این تحقیق مقایسه‌هایی نیز بین نمره آموذنیها در دو آزمون صحیح-غلط و چند-گزینه‌ای به تفکیک جنسیت و روش و نوع نمره‌گذاری سؤالات آزمون انجام خواهد شد. این روزها مسأله حضور یا عدم حضور دانشجو در کلاس

1- Coffman 2- Chase

3- Hughes, Keeling & Tucker

4- Hogan, Quellmalz & Capell

5- Chou 6- Odel

آزمون صحیح-غلط و چند-گزینه‌ای اشاره کرد^۲. با این که برای سؤالات آزمون چند-گزینه‌ای شاخصهای روانسنجی مطلوبی گزارش شده است، ولی بر سر این که کدام یک از این دو آزمون (صحیح-غلط و چند-گزینه‌ای) بهتر می‌تواند میزان پیشرفت دانش‌آموزان را بسنجد، بین روانسنان توافق کامل وجود ندارد. بررسی پیشینه تحقیق نشان می‌دهد که گروهی از روانسنان آزمونه‌های صحیح-غلط را در برخی جهات برتر از آزمونه‌های چند-گزینه‌ای می‌دانند و گاهی در برخی از ویژگیها آن را جلوتر از آزمونه‌های چند-گزینه‌ای قرار می‌دهند. به عنوان مثال، روانسنجانی چون لین^۳ (۱۹۸۳) بیرن بوم و تاتسوکا^۴ (۱۹۸۷)، ایبل (۱۹۷۵) و فریزی (۱۹۷۳) سؤالات آزمون صحیح-غلط را بسیار مناسب‌تر از سؤالات آزمون چند-گزینه‌ای می‌دانند. رواجی که استفاده از این دو نوع آزمون بین معلمان و استادان وجود دارد و علاقه‌ای که دانش‌آموزان و دانشجویان به این سؤالات نشان می‌دهند از یک سو، و

استفاده از آزمونه‌های شفاهی یا کتبی را توصیه می‌کردند (عیسی صدیق، ۱۳۳۶). شاید دلیل عمده استفاده از این آزمونها که به منظور آگاهی از میزان آموخته‌ها انجام می‌پذیرفت تعداد کم شاگردان بود که به معلم یا استاد فرصت می‌داد در وقت مناسب مفاهیم عمده درس را به طور کتبی و یا حتی به صورت شفاهی از شاگردان بپرسد. هم‌اکنون نیز استفاده از این نوع آزمون در بین معلمان مدارس و اساتید دانشگاه رایج است.

آزمونه‌های عینی بویژه آزمونه‌های صحیح-غلط و چند-گزینه‌ای سابقه چندان طولانی ندارند. این نوع آزمونها عمدتاً در پاسخ به محدودیتهایی که در سؤالات آزمونه‌های انشایی وجود دارد ساخته شده‌اند و در حال حاضر به صورت گسترده در گزینش افراد، در تصمیم‌گیریهای آموزشی، و در کلاس درس مورد استفاده قرار می‌گیرند. دانش‌آموزان مدارس و دانشجویان دانشگاه ترجیح می‌دهند که سؤالات امتحانی آنها از نوع عینی (چند-گزینه‌ای) باشد تا انشایی (مهرنز و Lehman^۱، ۱۹۸۴).

از میان آزمونه‌های عینی، آزمونه‌هایی که نمره‌گذاری سؤالات به صورت دو وجهی (صفر-یک) انجام می‌پذیرد، شهرت بیشتری دارند. از جمله این آزمونها می‌توان به سؤالات

1- Mehrens & Lehman

۲- در این تحقیق هر کجا به آزمون چند-گزینه‌ای اشاره می‌شود، مقصود آزمون چهار-گزینه‌ای است.

3- Linn

4- Biren Baum & Tatsuoka

عدم توافقی که بین روانسنگان بر سر استفاده از این آزمونها وجود دارد از سوی دیگر، به انجام تحقیق حاضر ضرورت می‌بخشد. در زیر پیشینه این دو آزمون مرور می‌شود.

آزمونهای چند-گزینه‌ای

آزمونهای چند-گزینه‌ای در مقایسه با سایر آزمونهای عینی به طور گسترده‌تری مورد استفاده قرار گرفته‌اند و در بین سازندگان آزمون از ارزش و اعتبار بالایی برخوردارند. سؤالات این نوع آزمون برای اندازه‌گیری سطوح مختلف یادگیری و نتایج مهم آموزشی از کارآیی مطلوبی برخوردارند. مهرنز و کامینسکی^۱ (۱۹۸۹) می‌گویند هر توانایی را که بتوان با سؤالاتی چون کوتاه- پاسخ، کامل- کردنی، جور- کردنی، صحیح- غلط، و یا حتی انشایی اندازه گرفت، به وسیله سؤالات آزمون چند-گزینه‌ای نیز قابل اندازه‌گیری است. علاوه بر این، نظر به تعدد گزینه در آنها، سؤالات این نسوع آزمون کمتر از سؤالات آزمون صحیح-غلط مظنون به خطای اندازه‌گیری ناشی از حدس و گمان می‌باشد.

سؤالات آزمون چند-گزینه‌ای به واسطه این که به وسیله افراد غیرماهر طوری ساخته می‌شوند که فقط به اندازه‌گیری جزئیات و سطوح پایین مفاهیم ذهنی می‌پردازند مورد

انتقاد قرار گرفته‌اند. منتقدان می‌گویند که تنها آزمونهای انشایی در شرایطی قرار دارند که قادرند مهارتهای سطوح عالی شناختی را اندازه‌گیری کنند. به علاوه آنها اشاره می‌کنند سؤالات آزمون چند-گزینه‌ای فقط به اندازه‌گیری مهارتهای ساده که از طریق حفظ کردن مطالب حاصل می‌شوند می‌پردازند. کیامنش (۱۳۷۱) از قول شیولسون و همکاران به دو معیار مهم آزمون یعنی توانایی درک مفاهیم و کسب مهارتهای لازم برای کاربست و انتقال این مفاهیم به یک موقعیت تازه و چگونگی تفسیر یافته‌ها، اشاره می‌کند. وی با توجه به ملاکهای فوق می‌گوید که "آزمونهای چند-گزینه‌ای به علت آن که اجزاء یک مسأله یا یک مفهوم را به صورت جداگانه و بدون ارتباط با سایر اجزاء مورد پرسش قرار می‌دهند دانش‌آموزان را با مسائل واقعی و آنچه مورد انتظار است یعنی روبرو شدن با مسائل آماده نمی‌کنند."

مولر و واسر^۲ (۱۹۷۷) می‌گویند ایرادهای منتقدان به ندرت به وسیله داده‌های پژوهشی بی‌طرف تأیید می‌شود. آنها اضافه می‌کنند با وجود این انتقاد که قدرت تمیز و پایایی نمره‌های آزمونهای چند-گزینه‌ای چندان

1- Mehrens & Kaminski

2- Mueller & Wasser

مطلوب نیست، ولی سازندگان ماهر و با تجربه آزمون نشان داده‌اند که نمره‌های این آزمون از دقت و ثبات و قدرت تفکیک مناسبی در اندازه‌گیری پیشرفت آزمودنیها برخوردار است. چنین استنباط می‌شود که شواهدی که در جهت عدم سود بخشی سؤالات آزمون چند-گزینه‌ای ارائه می‌شود بر پایه کلیاتی از موارد نادر است که سؤالات این آزمون ضعیف و برای اندازه‌گیری مفاهیم ساده ذهنی ساخته شده‌اند. بدون شک تمام فرمهای اندازه‌گیری از فواید و محدودیتهایی برخوردارند و سازندگان آزمونها نیز از این موضوع آگاهند. متخصصان آزمونهای روانی و تربیتی بر سر این مطلب که فرم واحدی از سؤالات آزمون برای تمام موقعیتهای آموزشی و روانی مناسبترین باشد توافق ندارند.

به طور کلی، در پاسخ به سؤالات انشایی آزمودنی بیشتر وقت خود را به فکر کردن و نوشتن صرف می‌کند، در حالی که در پاسخ به سؤالات آزمون چند-گزینه‌ای اکثراً به خواندن و فکر کردن مشغول می‌شود. برخی از روانسنجان بین این دو فرآیند تفاوت قائلند و معتقدند که شیوه فکر کردن و نوشتن فرآیند دشوارتری از خواندن و فکر کردن است (هافمن^۱ ۱۹۶۲، کرونباخ^۲ ۱۹۸۴). برخی دیگر از روانسنجان تفاوتی بین این دو فرآیند

نمی‌بینند (هوگان ۱۹۸۱، استیگینز^۳، روبل^۴ و کوئل مالز^۵ ۱۹۸۸). آنها معتقدند که در پاسخ به سؤالات چند-گزینه‌ای خوب آزمودنی لازم است از طریق تفکر منطقی، استدلال، حل مسأله، خلاقیت، اساس درستی برای انتخاب گزینه صحیح از بین سایر گزینه‌ها فراهم آورد. هوگان (۱۹۸۰) می‌گوید فراهم آوردن پاسخ به صورت نوشتاری الزاماً دلالت بر پیچیده‌تر بودن تکلیف نمی‌کند و یا نشانه بهتری از پیشرفت در مقایسه با انتخاب گزینه صحیح از بین سایر گزینه‌ها نیست. وی بر اساس نتایج تحقیقاتی که آزمونهای عینی و انشایی را با هم مقایسه کرده‌اند چنین نتیجه می‌گیرد که:

در بیشتر حالات، آزمونهای انشایی و عینی تا آنجا که به همبستگی درونیشان مربوط می‌شود پایایی یکسانی دارند. علاوه بر این، نمره‌های آزمونهای چند-گزینه‌ای تقریباً همیشه از نمره‌های آزمونهای انشایی پایاترند و نمره‌گذاری آسانتری دارند (هوگان، ص ۱۵۶)

با آن که نتایج بسیاری می‌توان در تأیید گفته‌های هوگان پیدا کرد، بسیاری از سازندگان

- | | |
|--------------|-------------|
| 1- Hoffman | 2- Cronbach |
| 3- Stiggins | 4- Rubel |
| 5- Quellmalz | |

سؤالات صحیح- غلط با دشواری مواجه‌اند، احتمالاً در بیان روشن و دقیق نقطه نظرات خود به صورت هر نوع سؤال دیگر نیز مشکل دارند.

مقایسه آزمونهای چند-گزینه‌ای و صحیح- غلط

تفاوت آشکار بین دو آزمون، در تعداد گزینه‌ها است. تفاوت دیگر به ویژگی تکلیفی که ارائه می‌دهند مربوط است. امکان دارد قضاوت درست یا غلط بودن یک قضیه از قضاوت در این که کدام گزینه بهترین گزینه پاسخ برای سؤال معین است دشوارتر باشد. ایبل و فریزی (۱۹۹۱) معتقدند سؤالات چند-گزینه‌ای حیطه مقایسات را برای آزمودنی‌ها در پیدا کردن پاسخ درست محدود می‌کند در حالی که به گفته آنها مجموعه مقایسات برای پیدا کردن پاسخ درست یک سؤال صحیح-غلط گسترده‌تر است. علی‌رغم تفاوتها و تشابهات زیادی که بین این دو نوع آزمون وجود دارد، آزمون چند-گزینه‌ای همانند آزمون صحیح-غلط، مفاهیم شناختی را اندازه می‌گیرد. سؤالات آزمون چند-گزینه‌ای که ماهرانه نوشته شده بود

آزمون این نتایج را نادیده می‌گیرند و بر این باور که در شرایط آنها دو نوع آزمون تواناییهای کاملاً متفاوتی را اندازه می‌گیرند اصرار می‌ورزند.

پیشینه تحقیق نشان می‌دهد با وجودی که سؤالات آزمون چند-گزینه‌ای از بونه آزمایش در تمام شرایط سرلند بیرون آمده است، ولی روانسنجانی هستند که توانایی سؤالات آزمون صحیح-غلط را نیز در اندازه‌گیری نتایج مهم آموزشی هم پای آزمونهای چند-گزینه‌ای می‌دانند. چنانچه ربط داشتن^۱ و کارآیی^۲ را دو ویژگی عمده سؤالات عینی بدانیم، ایبل (۱۹۹۱) می‌گوید آزمون صحیح-غلط این فایده را دارد که از لحاظ زمانی کارآمدتر است. وی می‌گوید تعداد پاسخهای قابل نمره‌گذاری مستقل برای هر هزار کلمه از آزمون یا در هر ساعت وقت اجرای آزمون به طور قابل ملاحظه‌ای در مقایسه با آزمون چند-گزینه‌ای بیشتر است. شواهد تحقیقاتی نشان می‌دهد که دانش‌آموزان می‌توانند در مقابل هر دو سؤال چند-گزینه‌ای به سه سؤال آزمون صحیح-غلط پاسخ دهند (فریزی؛ ۱۹۷۴، ۱۹۷۳). ایبل (۱۹۹۱) معتقد است که مهارت اصلی در نوشتن سؤالات صحیح-غلط با مهارت نوشتن هر نوع سؤال دیگر تفاوتی ندارد. ولی اضافه می‌کند آنهايي که در نوشتن

بیشتر باشد احتمال این که حدسهای او نزدیک به یقین باشد بیشتر خواهد بود. به بیان دیگر، فراگیران خوب می دانند که فکر کردن اساس مطمئن تری از حدس کور برای پیدا کردن پاسخ درست فراهم می آورد. در یک مطالعه، دانشجویان بهتر دانشگاه گزارش دادند که به طور متوسط در هر بیست سؤال احتمال دارد تنها به یک سؤال شانسی پاسخ دهند (هیلز و گلاذنی^۶، ۱۹۶۸)

ایراد دیگری که به سؤالات آزمون صحیح-غلط گرفته می شود این است که آزمودنی می تواند با استفاده از حدس کور با نمره نسبتاً خوب در آزمون قبول شود. این مورد از نظر قانون احتمال نیز به اثبات نرسیده است. اگر آزمونی فقط از یک سؤال غلط-صحیح تشکیل شده باشد و آزمودنی بخواهد با حدس کور از این سؤال نمره کامل (۱) بگیرد احتمال آن ۵۰ درصد است. این احتمال در آزمون دو سؤالی به ۲۵ درصد، در یک آزمون ۵ سؤالی به ۳ درصد و در یک آزمون ۱۰ سؤالی به ۰/۱ درصد کاهش می یابد. در یک آزمون ۱۰۰ سؤالی صحیح-غلط احتمال گرفتن نمره کامل (نمره صد) بسیار بسیار اندک

هنگامی که به سؤالات صحیح-غلط برگردانده شد و هر دو آزمون در مورد گروهی از آزمودنیها اجرا گردید، ضریب پایایی نمره های دو آزمون بهم نزدیک بود، (فریزی، ۱۹۸۸). در تحقیق دیگری محتوای هم وزن آزمونهای چند-گزینه ای و صحیح-غلط به طور شفاهی در مورد دانش آموزان عقب مانده ذهنی مدارس راهنمایی اجرا گردید، تفاوتی در دشواری یا ضریب پایایی آنها گزارش نگردید (ایروین^۱، هالپرن^۲ و لاندمن^۳، ۱۹۸۰).

با وجودی که ایراداتی از جانب روانسنجان، سازندگان سؤال و مجریان آزمون نسبت به سؤالات آزمون صحیح-غلط نسبت داده می شود، ایبل (۱۹۹۱) معتقد است که مطالعات تجربی دقیق کمی در تأیید این ایرادات صورت پذیرفته است. یکی از این ایرادها که برخی آن را بسیار جدی تلقی می کنند این است که حدس و شانس در پاسخ به سؤالات صحیح-غلط نقش مؤثری دارد و موجب خطای اندازه گیری در نمره آزمودنی می شود. فریزی (۱۹۸۸) می گوید باید بین حدس آگاهانه^۴ و حدس کور^۵ تفاوت قائل شد. وی معتقد است که حدس کور چیز قابل توجهی به نمره آزمون اضافه نمی کند ولی در مقابل حدس آگاهانه نشانه های با اعتباری از پیشرفت فراهم می کند. هر چه آگاهی فراگیر

- | | |
|----------------|--------------------|
| 1- Irvin | 2- Halpern |
| 3- Landman | 4- Informed Guess |
| 5- Blind Guess | 6- Hills & Gladney |

برای پاسخ آزمودنی فراهم آورد. وی آزمون چند-گزینه‌ای را بر آزمون صحیح-غلط ترجیح می‌دهد و در مورد استفاده از سؤالات صحیح-غلط چنین اظهار می‌دارد:

.... همان طور که قبلاً نشان داده شد، بهترین زمانی که می‌توان از آزمونهای صحیح-غلط استفاده کرد موقعی است که سؤالات چند-گزینه‌ای، مناسب و مقتضی تشخیص داده نمی‌شوند (ص. ۴۶).

روش تحقیق

جامعه

دانشجویان رشته‌های دبیری دانشگاه شهید چمران جامعه آماری این تحقیق را تشکیل می‌دهند.

نمونه

به منظور دستیابی به اهداف این تحقیق کل ۳۰۴ آزمودنی که در نیمسال دوم سال تحصیلی ۷۴-۷۵ برای واحد درس سنجش و اندازه‌گیری در علوم تربیتی ثبت‌نام کرده بودند، به عنوان نمونه این تحقیق در نظر گرفته شد و به طور تصادفی به دو گروه مساوی ۱۵۲ نفری تقسیم شدند. در هر گروه تعداد ۱۰۰ نفر پسر و ۵۲ نفر دختر قرار گرفت، روش تقسیم تصادفی و انتخاب دو گروه نمونه برای مقایسه عملکرد

است. چنانچه فرض کنیم یک آزمودنی می‌خواهد نمره نسبتاً خوب، مثلاً ۱۵ در یک آزمون صد-سؤالی با حدس کور دریافت کند احتمال آن یک در هزار است (ایبل و فریزی، ۱۹۹۱).

ایراد دیگری که به سؤالات صحیح-غلط گرفته می‌شود این است که نمره‌های این نوع سؤالات به دلیل اینکه پاسخ به آنها تحت تأثیر حدس کور قرار دارند، پایایی بالایی ندارند. مهرنز و لهمان (۱۹۸۴) ضریب پایایی ۰/۸۵ تا ۰/۹۵ را برای آزمونهای صحیح-غلط در کلاس درس نشان داده‌اند. آنها می‌گویند این مقادیر تقریباً به همان اندازه که می‌توان برای هر نوع دیگر از آزمون کلاسی انتظار داشت بالا است. این نتایج بر این تأکید می‌کنند که آزمونهای صحیح-غلط خوب از جانب حدس کور تهدید نمی‌شوند.

در مقابل، برخی متخصصان علوم تربیتی معتقدند که دو نوع آزمون، (چند-گزینه‌ای و صحیح-غلط) دو چیز متفاوت را می‌سنجد و اظهار می‌دارند که نمی‌توان در تمام حالات آزمون صحیح-غلط را جایگزین آزمون چند-گزینه‌ای کرد. گرون لاند (۱۹۶۸)، ضمن اشاره به محدودیتهای آزمون صحیح-غلط، معتقد است زمانی می‌توان از این آزمون استفاده کرد که سؤال فقط دو گزینه ممکن (درست یا غلط)

جدول ۱. تعداد و درصد حجم نمونه‌ها به تفکیک گروه و جنسیت

آزمودنی	گروه اول	درصد	گروه دوم	درصد	کل	درصد کل
پسر	۱۰۰	۳۲/۸۹	۱۰۰	۳۲/۸۹	۲۰۰	۶۵/۷۸
دختر	۵۲	۱۷/۱۱	۵۲	۱۷/۱۱	۱۰۴	۳۴/۲۲
کل	۱۵۲	۵۰	۱۵۲	۵۰	۳۰۴	۱۰۰

آنها در دو آزمون بدین ترتیب انجام گرفت که پس از استقرار آزمودنیها در جلسه امتحان، دو نوع آزمون به طور متناوب (یک در میان) بین آنها توزیع گردید. این روش علاوه بر آنکه آزمودنیها را به دو گروه مساوی به طور تصادفی تقسیم کرد و جنسیت افراد هر گروه را نیز تصادفی از همدیگر مجزا ساخت، این امکان را نیز برای پژوهشگر فراهم آورد تا از بروز خطای تصادفی از جمله احتمال تقلب آزمودنیها در جلسه امتحان (یعنی زیر چشمی نگاه کردن به و یا رونویسی از روی سؤالات آزمون بغل دستی) جلوگیری نماید. جدول ۱ حجم نمونه را به تفکیک گروه و جنسیت نشان می‌دهد.

ابزار تحقیق

از دو فرم آزمون یکی چند-گزینه‌ای و دیگری صحیح-غلط به عنوان ابزارهای این تحقیق استفاده شد. برای تهیه این ابزارها ابتدا

یک آزمون صحیح-غلط شامل ۷۱ سؤال و یک آزمون چند-گزینه‌ای از یک حیطة یکسان شامل ۶۰ سؤال بر اساس سرفصل درس سنجش و اندازه‌گیری و با استفاده از جدول مشخصات ساخته شد. این دو آزمون در نیمسال اول سال تحصیلی ۷۴-۷۵ در مورد دانشجویانی که برای واحد درس سنجش و اندازه‌گیری ثبت نام کرده بودند اجرا شد. ۳۰۴ نفر دانشجویان رشته‌های دبیری در جلسه امتحان پایان ترم به طور تصادفی به دو گروه تقسیم شدند و هر گروه به یک فرم آزمون پاسخ دادند. برای انتخاب بهترین سؤالات ضریب دشواری هر دو آزمون مورد تجزیه و تحلیل قرار گرفت. نتایج تحلیل سؤالات دو آزمون موجب گردید تا برخی از سؤالات که سطوح دشواری آنها به حد ملاک مطلوب نرسیده بود از مجموعه سؤالات حذف گردند. در نتیجه، ۷۱ سؤال آزمون صحیح-غلط به ۶۰ سؤال و تعداد ۶۰ سؤال آزمون چند-گزینه‌ای به

شد. سؤالات آزمون پیشرفت که با این روش ساخته می‌شوند، برای آنچه که قرار است اندازه بگیرند دارای اعتبار هستند. برای برآورد پایایی سؤالات دو آزمون از فرمول کورد-ریچاردسون (K-R20) استفاده گردید. برآورد پایایی یا ثبات درونی با استفاده از فرمول فوق برای آزمون چند-گزینه‌ای و آزمون صحیح-غلط به ترتیب ۰/۶۹ و ۰/۶۵ بود. ایبل (۱۹۹۱) حداقل پایایی قابل قبول آزمونهای پیشرفت کلاس درس را ۰/۶۵ برآورد می‌کند. برای آگاهی از اینکه این ابزارها برای اندازه‌گیری هدفهای مورد نظر تناسب دارند از روش اعتباریابی محتوا که روانسجان آن را برای اعتباریابی آزمونهای پیشرفت مناسب‌ترین روش می‌دانند استفاده گردید.

شیوه نمره‌گذاری آزمونها

در سه وضعیت ممکن آزمودنی به سؤالات آزمونهای عینی پاسخ می‌دهد. در وضعیت نخست، آزمودنی پاسخ درست را می‌داند و آن را انتخاب می‌کند. در حالت دوم، آزمودنی پاسخ آن دسته از سؤالاتی را که می‌داند انتخاب می‌کند و در پاسخی که نمی‌داند از حدس کور استفاده می‌کند. در

۴۰ سؤال تقلیل یافتند. از این دو فرم آزمون به عنوان ابزارهای تحقیق حاضر در نیمسال دوم همان سال تحصیلی استفاده به عمل آمد. چون آزمودنی‌ها به زمان کمتری برای پاسخ دادن به سؤالات آزمون صحیح-غلط تا چند گزینه‌ای نیاز دارند، بنابراین مدت اجرا هر دو آزمون مدت ۵۰ دقیقه تعیین گردید. نتایج تحقیقات نشان می‌دهند که آزمودنیها می‌توانند در وقت امتحان به ازاء هر دو سؤال چند گزینه‌ای به ۳ سؤال آزمون صحیح-غلط پاسخ دهند مهرز و کامینسکی^۱ (۱۹۸۹) افزون بر این، مهرز و لهما (۱۹۸۶) زمان لازم برای پاسخ به یک سؤال آزمون چند-گزینه‌ای را برابر ۷۵ ثانیه و برای پاسخ به یک سؤال آزمون صحیح-غلط را برابر ۵۰ ثانیه پیشنهاد می‌نمایند. گران لاند (۱۹۹۰)، ایبل (۱۹۹۱) با پیشنهاد مهرز و لهما (۱۹۸۴) موافقت دارند. تعداد سؤالات و زمان پاسخ دادن به سؤالات بر اساس یافته‌های این تحقیقات استوار است.

ابتدا هدفهای تدریس به صورت رفتاری و متناسب با اندازه‌گیری هر یک از قلمروهای شناختی از سرفصل درس استخراج گردید. آن‌گاه اهمیت نسبی هر یک از این قلمروها، (دانش ۱۵ درصد، فهم ۲۰ درصد، کاربرد ۳۵ درصد تحلیل ۳۰ درصد)، تعیین و تعداد سؤالاتی متناسب با هدفهای هر قلمرو ساخته

وضعیت سوم، آزمودنی سؤالاتی را که می‌داند پاسخ می‌دهد و بقیه را که نمی‌داند بدون پاسخ می‌گذارد. فرمول نمره‌گذاری برای سه وضعیت فوق متفاوت است. در حالت اول، نمره آزمودنی عبارت خواهد بود از مجموعه پاسخیهای درستی که به سؤالات آزمون می‌دهد. در حالت دوم از فرمول $S=R - \frac{W}{K-1}$ و در حالت سوم از فرمول $S=R + \frac{O}{K}$ برای نمره‌گذاری آزمون استفاده می‌شود. در این فرمول R تعداد پاسخیهای درست، W تعداد پاسخیهای غلط، O تعداد پاسخیهای بدون جواب (حذفها)، K تعداد گزینه‌ها و S نمره آزمودنی است.

اکثر روانسنجان وضعیت اول را برای نمره‌گذاری آزمونهای پیشرفت تحصیلی توصیه می‌کنند (گرونلند^۱، ۱۹۹۰، تروویلگر^۲، ۱۹۸۹). ولی برای استفاده از دو روش دیگر دلائل منطقی وجود دارد که در زیر به طور خلاصه به آنها اشاره خواهد شد. منطق استفاده از فرمول نمره‌گذاری وضعیت دوم را می‌توان در حذف عامل شانس که در پاسخ به سؤالات چندگزینه‌ای و صحیح-غلط وجود دارد ذکر کرد. به بیان دیگر، چون عامل شانس در پاسخ به این نوع سؤالات نقش مهمی ایفا می‌نماید، بنابراین نمره آزمودنی هیچ‌گاه بر روی پیوستار توانایی از نقطه صفر آغاز

نمی‌شود. برای اینکه نمره آزمودنی تصحیح گردد و در روی پیوستار به جای واقعی خود برگردانده شود، از فرمول دوم استفاده می‌شود. اما، مواقعی نیز وجود دارد که آزمودنی پاسخ درست برخی از سؤالات را نمی‌داند ولی بواسطه ترس از نمره منفی آنها را بدون جواب باقی می‌گذارد. در این وضعیت، بر تعداد پاسخیهای درست او معادل کسر $\frac{O}{K}$ اضافه می‌گردد. منطق استفاده از این روش این است چنانچه آزمودنی سؤالات را بدون پاسخ نمی‌گذاشت و از عامل حدس برای پاسخ دادن به آنها استفاده می‌کرد این احتمال که کسری معادل $\frac{O}{K}$ به نمره او اضافه شود وجود داشت. اگر چه استفاده از فرمول‌های نمره‌گذاری به نمره‌های متفاوتی منجر خواهند شد، ترتیب رتبه‌های آزمودنیها در توانایی مورد نظر که به وسیله آزمون اندازه‌گیری می‌شود یکسان باقی می‌مانند. همچنین اگر دو فرمول را برای مجموعه سؤالات یکسانی به کار بریم، نتایج کاملاً با هم همبستگی دارند (تراب و همبلتون^۳، ۱۹۷۲). در تحقیق حاضر، افزون بر راهنماییهای لازم، به آزمودنیها گفته شد که نمره آنها با احتساب نمره منفی گزارش خواهد شد. ولی، برای آزمون این فرضیه که آیا

1- Gronlund 2- Terwilliger
3- Truub & Hambelton

جدول ۲. میانگین، انحراف معیار، حداقل و حداکثر نمره‌های آزمودنیها به تفکیک جنسیت در آزمون

صحیح-غلط

نمره جنسیت	متغیر	میانگین	انحراف معیار	حداقل	حداکثر	تعداد
دختر	نمره ترم بدون تصحیح	۱۲/۲۰	۱/۸۸	۷/۸۸	۱۷/۱۸	۵۲
	نمره ترم با تصحیح اول	۵/۰۹	۲/۸۹	۰/۲۸	۱۴/۳۶	
	نمره ترم با تصحیح دوم	۱۲/۵۹	۱/۴۶	۱۰/۱۴	۱۷/۱۸	
	غیبت	۱/۹۱	۱/۵۵	۰	۶	
پسر	نمره ترم بدون تصحیح	۱۲/۸	۲/۲۰	۷/۶	۱۵/۲۱	۱۰۰
	نمره ترم با تصحیح اول	۶/۲۵	۲/۶۷	۰/۲۸	۱۱/۵۵	
	نمره ترم با تصحیح دوم	۱۳/۱۳	۱/۳۳	۱۰/۱۴	۱۵/۷۷	
	غیبت	۱/۳۶	۱/۲۳	۰	۶	
کل	نمره ترم بدون تصحیح	۱۲/۰۴	۱/۹۲	۷/۶۰	۱۷/۱۸	۱۵۲
	نمره ترم با تصحیح اول	۵/۵۰	۲/۸۵	۰/۲۸	۱۴/۳۶	
	نمره ترم با تصحیح دوم	۱۲/۷۸	۱/۴۴	۱۰/۱۴	۱۷/۱۸	

غلط داده شد. حیطه‌ای که سؤالات هر دو آزمون از آن استخراج گردید یکسان بود. افزون بر این، آزمودنیها به لحاظ جنسیت نیز به گروههای فرعی تقسیم شدند. جدول ۱ تعداد و درصد حجم نمونه‌ها را به تفکیک جنسیت نشان می‌دهد.

جدول ۲ داده‌های مربوط به نمره ترم بدون تصحیح، نمره ترم با تصحیح اول، نمره ترم با تصحیح دوم، و تعداد غیبتهای آزمودنیها را به تفکیک جنسیت و در کل در آزمون صحیح-غلط نشان می‌دهد.

از ۱۵۲ نفری که به سؤالات آزمون چهار

اختلاف معنی‌داری بین فرمولهای مختلف نمره‌گذاری وجود دارد یا خیر از هر سه شیوه نمره‌گذاری استفاده گردید. برای اطلاع خواننده، در این تحقیق به روش اول نمره‌گذاری اصطلاحاً "نمره ترم بدون تصحیح" به روش دوم نمره‌گذاری اصطلاحاً "نمره ترم با تصحیح اول"، و به روش سوم نمره‌گذاری "نمره ترم با تصحیح دوم" گفته می‌شود.

تحلیل داده‌ها

به یکی از گروه‌های ۱۵۲ نفری آزمون چند-گزینه‌ای و به گروه دیگر آزمون صحیح-

جدول ۳. میانگین، انحراف معیار، حداقل و حداکثر نمره‌های آزمودنی‌های دختر و پسر در آزمون چهار-گزینه‌ای

تعداد	حداکثر	حداقل	انحراف معیار	میانگین	متغیر	نمره جنسیت
۵۲	۱۸	۱/۵۰	۲/۹۰	۱۱/۴۷	نمره ترم بدون تصحیح	دختر
	۱۷/۳۴	۰/۱۷	۳/۴۷	۹/۰۱	نمره ترم با تصحیح اول	
	۱۸/۳۵	۰/۵	۲/۷۷	۱۱/۷۹	نمره ترم با تصحیح دوم	
	۵	۰	۱/۵۷	۱/۸۹	غیبت	
۱۰۰	۱۸/۵	۴/۵	۲/۶۱	۱۲/۵۰	نمره ترم بدون تصحیح	پسر
	۱۱/۰۰	۲/۳۳	۳/۸۰	۹/۹۷	نمره ترم با تصحیح اول	
	۱۸/۵	۶/۷۵	۲/۳۲	۱۲/۴۸	نمره ترم با تصحیح دوم	
	۴	۰	۱/۲۳	۱/۸۸	غیبت	
۱۵۲	۱۸/۵	۱/۵۰	۲/۸۰	۱۱/۶۸	نمره ترم بدون تصحیح	کل
	۱۸/۰۰	۰/۰۷	۳/۳۷	۹/۳۳	نمره ترم با تصحیح اول	
	۱۸/۵۰	۱/۵	۲/۴۴	۱۲/۰۲	نمره ترم با تصحیح دوم	
	۵	۰	۱/۴۶	۱/۸۹	غیبت	

نمره‌های پیشرفت آزمودنیها در پاسخ به سوالات آزمونهای صحیح-غلط و چند-گزینه‌ای اختلافی وجود ندارد، نشان داد که هر دو آزمون از لحاظ اندازه‌گیری میزان پیشرفت آزمودنی‌های مورد تحقیق یکسان عمل کرده‌اند. داده‌های فرضیه فوق با کنترل غیبت‌های آزمودنیها در طول ترم مورد تجزیه و تحلیل قرار گرفتند. جدول ۴ نتایج تحلیل داده‌ها را با کنترل تعداد غیبت‌ها نشان می‌دهد.

بررسی داده‌های جدول ۴ که بر اساس روش اول، نمره ترم بدون تصحیح، (مراجعه شود به شیوه نمره‌گذاری ص. ۴۱) و کنترل متغیر غیبت به دست آمد نشان می‌دهد که

گزینه‌ای پاسخ دادند ۱۰۰ نفر پسر و ۵۲ نفر دختر بودند. جدول ۳ داده‌های مربوط به آن دسته از دختران و پسرانی را که به سوالات چند-گزینه‌ای پاسخ دادند نشان می‌دهد. در این جدول میانگین، انحراف معیار، و دامنه نمره‌هایی که با سه روش متفاوت نمره‌گذاری و هم چنین آماره‌های مشابهی برای تعداد غیبت‌های این گروه از آزمودنیها در طول یک نیمسال تحصیلی همراه با دامنه تغییرات آنها آورده شده است.

آزمون فرضیه‌های تحقیق

آزمون این فرضیه که گفته شده بود بین

جدول ۴. تحلیل کواریانس نمره آزمودنیها در آزمونهای صحیح-غلط و چند-گزینه‌ای با کنترل غیبت

منبع تغییر	مجموع مربعات	درجه آزادی	میانگین مربعات	F	P
درون گروهی	۱۷۳۸/۹۳	۲۹۹	۵/۸۲		
جنسیت	۷/۸۰	۱	۷/۰۸	۱/۲۲	۰/۲۷۱
آزمون	۵/۷۰	۱	۵/۷۰	۰/۹۸	۰/۳۲۳
جنسیت × آزمون	۴/۴۹	۱	۴/۴۹	۰/۷۷	۰/۳۸۰

اختلاف معنی‌داری در سطح احتمال $P < ۰/۰۵$ بین نمره پیشرفت آزمودنیها به لحاظ نوع آزمونی که در آن شرکت کرده بودند (صحیح-غلط یا چند-گزینه‌ای) مشاهده نشد. این نتایج صحت فرضیه اصلی این تحقیق را مبنی بر عدم اختلاف میان نمره پیشرفت آزمودنیها در دو آزمون صحیح-غلط و چند-گزینه‌ای تأیید می‌کند. چنانچه این بار اختلاف بین آزمودنیها را به تفکیک جنسیت و نوع آزمونی که در آن شرکت داشتند با ملاحظه داده‌هایی که با استفاده از روش تصحیح اول (مراجعه شود به شیوه نمره‌گذاری، ص. ۴۱) به دست آمد همراه با کنترل متغیر غیبت تحلیل نماییم، بین پیشرفت آزمودنیها به لحاظ نوع آزمون اختلاف معنی‌داری در سطح احتمال $P < ۰/۰۵$ وجود دارد (جدول ۵ ملاحظه شود). همچنین، گروه پسران وضعیت بهتری در نمره‌های حاصل از

جدول ۵. تحلیل کواریانس نمره‌های پیشرفت آزمودنیها با تصحیح اول در آزمودنی‌های صحیح-غلط و چند-گزینه‌ای با کنترل غیبت

منبع تغییر	مجموع مجدورات	درجه آزادی	میانگین مربعات	F	P
درون گروهی	۲۸۸۸/۰۷	۳۰۰	۹/۶۳		
جنسیت	۷۶/۹۱	۱	۷۶/۹۱	۷/۹۹	۰/۰۰۵
آزمون	۱۰۰۱/۶۷	۱	۱۰۰۱/۶۷	۱۰۴/۰۵	۰/۰۰۰
جنسیت × آزمون	۰/۶۱	۱	۰/۶۱	۰/۰۶	۰/۸۰۱

جدول ۶. تحلیل کواریانس نمره‌های پیشرفت آزمودنیها (دختر و پسر) با تصحیح دوم در آزمودنیهای صحیح-غلط و چند-گزینه‌ای با کنترل غیبت

منبع تغییر	مجموع مجذورات	درجه آزادی	میانگین مربعات	F	P
درون گروهی	۶۳۷/۴۱	۳۰۱	۲/۱۲		
جنسیت	۵	۱	۵/۳۴	۲/۵۲	۰/۱۱۳
آزمون	۴/۴۸	۱	۴/۴۸	۲/۱۱	۰/۱۴۷
جنسیت × آزمون	۵/۰۶	۱	۵/۰۶	۲/۳۹	۰/۱۲۳

روش تصحیح دوم اختلاف معنی‌داری بین آزمودنی‌های دختر و پسر در آزمودنیهای صحیح-غلط و چند-گزینه‌ای وجود دارد یا نه به چه نتیجه‌ای می‌رسیم؟ نتایج نشان داد که این اختلاف به لحاظ جنسیت، نوع آزمون و تعامل بین جنسیت و نوع آزمون معنی‌دار نبود. جدول شماره ۶ نتایج این تحلیل را نشان می‌دهد.

علاوه بر مقایسه‌های فوق، در تحلیل دیگری ضریب همبستگی بین تعداد غیبتهای آزمودنیهای دختر با نمره پیشرفت آنها در آزمودنیهای صحیح-غلط و چند-گزینه‌ای محاسبه گردید و معنی‌دار بودن رابطه بین آنها مورد آزمون قرار گرفت. افزون بر این، محاسبات مشابهی برای تعیین میزان همبستگی بین تعداد غیبتهای آزمودنیهای پسر

دو آزمون صحیح-غلط و چند-گزینه‌ای نسبت به گروه دختران از خود نشان دادند. از این معنی‌دار بودن اختلاف پیشرفت آزمودنیها که با روش تصحیح اول به دست آمد، می‌توان چنین استنباط کرد که وقتی به پاسخهای غلط آزمودنیها نمره منفی تعلق می‌گیرد، هم بین میانگینهای دو نوع آزمون تفاوت وجود دارد و هم دو جنس (دختر و پسر) الگوهای متفاوتی در پاسخ به سؤالات آزمون اتخاذ می‌کنند. ولی، تعامل بین جنسیت و نوع آزمون (صحیح-غلط و چند-گزینه‌ای) نشان داد که بین نمره‌های پیشرفت آنها، اگر چه با نمره منفی محاسبه شد، اختلاف معنی‌داری مشاهده نگردید.

حال اگر تعداد غیبتهای دانشجویان را ثابت نگه داریم و این فرضیه را آزمون کنیم که آیا با

جدول ۷. ماتریس ضریب همبستگی بین تعداد غیبت‌های آزمودنی‌های پسر و دختر با نمره آنها در آزمون

صحیح-غلط و چهارگزینه‌ای

آزمون	روش نمره‌گذاری	نمره بدون تصحیح	نمره با تصحیح اول	نمره با تصحیح دوم	غیبت
دختر صحیح-غلط	نمره ترم بدون تصحیح	۱	۰/۹۱۴۰***	۰/۹۱۳۱***	-۰/۰۸۸۸
	نمره ترم با تصحیح اول		۱	۰/۹۸۸۰***	-۰/۱۳۱۹
	نمره ترم با تصحیح دوم غیبت			۱	-۰/۱۰۰۶ ۱
پسر صحیح-غلط	نمره ترم بدون تصحیح	۱	۰/۸۳۶۹***	۰/۸۳۶۰***	-۰/۱۵۸۰
	نمره ترم با تصحیح اول		۱	۰/۹۹۹۵***	-۰/۲۷۷۲
	نمره ترم با تصحیح دوم غیبت			۱	-۰/۲۷۷۷ ۱
دختر چهار-گزینه‌ای	نمره ترم بدون تصحیح	۱	۰/۹۵۶۵***	۰/۹۷۷۳***	-۰/۱۶۸۷
	نمره ترم با تصحیح اول		۱	۰/۹۷۶۰***	-۰/۲۳۷۷
	نمره ترم با تصحیح دوم غیبت			۱	-۰/۱۹۳۱ ۱
پسر چهار-گزینه‌ای	نمره ترم بدون تصحیح	۱	۰/۹۸۶۶***	۰/۹۸۷۲***	-۰/۲۷۹۵
	نمره ترم با تصحیح اول		۱	۰/۹۹۹۳***	-۰/۲۶۱۸
	نمره ترم با تصحیح دوم غیبت			۱	-۰/۲۵۲۲ ۱

** در سطح احتمال $P < 0/001$ معنی دار است.

پیشرفت تحصیلی دانشجویان و تعداد غیبت‌های آنان در طول یک نیمسال تحصیلی چند نکته را روشن می‌سازد. اول، نتایج نشان می‌دهد که در تمام موارد نمره‌گذاری، رابطه بین پیشرفت تحصیلی دانشجویان و تعداد غیبت‌های آنان منفی و معکوس است. به بیان دیگر، هر چه بر تعداد غیبت‌های دانشجویان

با نمره پیشرفت آنها انجام پذیرفت. لازم به توضیح است که نمره پیشرفت هر دو گروه آزمودنی در دو آزمون با استفاده از سه روش متفاوت نمره‌گذاری صورت پذیرفت. جدول ۷ ماتریس ضریب همبستگی تحلیل‌های فوق را نشان می‌دهد. بررسی نتایج جدول ۷ مبنی بر رابطه بین

افزوده می‌شود نمره پیشرفت آنها کاهش می‌یابد. دوم، در مورد دانشجویان دختر، با این که رابطه بین نمره پیشرفت و تعداد غیبت‌های آنان منفی است ولی این ارتباط چه در آزمون صحیح-غلط و چه در آزمون چند-گزینه‌ای در سطح احتمال $\alpha=0/05$ معنی‌دار نبود. شاید به توان این عدم معنی‌دار بودن را به حضور مرتب‌تر دانشجویان دختر در کلاس‌های درس و تعداد کم آنان نسبت داد. سوم، موقعی که رابطه بین پیشرفت تحصیلی دانشجویان پسر در دو آزمون صحیح-غلط و چند-گزینه‌ای و تعداد غیبت‌های آنان در طول یک نیمسال تحصیلی محاسبه گردید، نتایج نشان می‌دهد که ضرایب همبستگی معکوس و در سطح احتمال $\alpha=0/05$ معنی‌دار شدند.

بحث و نتیجه‌گیری

این تحقیق به دنبال آزمون این فرضیه اساسی بود که آیا اختلاف معنی‌داری بین نمره‌های پیشرفت دانشجویان در پاسخ به سؤالات آزمونهای چند-گزینه‌ای و صحیح-غلط وجود دارد یا نه. افزون بر این، بررسی تأثیر تعداد غیبت‌های دانشجویان بر عملکرد تحصیلی آنها در طول یک نیمسال تحصیلی

فرضیه عمده دیگر این تحقیق را تشکیل می‌داد که مورد آزمون قرار گرفت:

تحلیل آماری فرضیه اصلی این تحقیق نشان داد در صورتی که نمره آزمودنی‌های پسر و دختر زمانی که به سؤالات آزمونهای چند-گزینه‌ای و صحیح-غلط پاسخ می‌دهند و نمره پیشرفت آنها در پایان ترم با روش نمره‌گذاری معمولی (نمره ترم بدون تصحیح) مورد مقایسه قرار گیرند، اختلاف معنی‌داری بین عملکرد تحصیلی آنها در امتحان پایان ترم مشاهده نمی‌گردد. این نتیجه با یافته پژوهشهای روانسنجان به ویژه ایبل و فریزی (۱۹۹۱)، فریزی (۱۹۸۸)، و هوگان (۱۹۸۰) همخوانی دارد. از بررسی نتایج این تحقیق و آنچه از یافته‌های دیگران در این زمینه به دست آمده می‌توان چنین استنباط کرد که نتایج دو آزمون (چند-گزینه‌ای و صحیح-غلط) در ارزشیابی میزان پیشرفت تحصیلی دانشجویان با هم متناظرند. به سخن دیگر، چنانچه هدف از سؤالات آزمونهای پیشرفت نمونه‌گیری از محتوی حیطة مورد نظر باشد و بخواهیم نتایج آنها را به حیطة گسترده‌تری تعمیم بدهیم، این تحقیق نشان داد که بین سؤالات آزمونهای چند-گزینه‌ای و صحیح-غلط اختلاف

معنی داری مشاهده نمی‌شود. مطالبی که وسیله سؤالات آزمون چند-گزینه‌ای پرسیده می‌شوند امکان پرسیدن آنها نیز توسط سؤالات آزمون صحیح-غلط وجود دارد. متخصصان آزمون سازی روشهای متفاوتی برای نمره‌گذاری آزمونهای پیشرفت ارائه کرده‌اند. یافته‌های این تحقیق نشان داد چنانچه نمره‌گذاری سؤالات آزمونهای چند-گزینه‌ای و صحیح-غلط با روش تصحیح اول یعنی با احتساب نمره منفی (جریمه کردن آزمودنی‌ها به خاطر پاسخ‌های غلط) و با کنترل تعداد غیبت‌ها انجام پذیرد، اختلاف معنی داری در سطح احتمال $P < 0/05$ بین میانگینهای آزمونهای چند-گزینه‌ای و صحیح غلط از یکسو و بین عملکرد دانشجویان دختر و پسر از سوی دیگر وجود دارد. علاوه بر این، یافته‌های تحقیق با روش دیگری نیز مورد تجزیه و تحلیل قرار گرفت. در این روش که کسری معادل $\frac{0}{k}$ (به خاطر پاسخ‌هایی که بدون جواب گذاشته شده‌اند) به نمره آزمودنی‌ها اضافه شد، اختلاف معنی داری بین میانگینهای آزمونهای چند-گزینه‌ای و صحیح-غلط و بین عملکرد دانشجویان دختر و پسر مشاهده نگردید.

معمولاً به ندرت از این دو روش در نمره‌گذاری آزمونهای عینی به منظور اندازه‌گیری پیشرفت تحصیلی آزمودنیها در کلاس درس استفاده می‌شود. متخصصان آزمون سازی نیز پیشنهاد می‌کنند که حتی‌المقدور از این روشها برای نمره‌گذاری آزمونها استفاده نشود (مهرنر، لهما، ۱۹۸۴). زمانی که به پاسخهای غلط نمره منفی تعلق می‌گیرد نتایج تحلیل داده‌های این تحقیق نشان داد که بین پیشرفت آزمودنی‌های پسر و دختر اختلاف وجود دارد. آزمودنی‌های پسر در مقایسه با آزمودنی‌های دختر عملکرد بهتری در هر آزمون نشان دادند. شاید این اختلاف را بتوان به الگوهای متفاوتی که آزمودنی‌ها در رابطه با نمره منفی اتخاذ می‌کنند نسبت داد. دانشجویان پسر با اطمینان خاطر بیشتری به پیدا کردن گزینه صحیح می‌پردازند در حالی که دانشجویان دختر از بیم گرفتن نمره منفی بسیاری از سؤالات را بدون پاسخ می‌گذارند. چنانچه بتوان نتیجه این تحقیق را به آزمونهای چند-گزینه‌ای دیگر مثلاً (آزمون سراسری) که در آنها آزمودنیها برای پاسخ غلط جریمه می‌شوند تعمیم بدهیم دختران بیش از پسران متضرر می‌شوند. این نتیجه‌گیری به تحقیقات

بیشتری نیاز دارد.

معنی داری نرسیده است. بررسی ضرایب همبستگی بین غیبت‌ها و عملکرد دانشجویان تصویر روشنتری ترسیم می‌کند بدین معنا که اغلب این ضرایب همبستگی منفی و به سطح معنی داری $\alpha=0/05$ رسیده‌اند و حاکی از آن است که این دانشجویان از عدم حضور در کلاس متضرر می‌شوند. شاید عدم معنی دار بودن رابطه پیشرفت دانشجویان دختر و تعداد غیبت‌های آنان را در مقایسه با معنی دار بودن این ارتباط در مورد دانشجویان پسر بتوان به تعداد کم دختران در مقایسه با تعداد بیشتر پسران شرکت کننده در تحقیق نسبت داد.

بررسی رابطه پیشرفت تحصیلی دانشجویان و میزان حضور آنان در یک نیمسال تحصیلی در کلاس درس چشم‌انداز روشنی پیش روی مسئولان دانشگاهی قرار می‌دهد. به بیان دیگر، نتایج این تحقیق نشان می‌دهد که نگرانی مسئولان دانشگاه از عدم حضور مرتب دانشجویان را در کلاسهای درس باید جدی تلقی کرد. باید شرایطی فراهم کرد تا دانشجویان حضور مرتب‌تری در کلاسهای درس داشته باشند. توجه به رفتار ورودی شناختی و رفتار ورودی عاطفی دانشجویان توسط اساتید دانشگاه می‌توانند در ترغیب

در بررسی پیشینه موضوع مطالعاتی که تأثیر غیبت‌های دانشجویان را بر پیشرفت تحصیلی آنان مورد بررسی قرار می‌دهند بدست نیامد. چون حضور مرتب دانشجویان در کلاسهای درس مورد علاقه و عدم حضور آنان نگرانی مسئولان را به همراه دارد، بنابراین رابطه پیشرفت تحصیلی دانشجویان و تعداد غیبت‌های آنان در طول یک نیمسال تحصیلی مورد آزمون قرار گرفت.

تحلیل نتایج که در جدول‌های ماتریس ضرایب همبستگی آورده شده‌اند، به نکات جالب توجهی اشاره دارند. نتایج نشان می‌دهد که بین نمره ترم تمام دانشجویان و تعداد غیبت‌های آنان در طول ترم و در همه روشهای نمره‌گذاری رابطه معکوس وجود دارد. این رابطه به لحاظ شدت همبستگی گاهی معنی دار است و گاهی نیست. به عنوان مثال، تحلیل ضرایب همبستگی به لحاظ جنسیت دانشجویان نکات متفاوتی را روشن می‌سازد. بدلیل حضور مرتب دانشجویان دختر در کلاسهای درس، با اینکه رابطه بین نمره پیشرفت و تعداد غیبت‌های آنان منفی است، ولی ایسن رابطه در $\alpha=0/05$ به سطح

دانشجویان برای مشارکت فعال تر در تکلیف درسی مؤثر باشد. نتایج تحقیق حداکثر ۰/۵۲ همبستگی بین میزان علاقه و تمایل به ماندن در کلاس درس و کسب تجارب بیشتر را گزارش می کنند (بلوم، ۱۹۶۳؛ ص. ۹۲).

پیشنهادها

بر اساس نتایج این تحقیق، امکان ارایه پیشنهادهای زیر میسر می گردد.

۱. دو آزمون قادرند از حیطة شناختی سؤالات متفاوتی را نمونه برداری کنند. به بیان دیگر، سؤالات این دو آزمون وقتی که به روش معمولی و بدون اعمال تصحیح نمره گذاری شدند، به لحاظ اندازه گیری پیشرفت فراگیران یکسان عمل می کنند. و لذا می توان یکی را به جای دیگری مورد استفاده قرار داد.
۲. گاهی معلم به علت کثرت فراگیران قصد دارد از بروز برخی خطاهای تصادفی در جلسه امتحان جلوگیری کند. برای حصول به این مطلب پیشنهاد می شود که از یک محتوا دو فرم آزمون همتا یکی صحیح-غلط و دیگری چند-گزینه ای ساخته شود. استفاده از این دو فرم همتای یک آزمون به معلم فرصت می دهد تا در مورد تقلب فراگیران نگرانی به خود راه

ندهد.

۳. معلمانی که آزمونهای عینی را بهترین ابزار برای اندازه گیری پیشرفت آزمودنیهای خود انتخاب کرده اند حتی الامکان باید از دادن نمره منفی اجتناب ورزند. روان سنجان به طور کلی این روش ارزشیابی یعنی جریمه کردن آزمودنی را به خاطر انتخاب گزینه غلط به ویژه در کلاسهای درس پیشنهاد نمی کنند. توصیه روانسنجان با نتایج این تحقیق هم خوانی دارد. اهداف آزمونهای ورودی دانشگاهها با اهداف آزمونهایی که برای اندازه گیری پیشرفت آزمودنی در کلاس درس انجام می پذیرد متفاوتند.

۴. این که عملکرد آزمودنیهای پسر و دختر در پاسخ به سؤالات آزمونهای چند-گزینه ای و صحیح-غلط تفاوت معنی داری نشان نمی دهند هیچ گونه مطلبی درباره اهمیت سؤالات و درباره خود آزمونها بیان نمی کند. این مطلبی است که به شاخصهای روان سنجی سؤالات آزمونها و روش آموزش و تدریس معلم در کلاس درس بستگی دارد. بنابراین، پیشنهاد می شود برای این که آزمون انعکاسی از روش تدریس مطلوب باشد، باید توجه داشت که بازده یادگیری از یکسو به رفتارهای

ورودی شناختی و رفتارهای ورودی عاطفی آزمودنی و از سوی دیگر به تکالیف یادگیری و کیفیت آموزشی که معلم فراهم می‌سازد بستگی دارد.

به سخن دیگر، توافق معلم بر سر این که از کدام یک از انواع سؤالات آزمون استفاده کند کار چندان دشواری نیست. آنچه در فرآیند آزمون با اهمیت و درخور تعمق است برقراری تناسب بین سؤالات آزمون با سطوح تواناییهای شناختی آزمودنیها است. چنانچه هدف از اجرای آزمون صرفاً انجام وظیفه و برگزاری آزمون و گزارش نتایج در پایان ثلث یا ترم باشد تفاوت چندانانی ندارد که معلم از چه نوع سؤالاتی برای آزمون خود استفاده کند. آزمونی که نمره سؤالات آن فقط برای اطلاع آزمودنیها در کلاس خوانده می‌شود و دانش‌آموزان حتی موفق به دیدن اوراق امتحانی خود نمی‌شوند. نمره آزمونی که آزمودنیها نه قادرند عملکرد خود را با ملاک و نه با هنجار مقایسه کنند، و بالاخره با نمره آزمونی که شاخصهای روانسنجی سؤالات آن معلوم نیست، تفاوتی ندارد که از آزمونهاى انشایی و یا از آزمونهاى عینی به دست آمده باشد. مدت زمانی که معلم پیش از طرح

سؤالات و مدت زمانی که پس از اجرای آزمون صرف می‌کند که در یک پژوهش بیش از ۳۰ درصد کل فعالیت‌های آموزشی معلم تخمین زده می‌شود حایز کمال اهمیت است. مثلاً چنانچه قصد داریم سؤالات آزمون نمونه کاملی از مطالب تدریس باشد، یا چنانچه قصدمان از اجرای آزمون این است که نقطه قوت دانش‌آموز را اعتلا بخشیم و ضعف او را اصلاح کنیم، یا قصد از اجرای آزمون پی بردن به فرآیند ذهنی است که دانش‌آموز در پاسخ به یک سؤال یا حل یک مساله به کار می‌برد و این که چرا برخی از فراگیران از درک و فهم برخی مسائل عاجزند و بالاخره اگر می‌خواهیم دشواری سؤالات آزمون را کنترل کنیم و توزیع نمره‌های دانش‌آموزان را در سطح نرمال قرار دهیم، در چنین موقعیتی است که "از چه نوع سؤالی استفاده کنیم" اهمیت پیدا می‌کند.

۵. ماهیت برخی از دروس ایجاب می‌کند که آزمونهاى چند-گزینه‌ای و صحیح-غلط در آن قابل اجراء باشد و بنابراین بر اساس یافته‌های این تحقیق تفاوت معنی‌داری بین این دو دسته آزمون در اندازه‌گیری پیشرفت تحصیلی دانش‌جویان مشاهده نمی‌شود. بنابراین،

دانشجویان برای مشارکت فعال تر آنان در کلاسهای درس میزول نمایند.

۷. نتایج این تحقیق حداقل در مورد ۳۰۴ دانشجو (آزمونهای این تحقیق) صادق است. گرچه نتایج این تحقیق با نتایج تحقیق سایرین در این زمینه همخوانی دارد، ولی بدون شک تعمیم نتایج آن باید با احتیاط تلقی گردد. پیشنهاد می شود مطالعات مشابه دیگری در مورد سایر دروس با سایر آزمونها به وسیله اساتید دیگری انجام پذیرد، تا بتوان پاسخهای منطقی تری در تأیید یافته های این تحقیق پیدا کرد.

پیشنهاد می شود که نتایج این تحقیق در مورد دروس دیگر که امکان دارد کاربرد آزمونهای دیگری مناسبتر تشخیص داده شود، با احتیاط تلقی شود. ۶. باید عدم حضور مرتب دانشجویان در کلاس درس را جدی تلقی کرد. مسئولان برنامه ریزی دانشگاه از یکسو و استادان از سوی دیگر باید تمهیداتی بیندیشند تا دانشجویان با میل و رغبت بیشتری در کلاسهای درس حضور پیدا کنند. در این زمینه، پیشنهاد می شود اساتید توجه بیشتری به رفتارهای ورودی شناختی و عاطفی

منابع

فارسی

- بلوم، بنجامین (۱۳۶۳). ویژگیهای آدمی و یادگیری آموزشی. ترجمه علی اکبر سیف، تهران. مرکز نشر دانشگاهی.
- صدیق، عیسی (۱۳۳۶). تاریخ فرهنگ ایران. انتشارات دانشگاه تهران، شماره ۴۲۴.
- کیامنش، علیرضا (۱۳۷۱). آزمون چند-گزینه‌ای و تأثیر آن بر یادگیری دانش‌آموزان، محتوی آموزشی و روش تدریس. فصلنامه تعلیم و تربیت، شماره ۳۱.

خارجی

- Ahman, J. and Glock, M. (1971). Evaluating pupill growth: Principles of tests and measurements. (4th ed.) Boston: Allin and Bacon.
- Birnbaum, M. and Tatsuoaka, K.K. (1987). Open versus multiple choice response formats. *Journal of Applied Psychological Measurement*, 11, 385-95.
- Chase, C. (1979). Relative length of options and response set in multiple choice items. *Journal of Educational Measurement*. 16, 38-45.
- Chou, L.M. (1991). Ordering power of separate versus grouped true-false tests: interaction on type of test with knowledge level of examinees. *Applied Psychology Measurement*. 3, 529-36.
- Coffman, W.E. (1971). Essay examinations. in R.L. Thondike (ed.) Educational Measurement. Washington, D.C: American Council on Education.
- Cronbach, L.J. (1984). Further evidence on response sets and test design. *Educational and Psychological Measurement*. 10, 3-31.
- Crocker, L., and Algina, J. (1986). Introduction to classical and modern test theory. New York: Holt, Rinehart, and Winston Inc.
- Ebel, A.L. (1968). Blind guessing on objective achievement tests. *Journal of Educational Measurement*. 38,37-44.
- Ebel, L.R. and Frisbie, A.D. (1991). Essentials of Educational

- measurement. Englewood Cliffs, New Jersey, Prentice Hall.
- Ebel, L.R. (1975). Can teachers write good true false test items? *Journal of Educational Measurement*. 12, 31-36.
- Frisbie, D.A. (1973). Multiple choice vs. true-false: A comparison of reliabilities and construct validities. *Journal of Educational Measurement*. 10, 297-304.
- Frisbie, D.A. (1974). The effect of item format on reliability and validity. A study of multiple choice and true-false achievement test. *Educational and Psychological Measurement*. 34, 883-92.
- Frisbie, D.A. (1988). Reliability of scores from teacher-made tests. *Educational Measurement: Issues and Practice*. 7(1), 25-35.
- Gronlund, N.E. (1968). Constructing achievement tests. Nj, Prentice-Hall, Inc., Englewood Cliffs.
- Gronlund, N.E. and Linn, R.L. (1990). Measurement and evaluation in teaching. (6th ed.) New York: Macmillan Publishing Inc.
- Hills, J.R. and Gladney, M.B. (1968). Predicting grades from below chance test scores. *Journal of Educational Measurement*. 5, 45-53.
- Hogan, T.D. (1981). Relationship between free-response and choice type tests of achievement. ERIC Document Reproduction Service No. ED 224-811.
- Hogan, T.D., Quellmalz, F.S. and Capell, M.N. (1980). Effects of discourse response made on the measurement of writing competence. *Journal of Educational Measurement*. 19, 241-58.
- Hoffman, B. (1962). The tyranny of testing. New York: Collier Macmillan Inc.
- Huges, D.C., Keeling, B. and Tucker, B.F. (1983). Effects of achievement expectations and handwriting on scoring essays. *Journal of Educational Measurement*. 20, 65-70.
- Irvin, L.K. Halpern, A.S. and Landman, J.t. (1980). Assessment of retarding student achievement with standardized true-false and multiple-choice tests. *Journal of Educational Measurement*. 17, 51-58.
- Linn, R.I. (1983). Testing and instruction: links and distinction review. *Education*. 81, 7-23.
- Mehrens, A.W. and Lehman, J.I. (1984). Measurement and evaluation in

education and psychology. Holt, Rinehart and Winston Inc.

Mehrens, A.W. and Kaminski, L.M. (1989). Methods for improving standardized test scores. *Educational Measurement: Issues and Practice*. 8, 14-22.

Mueller, D.J. and Wasser, V. (1977). Implications of changing answers on objective test items. *Journal of Educational Measurement*. 14, 9-14.

Odel, C.W. (1981). Scales for rating pupils' answers to nine types of thought questions in English literature. IL: University of Illinois, Bureau of Educational Research.

Stanley, J.C. (1964). Measurement in today's school (4th ed.) Englewood Cliffs, New Jersey Prentice Hall Inc.

Stiggins, R.J. Rubel, E. and Quellmalz (1988). Improving thinking skills in classroom. Washington D.C: National of Educational Association.

Terwilliger, J.S. (1989). Classroom standard setting scoring practices. *Educational Measurement Issues and Practice*. 8(2), 15-19.

Traub, R.E. and Hambleton, R.k. (1972). The effect of scoring instructions and degree of speedness on the validity and reliability of multiple-choice tests. *Educational and Psychological Measurement*. 32, 737-758.