

تأثیر عامل حدس بر عملکرد تحصیلی دانش‌آموزان دوره متوسطه اول

The influence of predicting academic performance of high school students first

سیه بخشی پریخانی*^۱، صادق حامدی نسب^۲

Abstract:

The aim of this study was to investigate the influence of predicting academic performance of high school students first. The method used in this research is descriptive and correlational. The study population included all high school female students in the academic year ۱۳۹۵-۱۳۹۴ the city Meshgin is studying. Multistage cluster sampling of ۴۲۱ students were chosen girl. To collect data for this study from a test of mathematics achievement of first-choice school were used. In order to analyze the collected data, the descriptive statistics, mean and standard deviation and inferential statistics, methods of statistical correlation and t-test was used. For analysis software SPSS, BILOG was used. The results showed that the test instructions and Btb scoring methods according to the amount of C parameter estimation, and this effect is significant. Low levels of ability (۳ to ۵/۱) the performance of students in the first run (without guessing) compared to the second run (with guess), but is better in the middle and upper levels will be able to see the result. The validity of test scores (۲۵/۰ low) and lower levels (۲۵/۰ high) better estimation of ability in the first run (without guessing) in the second

چکیده:

هدف از انجام این پژوهش بررسی تأثیر عامل حدس بر عملکرد تحصیلی دانش‌آموزان دوره متوسطه اول است. روش تحقیق مورد استفاده در این پژوهش توصیفی از نوع همبستگی است. جامعه آماری، شامل کلیه دانش‌آموزان دختر دوره متوسطه اول شهرستان مشگین‌شهر که در سال تحصیلی ۱۳۹۴-۱۳۹۵ مشغول به تحصیل می‌باشد. با روش نمونه‌گیری خوشه‌ای چند مرحله‌ای تعداد ۴۲۱ نفر از دانش‌آموزان دختر به‌عنوان نمونه انتخاب شدند. برای گردآوری داده‌ها، برای انجام این پژوهش از یک آزمون چهارگزینه‌ای پیشرفت تحصیلی مربوط به درس ریاضیات اول دبیرستان استفاده شد. به‌منظور تجزیه و تحلیل داده‌های جمع‌آوری‌شده، از شاخص‌های آمار توصیفی؛ میانگین و انحراف معیار و از آمار استنباطی؛ روش‌های آماری همبستگی پیرسون و آزمون t وابسته استفاده شده است. برای تحلیل از نرم‌افزارهای SPSS، BILOG استفاده شد. یافته‌ها نشان داد که دستورالعمل آزمون و به‌تبع آن شیوه نمره‌گذاری متناسب با آن بر میزان برآورد پارامتر C مؤثر است و این تأثیر معنادار است. در سطوح پایین توانایی (۳ تا ۱/۵-) عملکرد دانش‌آموزان در اجرای اول (بدون حدس) نسبت به اجرای دوم (همراه با حدس)، بهتر است اما در سطوح میانی

^۱. کارشناسی ارشد تحقیقات آموزشی، دانشگاه خوارزمی (نویسنده مسئول: samabp^{۸۸}@gmail.com)
^۲. دانشجوی دکتری برنامه ریزی درسی، دانشگاه بیرجند

run (with guess) is.

Keywords: Operating guess, academic performance, female students, first year high school mathematics.

و بالایی توانایی این نتیجه عکس می‌شود. روایی نمره‌های آزمون در سطوح (۰/۲۵ پایین) کمتر و در سطوح (۰/۲۵ بالا) بهتر است و دقت برآورد توانایی در اجرای اول (بدون حدس) بیشتر از اجرای دوم (همراه با حدس) است. بررسی و مقایسه واریانس‌های وابسته در دو اجرا نشان می‌دهد که گروه قوی نسبت به گروه ضعیف، به‌طور یکسان عمل نکرده‌اند و میزان برآورد توانایی آن‌ها پراکندگی بیشتری را در هر اجرا نشان می‌دهد.

واژه‌های کلیدی: عامل حدس، عملکرد تحصیلی، دانش‌آموزان دختر، ریاضیات پایه اول دبیرستان.

مقدمه

یکی از اهداف روانسنجی ارزشیابی و تجزیه و تحلیل آزمون‌هایی است که به‌گونه‌ای به قصد اندازه‌گیری خصیصه‌در انسان‌ها تهیه می‌شوند (مرتاض هجری و همکاران، ۱۳۹۳). یکی از نگرانی‌هایی که همواره در مورد سؤالات چندگزینه‌ای وجود داشته است، پاسخ دادن حدسی آزمون‌شوندگان به سؤالات آزمون است. به‌طور کلی دو الگو در حدس زدن دانش‌آموزان مشهود است. در الگوی اول که حدس زدن ناآگاهانه یا تصادفی^۱ است، فراگیر نسبت به‌هیچ یک از گزینه‌ها هیچ نظری ندارد و یکی از گزینه‌ها را به صورت تصادفی انتخاب می‌کند. در الگوی دوم که حدس آگاهانه یا آموخته شده^۲ است، پیش‌فرض این است که دانش‌آموز با استفاده از استدلال منطقی یا با تکیه بر دانش نسبی^۳ خود سعی می‌کند با رد گزینه‌ها، جواب صحیح سؤال را حدس بزند (وایس و سوپ^۴، ۲۰۰۸). در باور عمومی جواب دادن حدسی سؤالات کار مذمومی است و نباید انجام شود؛ زیرا فرض بر آن است که از این کار موجب افزایش نمره دانش‌آموز به ناحق و کاهش روایی و پایایی آزمون می‌گردد (کریمی و همکاران، ۱۳۹۰). هر چند مطالعات تجربی در مورد این پیش‌فرض‌ها به نتایج یکسانی نرسیده‌اند، از دیرباز راه‌کارهایی مورد استفاده قرار گرفته‌اند که هدف عمده آن‌ها کاهش میزان حدس زدن دانش‌آموزان، تعدیل نمره آنان به میزان واقعی و بهبود ویژگی‌های سایکومتریک آزمون

^۱. Blind guess, Wild guess

^۲. Informed guessing, Educated guessing

^۳. Partial knowledge

^۴. Vyas & Supe

است (خین-چی - یه^۱، ۲۰۰۷). از این رو، این پژوهش درصدد بررسی راهبردی که در جهت کم کردن اثر عامل حدس دخیل‌اند، می‌پردازد.

تحقیق روی نخستین آزمون‌ها سبب شد که مفاهیم و روش‌های آماری در آزمون‌سازی و تفسیر نمره‌ی آزمون‌ها مورد استفاده قرار گیرد (کریمی خرامه، ۱۳۹۰). هر پژوهشگری خواه اجرای یک مطالعه اساساً تجربی مانند انجام تجربه آزمایشگاهی درباره یادگیری، انگیزش و تفکر را عهده‌دار باشد، خواه یک سلسله عملیات اساساً آماری را هدایت کند الزاماً به‌نوعی نمره‌های آزمونی دست می‌یابد (همبلتون^۲، ۲۰۱۳). روشن است که برای این‌گونه عملیات در زمان حاضر چندین روش متمایز وجود دارد. همیشه این روش‌ها، مفاهیم و اندیشه‌های مفیدی را که پایه‌ی فعالیت‌های مربوط به آزمون است و قیاس‌هایی را که در مورد آن‌ها توافق کلی وجود دارد به‌دست می‌دهد (کیمیایی، ۲۰۱۲). همه‌ی این روش‌ها با معانی روان‌شناختی و آماری نمره‌های آزمونی سروکار دارد و همه آن‌ها این سؤال کلی را مورد توجه قرار می‌دهند که: «نمره‌های آزمونی چه چیز را با چه درجه‌ای از دقت می‌سنجند؟» (شریفی، ۱۳۷۴). این روش‌ها هدف و مقصود بنیادی همه سنجش‌های آزمونی را که متضمن ارزشیابی افراد و کارکردهای افراد در پیوستارهای معروف است را دنبال می‌کنند و از فایده تفکر ریاضی محض به‌عنوان وسیله‌ای برای رسیدن به قیاس‌های قابل اعتماد درباره ویژگی‌های آزمون و نتایج آزمون‌ها سود می‌برند (لرد^۳، ۱۹۷۴).

تنوع و گستردگی استفاده از آزمون‌ها باعث شده است که انواع مختلفی از آن‌ها ساخته شود که هر کدام برای هدف خاصی به‌کار می‌روند در بین انواع آزمون‌ها، آزمون‌های چندگزینه‌ای از همه انواع دیگر مفیدتر و قابل انعطاف‌ترند و به کمک آن‌ها می‌توان تعداد زیادی از افراد را در سطح گسترده-ای از دانش، مورد آزمون قرار داد. سیف (۱۳۷۵) در کتاب روش‌های اندازه‌گیری و ارزشیابی خود ذکر می‌کند که یک سؤال چندگزینه‌ای است که آزمون‌شونده را با تکلیفی که هم مهم و هم کاملاً برای او قابل فهم است مواجهه می‌سازد، هر سؤال باید طوری نوشته شود که به‌وسیله کسانی که جواب سؤال را می‌دانند به درستی پاسخ داده شود و کسانی که تسلط کامل بر مطالب ندارند نتوانند به آن پاسخ درست بدهند (ص ۱۲۳). از آن‌جا که آزمون‌های پیشرفت‌تحصیلی برای اندازه‌گیری بازده فعالیت‌های آموزشی و تحصیلی به‌کار می‌رود، مهم است که فرایند نمره‌گذاری آن درست و دقیق انجام شود، یک مورد مهم و اساسی در آزمون‌های چندگزینه‌ای که باعث نگرانی شده است، تأثیر عامل حدس بر روی نمرات آزمودنی‌ها و مخدوش شدن دقت نمره‌گذاری و برآورد توانایی افراد است، بنابراین ضرورت بررسی راه‌هایی که به کم کردن اثر حدس بر روی نمره آزمودنی‌ها مخصوصاً آزمودنی‌های پیشرفت‌تحصیلی که تصمیماتی در پی دارد، احساس می‌شود (مؤمنی، ۱۳۷۹).

^۱. Chien-chi-yeh

^۲. Hambleton

^۳. Lord

افراد و نهادها زمانی از سنجش‌های آزمونی سود می‌برند که در رسیدن به هدف‌های خود از آن بهره گیرند و جامعه نیز موقعی از سنجش‌های آزمونی بهره می‌برد که پیشرفت هدف‌های فردی و سازمانی به هدف کلی آن کمک کند (فراری و زمیمرمان^۱، ۱۹۷۰). نتایج آزمون‌های مناسب که با دقت انتخاب شده باشد وقتی به‌گونه‌ای مناسب تفسیر شود می‌تواند اطلاعات مفیدی درباره پیشرفت در زمینه‌های مختلف در اختیار دست‌اندرکاران قرار دهد (شریفی، ۱۳۷۴). به‌عنوان مثال یک آزمون پیشرفت تحصیلی می‌تواند اطلاعات مفیدی درباره پیشرفت عمومی تحصیلی شاگردان در اختیار مجریان و مدیران آموزشی قرار دهد و از این طریق در زمینه‌ی آموزش صحیح به آن‌ها کمک کند. هم‌چنین نمره‌های آزمون‌ها می‌تواند معلمان، شاگردان و والدین را نیز در شناسایی نقاط قوت و ضعف شاگردان و به‌منظور کنار زدن آنان به کلاس‌های ویژه یا انجام اقدامات جبرانی و توجه خاص یاری دهد (مرتاض هجری و همکاران، ۱۳۹۳).

در بیشتر جوامع یکی از انواع رایج آزمون‌ها که به‌منظورهای گفته شده در بالا به‌کار می‌رود آزمون‌های چندگزینه‌ای است که علی‌رغم تمامی مزیت‌های گفته شده در مورد آن، یک مشکل اساسی در اجرای آن‌ها وجود دارد. آن مسئله اساسی امکان حدس زدن پاسخ صحیح، بدون داشتن دانش مربوط به آن است (کروس و فراری^۲، ۱۹۷۷)، در نتیجه نمرات دانش‌آموزانی که به این عمل دست می‌زنند به‌طور واقعی نشان‌دهنده میزان آگاهی دانش‌آموزان از موضوع مورد نظر نیست و این در تصمیماتی که با استفاده از نتایج این آزمون‌ها صورت می‌گیرد ایجاد اشکال می‌کند (دیاموند و ایونس^۳، ۱۹۷۳). تحقیقات بسیاری برای بررسی جنبه‌های مختلف دخیل در این مسئله انجام شده که تحقیق فعلی هم به جنبه‌ای از این مسئله می‌پردازد.

در یکی از اولین مطالعات در این زمینه راش و اسودارد^۴ (۱۹۲۵) مطرح کردند که استفاده از تصحیح در بیشتر موارد پایایی را کاهش می‌دهد، اما سال بعد راش و دگراف^۵ (۱۹۲۶) ارتباط بین دستورالعمل‌های آزمون و پایایی و نتایجی که اگر دستورالعمل « حدس نزنید^۶ » را به‌دست می‌آید را مطالعه کردند، به‌طور کلی پایایی نمره‌های تصحیح شده^۸ و تصحیح نشده^۹ همان بود، راش و دگراف به این نتیجه رسیدند که وقتی دستورالعمل حدس زدن داده شده بود نمره‌های تصحیح شده، پایایی بالاتری داشتند با این وجود بسیاری از نویسندگان مثل دیویس^{۱۰} (۱۹۶۴) مرغوبیت

^۱. Frary, & Zimmerman

^۲. Multiple- Choise

^۳. Cross&Frary

^۴. Diamond& Evans

^۵. Stoddard

^۶. Degraff

^۷. Do not guess

^۸. Corrected

^۹. un Corrected

^{۱۰}. Dives

نمره‌گذاری آزمون‌ها را در تعیین راهکار مورد نظر در دستورالعمل‌ها می‌دانند (مؤمنی، ۱۳۷۹؛ دووینیق^۱، ۲۰۰۳).

تراب، همبلتون و سینگ^۲ (۱۹۶۹) اثر اینکه به دانش‌آموزان گفته شود که: «نمراتشان براساس تعداد پاسخ‌های صحیح به‌علاوه $\frac{1}{8}$ تعداد سؤال‌های پاسخ نداده^۳ (سفید) است» مورد بررسی قرار دادند. این دستورالعمل یکی از چهار دستورالعملی بود که آن‌ها در تحقیقات‌شان به آزمودنی‌ها داده بودند، سه‌تای دیگر چیزی شبیه جملات زیر بود، «حدس نزنید»، «حدس بزنید^۴» و «در مورد حدس زدن چیزی گفته نشده بود»^۵ دو فرم آزمون دومینوی لغات بر روی ۶۶۷ دانش‌آموز پایه نهم اجرا شد. گروه‌های مستقل دانش‌آموزان در شرایط برابر آزمون‌ها را دریافت کردند و همه برگه‌ها با استفاده از نمره‌گذاری فرمولی نمره داده شد و همچنین تعداد پاسخ‌های صحیح، غلط و سفید لحاظ شدند. اگر چه محققین چندین برآورد پایایی فرم‌های همتا را ارائه کردند اما در مواردی که تکنیک نمره‌گذاری و دستورالعمل آزمون سازگار بودند پایایی‌ها هیچ اختلافی نشان نمی‌دادند. تحت دستورالعمل «حدس نزنید» با نمره‌گذاری فرمولی پایایی ۰/۸۹ بود، برای دستورالعمل «هیچ اشاره‌ای به حدس زدن نشده بود» با نمره‌گذاری «فقط صحیح‌ها» برآورد پایایی ۰/۹۵ بود، برای هر هشت ضریب پایایی (چهار مجموعه دستورالعمل و دو متد نمره‌گذاری) بالاترین برآورد پایایی برای دستورالعمل زیر به دست آمد: شما ممکن است پاسخ را با کلمات ناآشنا حدس بزنید، نمره‌ی شما در این آزمون براساس تعداد پاسخ‌های صحیح‌تان خواهد بود».

لرد^۶ (۱۹۷۴) قانع‌کننده‌ترین دلیل را برای استفاده از فرمول تصحیح به‌منظور افزایش روایی ارائه کرد. او گفت، اگر چه استفاده از فرمول‌های تصحیح فقط اندکی روایی را افزایش می‌دهد (۰/۰۳ تا ۰/۰۲)، قصور در استفاده از آن منجر به کاهش روایی می‌شود. شاید برابر باشد با چشم پوشیدن از نیمی از سؤالات و نیمی از زمان آزمون یا برابر باشد با پوشیدن از نیمی پاسخ‌های هر آزمون. با این حال اثر فرمول تصحیح در ضریب روایی کم است، لرد پیشنهاد کرد که استفاده از آن محدود شود به جائیکه، (a) تمایل به حدس زدن به‌طور قابل ملاحظه‌ای از شخصی به شخص دیگر تغییر می‌کند. (B) تعداد گزینه‌ها کمتر از ۵ تا است. (c) آزمون خیلی مشکل است.

لرد در مطالعه بعدی (۱۹۶۴) به دانش‌آموزان برای پاسخ‌گویی به هر سؤال اخطار داد، از آن‌جائیکه این دستورالعمل‌ها (دستورالعمل‌های بدون اخطار) ممکن است منجر به خطای اندازه‌گیری از طریق اجبار به حدس تصادفی شود، روایی را هم کاهش می‌دهد، با این حال در صورتی که رفتار حدس

^۱.Downing

^۲.Traub, Hambleton & Singh

^۳. Omitted

^۴. guess

^۵. No reference- to guessing

^۶.Doppelt

زدن غیرتصادفی باشد ممکن است منجر به افزایش روایی شود. هالادینا و داوینینق^۱ (۱۹۸۹) گفته لرد را مورد مطالعه قرار دادند، آن‌ها ضریب روایی پیش‌بینی را به‌دست آوردند که هیچ برتری را برای نمره‌های تصحیح شده و تصحیح نشده نشان نمی‌داد.

چانتاریلک^۲ (۲۰۱۱) اشاره کرد که دستورالعمل حدس نزنید یک برتری را برای آزمودنی‌های پرخاشگر در برابر غیر پرخاشگرها قائل می‌شود حتی اگر توانایی هر دو گروه برای پاسخ صحیح دادن به سؤالات پاسخ داده نشده، برابر باشد. کرونباخ^۳ (۱۹۴۶) نیز در تحقیقی که انجام داد به این نتیجه رسید که مقدار حدس آزمودنی‌ها در سؤال‌های چند گزینه‌ای با سطح دانش آن‌ها به خوبی فاکتورهای شخصیت‌شان مرتبط است.

براساس آنچه گفته شد پژوهش حاضر با هدف بررسی تأثیر عامل حدس بر عملکرد تحصیلی دانش‌آموزان دوره متوسطه اول انجام شد تا راهبردهایی در جهت کم کردن اثر عامل حدس ارائه دهد. از این رو پژوهش حاضر درصدد پاسخ‌گویی به سؤال‌های زیر می‌باشد.

۱- آیا دستورالعمل آزمون (حدس نزنید، حدس نزنید) تأثیری بر میزان برآورد پارامتر C دارد؟

۲- آیا استفاده از فرمول تصحیح نمره‌گذاری^۴ در مقایسه با روش نمره‌گذاری معمول پایایی نمره‌های آزمون را افزایش می‌دهد؟

۳- آیا استفاده از فرمول تصحیح نمره‌گذاری در مقایسه با روش نمره‌گذاری معمول روایی دارد؟

۴- آیا نمره‌گذاری فرمول تصحیح برآورد بهتری از توانایی آزمودنی می‌دهد؟

روش تحقیق

روش تحقیق در پژوهش حاضر از منظر هدف توصیفی-همبستگی است.

جامعه، نمونه و روش نمونه‌گیری

جامعه آماری در پژوهش حاضر عبارت است از کلیه دانش‌آموزان دوره متوسطه اول شهرستان مشگین‌شهر که در سال تحصیلی ۱۳۹۵-۱۳۹۴ مشغول به تحصیل می‌باشند. برای تعیین حجم نمونه، از روش نمونه‌گیری خوشه‌ای استفاده شد. برای این منظور ابتدا مدارس دخترانه متوسطه اول شهر مشگین‌شهر به چهار قطب تبدیل شدند بعد از هر قطب به صورت تصادفی شش دبیرستان برای نمونه‌گیری انتخاب شدند. تعداد حجم نمونه شامل ۴۲۱ دانش‌آموز دختر مقطع اول دبیرستان که در سال تحصیلی ۱۳۹۵-۱۳۹۴ مشغول تحصیل‌اند نمونه ما را تشکیل دادند.

ابزارهای پژوهش و روش اجرا

^۱. Haladyna & Downing

^۲. Chandratilake

^۳. Cronbach

^۴. $R = \frac{W}{A-1}$

ابزار اندازه‌گیری در این پژوهش یک آزمون ۳۰ سؤالی چهارگزینه‌ای پیشرفت تحصیلی درس ریاضیات است که طبق نظر کارشناسان از روایی صوری مناسب برخوردار است. به‌همراه دو دستورالعمل که به‌طور جداگانه به دانش‌آموزان داده می‌شود. نحوه اجرا و ارزشیابی آزمون به این صورت است که، آزمون به‌صورت گروهی اجرا می‌شود (دبیرستان به دبیرستان)، سعی می‌شود شرایط اجرای آزمون برای همه یکسان باشد. پس از آماده نمودن شرایط اجرای آزمون و دادن توضیحاتی مبتنی بر اینکه آزمون موردنظر به‌منظور جمع‌آوری داده‌هایی برای یک پژوهش اجرا می‌شود و اطمینان از این‌که دانش‌آموزان همکاری می‌کنند، آزمون اجرا شد. برای شروع مرحله دوم، اجرای آزمون در دو مرحله و بدون وقفه انجام شد. ابتدا دستورالعمل شماره ۱ به‌همراه دفترچه سؤالات و یک برگ پاسخ‌نامه، به هر دانش‌آموز داده شد از آن‌ها خواسته شد ابتدا دستورالعمل شماره یک را به‌دقت بخوانند و با توجه به آن به سؤالات پاسخ دهند، در این مرحله محدودیت زمانی لحاظ نشد و زمان لازم برای این‌که هر دانش‌آموز به سؤالات پاسخ دهد، داده شد (حداکثر ۳۰ دقیقه طول کشید).

در رابطه با شیوه نمره‌گذاری، به‌دلیل این‌که آزمون در دو مرحله (بدون وقفه) انجام شد و با توجه به دستورالعمل‌های داده شده هر دانش‌آموز دو نمره دارد - یکی برای مرحله اول که در دستورالعمل به دانش‌آموزان گفته شده بود سؤالات را با حدس پاسخ ندهند و پاسخ‌های غلط نمره منفی دارند و نمره‌گذاری در مرحله‌ای که از فرمول اصلاح نمره $(R - \frac{W}{A-1})$

استفاده شده بود. (R) ، تعداد پاسخ‌های صحیح؛ W ، تعداد پاسخ‌های غلط و A ، تعداد گزینه‌های هر سؤال است) و دیگری برای مرحله دوم که دستورالعمل دیگری به دانش‌آموزان داده شده بود مبنی بر این‌که با خودکار قرمز سؤالاتی را که پاسخ نداده‌اند به‌صورت حدسی پاسخ دهند و نمره هر آزمودنی در این مرحله با توجه به تعداد سؤالاتی که درست پاسخ داده‌اند است، برای هر پاسخ درست یک نمره در نظر گرفته شد.

به‌طور کلی برای تجزیه و تحلیل اطلاعاتی که منجر به بررسی سؤالاتی (مثل برآورد پارامتر C ، تابع آگاهی و...) که مباحث نظریه‌های جدید اندازه‌گیری را پوشش می‌دهند، می‌شود از نرم‌افزار **BILOG** و **spss** استفاده کرد و برای بررسی سؤالاتی که مباحث نظریه کلاسیک اندازه‌گیری را پوشش می‌دهند از نرم‌افزار **spss** استفاده شد.

یافته‌های پژوهش

هدف این پژوهش، بررسی رابطه تأثیر عامل حدس بر عملکرد تحصیلی دانش‌آموزان دوره متوسطه اول است. به منظور تجزیه و تحلیل اطلاعات جمع‌آوری شده برای پاسخ‌گویی به سؤالات تحقیق ابتدا اطلاعات پاسخ‌نامه‌ها در دو فایل جداگانه ریخته شد، فایل اول داده‌های مربوط به اجرای اول و فایل دوم داده‌های مربوط به اجرای دوم. به منظور گزارش یافته‌ها، ابتدا آماره‌های توصیفی هر دو اجرا به صورت مجزا برای گروه نمونه در جدول (۱) گزارش شد، لازم به توضیح است که نمرات اجرای اول (بدون حدس با نمره منفی) از طریق فرمول $(R - \frac{W}{A-1})$ به دست آمده است و نمرات اجرای دوم (با حدس، بدون نمره منفی) با توجه به تعداد پاسخ‌های صحیح بوده که به هر پاسخ یک نمره تعلق گرفته است. همان‌طور که مشاهده می‌شود در نمرات اجرای دوم، میانگین نمرات بالا رفته و در عین حال انحراف معیار و واریانس (پراکندگی) کمتری را شاهد هستیم، همان‌طور که مشاهده می‌شود کجی^۱ در اجرای اول و دوم هر دو منفی است و این نشان‌دهنده این است که منحنی متقارن نیست و دنباله‌ی طولانی منحنی به طرف چپ است (میانگین کوچک‌تر از میانه و نما)، اما با وجود منفی بودن کجی در هر دو اجرا در اجرای دوم کجی بیشتر از اجرای اول است و این نشان‌دهنده‌ی عدم تقارن بیشتر است. میانه، مد، کشیدگی^۲ ماکزیمم و مینیمم نمرات را در هر دو اجرا (جدول ۱) می‌توان دید و موارد مشابه را می‌توان با هم مقایسه کرد.

جدول ۱: آماره‌های توصیفی نمرات دانش‌آموزان گروه نمونه مربوط به اجرای اول و دوم

اجرای اول	اجرای دوم	
۴۲۱	۴۲۱	تعداد
۱۴/۲۶	۱۷/۳۳	میانگین
۰/۳۱	۰/۲۵	خطای استاندارد از میانگین
۱۴/۶۵	۱۸/۰۰	میانه
۷/۳۲	۴/۹۷	انحراف استاندارد
۴۰/۰۴	۲۴/۷۴	واریانس
-۰/۳۸	-۰/۵۵	کجی
۰/۱۱۸	۰/۱۱۸	خطای استاندارد کجی
۰/۰۵۳	۰/۱۱۵	کشیدگی
۰/۳۳۷	۰/۳۳۷	خطای استاندارد کشیدگی
۳۲/۰۰	۲۸/۰	دامنه
-۶/۰۰	۳/۰	نمره مینیمم
۲۸/۰۰	۳۰/۰	نمره ماکزیمم

^۱.Skweness

^۲.Kurtosis

نتایج تحلیل داده‌های مذکور در چارچوب ۴ سؤال تحقیق به شرح زیر ارائه می‌شود:
 سؤال ۱: آیا دستورالعمل آزمون (شیوه نمره‌گذاری متناسب با آن) تأثیری بر میزان برآورد پارامتر C دارد؟

همان‌طور که در این جدول (۳) هویداست در اجرای اول که دستورالعمل حدس نزنیید و شیوه نمره‌گذاری $R = \frac{W}{A-1}$ به کار رفته است، میانگین و انحراف استاندارد پارامتر C به این قرار است ($\mu = 0.17$ و $S = 0.057$) است و در اجرای دوم که دستورالعمل حدس نزنیید و شیوه نمره‌گذاری تعداد صحیح‌ها به کار رفته، میانگین و انحراف استاندارد به این قرار است ($\mu = 0.22$ ، $S = 0.050$)؛ با مراجعه به جدول شماره (۲) و مشاهده نتیجه t-test وابسته، بین دو اجرا معناداری تفاوت دو میانگین را می‌توان دید. این نتیجه نشان‌دهنده تفاوت معنادار نمرات افراد گروه نمونه زمانی که دستورالعمل آزمون و شیوه نمره‌گذاری پاسخ‌نامه تغییر می‌کند، است ($t = 0.01$ ، $p < 0.05$). بنابراین با توجه به جدول (۲) آزمون t بین میانگین‌های پارامتر C در دو اجرا تفاوت معناداری را نشان می‌دهد و به‌طور اجمال می‌توان گفت دستورالعمل آزمون و به‌تبع آن شیوه نمره‌گذاری متناسب با آن بر میزان برآورد پارامتر C مؤثر است و این تأثیر معنادار است.

جدول ۲: آزمون t وابسته بین میانگین دو اجرا

سطح معناداری	درجه آزادی	آزمون t	فاصله اطمینان ۹۵٪	خطای استاندارد میانگین	انحراف استاندارد	میانگین	اجرای اول و اجرای دوم
۰/۰۰۱	۴۲۰	۴۸/۵۰	۴/۷۶	۵/۱۶	۰/۱۰۲	۲/۱۰	۴/۹۶

جدول ۳: میانگین و انحراف استاندارد برآورد پارامترهای سؤال a,b,c (اجرای اول و اجرای دوم)

اجرای اول		اجرای دوم	
میانگین	انحراف استاندارد	میانگین	انحراف استاندارد
۰/۱۷۹	۰/۰۵۷	۰/۲۲۱	۰/۰۵۰
۱/۳۲۶	۰/۶۵۰	۱/۳۳۶	۰/۵۳۹
۰/۵۷۷	۱/۴۹۸	۰/۱۸۸	۱/۳۷۵

سؤال ۲: آیا استفاده از فرمول نمره‌گذاری $R = \frac{W}{A-1}$ (که عامل حدس را در نظر می‌گیرد) در مقایسه با روش نمره‌گذاری معمول (تعداد صحیح‌ها) پایایی نمره‌های آزمون را افزایش می‌دهد؟

به منظور بررسی این موضوع با تکیه بر همسانی درونی از روش ضریب آلفای کرونباخ و همچنین روش دونیمه کردن آزمون استفاده شد. ضریب آلفا برای اجرای اول ۰/۹۱ و برای اجرای دوم ۰/۸۰ به دست آمد (جدول ۴). همچنین ضریب همبستگی بین دو نیمه سؤالات زوج و فرد برای اجرای اول ۰/۸۹ و برای اجرای دوم ۰/۷۰ به دست آمد (جدول ۵)، که نتایج هر دو روش (ضریب آلفا و دونیمه کردن) منتهی به پاسخ مثبت به سؤال فوق می‌شوند.

جدول ۴: ضرایب آلفای کرونباخ برای اجرای اول و دوم

اجرای اول	اجرای دوم	
۴۲۱	۴۲۱	تعداد آزمودنی‌ها
۳۰	۳۰	تعداد سؤال
۰/۸۰	۰/۹۱	ضریب آلفای کرونباخ

جدول ۵: ضرایب پایایی به روش دونیمه کردن اجرای اول و دوم

اجرای اول	اجرای دوم	
۴۲۱	۴۲۱	آزمودنی‌ها
۳۰	۳۰	تعداد سؤال‌ها
۰/۸۱	۰/۵۴	همبستگی بین دو فرم
۰/۸۹	۰/۷۰	ضریب همبستگی قسمت اول شامل ۱۵ سؤال
۰/۸۹	۰/۷۰	ضریب همبستگی قسمت دوم شامل ۱۵ سؤال
۰/۸۹	۰/۷۶	ضریب آلفا برای قسمت اول
۰/۸۱	۰/۶۲	ضریب آلفا برای قسمت دوم

سؤال ۳: آیا استفاده از فرمول نمره‌گذاری $R = \frac{W}{A-1}$ در مقایسه با روش نمره‌گذاری معمول روایی دارد؟

برای بررسی اثر دستورالعمل و شیوه نمره‌گذاری در روایی آزمون، ضریب همبستگی نمرات دو اجرا و ضریب همبستگی هر اجرا با نمرات ریاضیات پایان ترم دانش‌آموزان گروه نمونه (به عنوان متغییر ملاک) محاسبه شده که با مراجعه به جدول (۶) می‌بینیم که بین نمرات دو اجرا همبستگی بالایی وجود دارد (۰/۹۵) و ضریب همبستگی نمرات اجرای اول

(بدون حدس) با متغیر ملاک (نمره ریاضیات) ۰/۳۴۶ و ضریب همبستگی نمرات اجرای دوم (با حدس) با متغیر ملاک (نمره ریاضیات) ۰/۳۴۲ است در مقایسه مشاهده می‌شود نمرات اجرای دوم با متغیر ملاک نسبت به اجرای اول با متغیر ملاک همبستگی کمتری دارند. (به- دلیل اینکه نمره ریاضیات پایان ترم دانش‌آموزان گروه نمونه که توسط دبیران مربوطه ارائه شده نمره اطلاعات کلی از دانش ریاضی دانش‌آموز است و آزمون مورد نظر اطلاعات جزئی- تری از این دانش را مدنظر داشته، همبستگی بالایی مشاهده نمی‌شود اما همین مقدار به- دلیل یاد شده برای منظور ما مناسب است).

جدول ۶: ضریب همبستگی محاسبه شده بین دو اجرا

اجرای اول	سطح معنی‌داری	ضریب همبستگی	تعداد
و اجرای دوم	۰/۰۰۱	۰/۹۵۹	۴۲۱

جدول ۷: جدول ضرایب همبستگی (دو اجرا باهم و هر اجرا با متغیر ملاک)

ریاضیات	اجرای دوم	اجرای اول	
۰/۳۴**	۰/۹۵**	۱/۰۰	ضریب همبستگی پیرسون برای اجرای اول
۰/۰۰۱	۰/۰۰۱		سطح معنی‌داری
۴۲۱	۴۲۱	۴۲۱	تعداد
۰/۳۴**	۱/۰۰	۰/۹۵**	ضریب همبستگی پیرسون برای اجرای دوم
۰/۰۰۱		۰/۰۰۱	سطح معنی‌داری
۴۲۱	۴۲۱	۴۲۱	تعداد
۱/۰۰	۰/۳۴۲**	۰/۳۴۶**	ضریب همبستگی پیرسون برای ریاضیات
	۰/۰۰۱	۰/۰۰۱	سطح معنی‌داری
۴۲۱	۴۲۱	۴۲۱	تعداد

** ضریب همبستگی در ۰/۰۱ معنی دار است.

سؤال ۴: آیا نمره‌گذاری فرمولی $R - \frac{W}{A-1}$ برآورد بهتری از توانایی آزمودنی می‌دهد؟ برای بررسی دقت برآورد توانایی در هر اجرا، خطای استاندارد سطوح مختلف توانایی مشخص شد (جدول ۸). نتیجه این‌که در سطوح پایین (۲/۵- به پایین) خطای استاندارد بیشتری را در اجرای دوم نسبت به اجرای اول مشاهده می‌کنیم این نشان می‌دهد که در

سطوح یاد شده دقت برآورد توانایی در اجرای اول بهتر از اجرای دوم است، در سطوح دیگر تنها نتیجه عکس می‌شود و میزان خطای استاندارد اجرای اول بیشتر از اجرای دوم است پس دقت برآورد توانایی در این سطوح در اجرای دوم بهتر از اول است. لازم به ذکر است که واریانس‌ها در چند سطح تنها در دو اجرا مورد مقایسه قرار گرفتند، (استفاده از روش مقایسه واریانس‌ها) که در جدول (۸) آمده است. نتیجه اینکه سطوح ۱، ۲، ۳ از تنها تفاوت واریانس‌ها معنی‌دار است این نشان می‌دهد که میزان پراکندگی در سطوح یاد شده در دو اجرا به‌طور معنی‌داری با هم تفاوت دارند در حالیکه در سطوح پایین تنها این تفاوت‌ها معنی‌دار است. یعنی دانش‌آموزان گروه قوی نسبت به گروه ضعیف در دو اجرا به‌طور یکسان عمل نکرده‌اند و میزان برآورد توانایی آن‌ها در دو اجرا پراکندگی بیشتری را نشان می‌دهد.

جدول ۸: مقایسه میزان خطای استاندارد واریانس‌های وابسته در سطوح مختلف

توانایی در دو اجرا

t	σ_2^2	σ_1^2	SE_2	SE_1	θ
					-۳/۵
					-۳
			۱/۰۴	۰/۹۶	-۲/۵
-۱۰/۵۵	۰/۵۹۸	۰/۴۳	۰/۶۸	۰/۶۰	-۲
			۰/۴۷	۰/۴۳	-۱/۵
۰/۳۳۹	۰/۱۳۸	۰/۱۴۱	۰/۳۶	۰/۳۸	-۱
			۰/۳۹	۰/۳۸	-۰/۵
۲/۶۱	۰/۱۶۸	۰/۱۸۲	۰/۴۲	۰/۴۳	۰
			۰/۴۷	۰/۴۸	۰/۵
۷/۲۳	۰/۲۱۵	۰/۲۶۸	۰/۵۱	۰/۵۳	۱
			۰/۴۸	۰/۵۸	۱/۵
۱۱/۰۸	۰/۳۳۶	۰/۴۷۰	۰/۵۶	۰/۷۰	۲
			۰/۶۸	۰/۷۵	۲/۵
۱۱/۰۰۳	۰/۶۷۳	۰/۹۴۰	۰/۸۲	۰/۹۵	۳
					۳/۵

بحث و نتیجه‌گیری

یافته‌ها بیانگر تأثیر عامل حدس زدن بر عملکرد تحصیلی دانش‌آموزان دوره متوسطه اول است. همانطور که ملاحظه شد دو اجرا روی یک گروه نمونه داشتیم که هر اجرا دستورالعمل خاص خود را داشت و شیوه نمره‌گذاری پاسخ‌نامه هم در دو اجرا متفاوت بود. با توجه به

یافته‌های سؤال (اول) که نشان داده شد که دستورالعمل آزمون و به تبع آن شیوه نمره‌گذاری متناسب با آن بر میزان برآورد پارامتر C مؤثر است و این تأثیر توجه معنادار است. در ادامه باید گفت که نتایج موجود در جدول (۳) آمده است دستورالعمل حدس بزنیید به میزان زیادی از رفتار حدس زدن آزمودنی‌ها را کاسته است. به صورتی که برای یک گروه در دو اجرا به تفاوت معنی‌داری از اختلاف میانگین برآورد پارامتر C برمی‌خوریم؛ بنابراین آزمون‌گران باید این موضوع را مدنظر داشته باشند. این نتایج با تحقیق مؤمنی (۱۳۷۹)، مبنی بر «تأثیر برآورد پارامتر C بر شیوه نمره‌گذاری متناسب بر آن» همسو می‌باشد.

در سؤال (دوم)، نتایج پایایی که به وسیله هر دو روش (ضریب آلفا و دونیمه کردن) انجام شد نشان داد که استفاده از فرمول نمره‌گذاری $R - \frac{W}{A-1}$ در مقایسه با روش نمره‌گذاری معمول پایایی نمره‌های آزمون را افزایش می‌دهد. این نتایج همسو با نتایجی است که کراس و فراری (۱۹۷۷) و مؤمنی (۱۳۷۹) به دست آوردند، فراری و کراس با شیوه اجرایی مشابه این پژوهش به ضریب پایایی ۰/۶۶ برای اجرای اول (بدون حدس) و ۰/۶۵ برای اجرای دوم (همراه با حدس) و مؤمنی ضریب پایایی ۰/۶۸ برای اجرای اول (بدون حدس) و ۰/۷۲ برای اجرای دوم (همراه با حدس) به دست آوردند.

در سؤال (سوم)، با محاسبه ضریب همبستگی بین نمرات هر اجرا مشخص شد که این تفاوت در حد چند هزارم است اما می‌توان گفت که افزایش روایی حتی در صدم و هزارم هم ارزشمند است؛ بنابراین می‌توان گفت اگر نیاز به افزایش روایی نداشته باشیم می‌توانیم از روش اول استفاده کرد و حتی تا $\frac{1}{n}$ (به وسیله روش‌های آماری مشخص می‌شود) کوتاه‌تر کنیم. نتایج بالا مشابه نتایج کراس و فراری^۱ (۱۹۷۷) و مؤمنی (۱۳۷۹) می‌باشد. آن‌ها طی پژوهشی که شیوه اجرای آن مشابه شیوه اجرای پژوهش حاضر است. به ترتیب به روایی ۰/۴۶، ۰/۶۲ و ۰/۵۹ برای اجرای اول (بدون حدس) و ۰/۴۶، ۰/۶۰ و ۰/۵۸ برای اجرای دوم (همراه با حدس) دست یافتند.

سؤال (چهارم) نیز نشان داد که برای بررسی دقت برآورد توانایی در هر اجرا، در سطوح پایین (۲/۵- به پایین) خطای استاندارد بیشتری را در اجرای دوم نسبت به اجرای اول مشاهده می‌کنیم این نشان می‌دهد که در سطوح یاد شده دقت برآورد توانایی در اجرای اول بهتر از اجرای دوم است، در سطوح دیگر نتایج عکس می‌شود و میزان خطای استاندارد اجرای اول بیشتر از اجرای دوم است پس دقت برآورد توانایی در این سطوح در اجرای دوم بهتر از اول است. لازم به ذکر است که واریانس‌ها نیز در چند سطح تتا در دو اجرا مورد مقایسه قرار گرفتند، نتیجه اینکه سطوح ۱، ۲، ۳ از تتا تفاوت واریانس‌ها معنی‌دار است این نشان می‌دهد که میزان پراکندگی در سطوح یاد شده در دو اجرا به طور معنی‌داری با هم

^۱.Cross&Frery

تفاوت دارند در حالیکه در سطوح پایین تا این تفاوت‌ها معنی‌دار است. یعنی دانش‌آموزان گروه قوی نسبت به گروه ضعیف در دو اجرا به‌طور یکسان عمل نکرده‌اند و میزان برآورد توانایی آن‌ها در دو اجرا پراکندگی بیشتری را نشان می‌دهد. این نتایج با تحقیق مؤمنی (۱۳۷۹) مبنی بر «تأثیر برآورد توانایی آزمودنی بر نمره‌گذاری فرمولی» همسو می‌باشد. آزمون‌های پیشرفت تحصیلی برای اندازه‌گیری بازده فعالیت‌های آموزشی - تحصیلی دانش‌آموزان به‌کار می‌رود از این رو، مهم است که فرایند نمره‌گذاری آن درست انجام شود. تأثیر عامل حدس بر نمرات آزمودنی‌ها و مخدوش شدن دقت نمره‌گذاری و برآورد توانایی افراد است که باعث نگرانی شده است. دستورالعمل آزمون و شیوه نمره‌گذاری متناسب با آن بر میزان برآورد پارامتر تأثیر دارد. از نمره‌گذاری فرمولی استفاده می‌شود. در سطوح پایین توانایی عملکرد دانش‌آموزان زمانی که از آن‌ها خواسته می‌شود حدس زنید نسبت به زمانی که از آن‌ها خواسته می‌شود، پاسخ‌ها را حدس بزنند بهتر است در سطوح میانی و بالای توانایی عکس می‌شود. این نتیجه شامل دقت برآورد توانایی هم می‌شود. از این رو، پیشنهاد می‌شود مشابه این پژوهش به‌همراه اجرای یک آزمون شخصیتی انجام شود و نتایج به تفکیک تیپ شخصیتی مورد تجزیه و تحلیل قرار گیرد (به این معنی که چه تیپ‌هایی بیشتر از حدس زدن برای پاسخ‌گویی به آزمون‌ها استفاده می‌کنند) و یا پیشنهاد می‌شود مشابه این پژوهش با در نظر گرفتن یک گروه کنترل انجام شود.

منابع

- آلن، ام، جی،، وین، دلبلیو، ام. (۱۳۸۳). مقدمه‌ای بر نظریه‌های اندازه‌گیری (روانسنجی) ترجمه علی دلاور. تهران: سمت، چاپ سوم.
- ثرندایک، رابرت ال. (۱۳۶۶). روانسنجی کاربردی. ترجمه حیدر علی هومن. تهران: دانشگاه تهران.
- سید مؤمنی، سیده طاهره. (۱۳۷۹). تأثیر عامل حدس (حدس زدن) بر نمره آزمون‌های چندگزینه‌ای پیشرفت تحصیلی. پایان نامه ارشد. دانشگاه علامه طباطبایی.
- سیف، علی‌اکبر. (۱۳۷۵). اندازه‌گیری، سنجش و ارزشیابی و آموزشی. تهران: دوران.
- سیف، علی‌اکبر. (۱۳۶۵). اندازه‌گیری پیشرفت تحصیلی، تهران: آگاه.
- شریفی، حسن پاشا. (۱۳۷۴). اصول روانسنجی و روان‌آزمایی. تهران: رشد.
- کریمی خرامی، بهمن. (۱۳۹۰). تأثیر تعداد گزینه‌های سؤال‌الدرویزگی‌ها بر روانسنجی آزمون توانایی برآورد شده در مدل‌های پیرسشپاسخ‌کو لاسیکان‌داز‌ه‌گیری. پایان نامه ارشد. دانشگاه علامه طباطبایی.

کریمی، بهنام، فلسفی نژاد، محمدرضد و درتاج، فریبرز (۱۳۹۰). تأثیر تعداد گزینه‌های سؤال در ویژگی‌های روان‌سنجی آزمون و توانایی برآورد شده در مدل‌های پرسش پاسخ و کلاسک اندازه‌گیری. فصلنامه اندازه‌گیری تربیتی شماره ۲ (۶).

مرتاض هجری، س، خباز مافی نژاد، م. جلیلی، م (۱۳۹۳). پاسخ حدسی به سوالات چندگزینه‌ای: چالش‌ها و راه کارها. مجله ایرانی آموزش در علوم پزشکی. ۱۴ (۷).
<http://ijme.mui.ac.ir>

References

- Betts LR, Elder TJ, Hartley J, Trueman M.* (۲۰۰۹). Does Correction For Guessing Reduce Students Performance On Multiple Choice Examinations? *Yes? No? Sometimes? [Abstract]. Assessment & Evaluation In Higher Education.* ۳۴(۱): ۱-۱۵.
- Chandratilake M, Davis M,* (۲۰۱۱). Ponnampereuma G. Assessment of medical knowledge: the pros and cons of using true/false multiple choice questions. *Natl Med J India.* ۲۴(۴): ۲۲۵-۸.
- Chien-chi-yeh* (۲۰۰۷). *The effect of guessing on assessing dimensionality in multiple choice test*, university of Pittsburgh.
- Cronbach, L. J.* (۱۹۴۶). Response sets and test validity. *Educational and psychological measurement*, ۶(۴), ۴۷۵-۴۹۴.
- Cross, L. H., & Frary, R. B.* (۱۹۷۷). An empirical test of lord's theoretical results regarding formula scoring of multiple-choice tests. *Journal of Educational Measurement*, ۱۴(۴), ۳۱۳-۳۲۱.
- Diamond, J., & Evans, W.* (۱۹۷۳). The Correction for Guessing ۱. *Review of Educational Research*, ۴۳(۲), ۱۸۱-۱۹۱.
- Downing SM.* (۲۰۰۳). Guessing on selected-response examinations. *Med Educ.* ۳۷(۸): ۶۷۰-۱.
- Frary, R. B., & Zimmerman, D. W.* (۱۹۷۰). Effect of variation in probability of guessing correctly on reliability of multiple-choice tests. *Educational and Psychological Measurement*, ۳۰(۳), ۵۹۵-۶۰۵.

- Haladyna, T. M., & Downing, S. M.* (1989). Validity of a taxonomy of multiple-choice item-writing rules. *Applied measurement in education*, 2(1), 51-78.
- Hambleton, R. K., & Swaminathan, H.* (2013). Item response theory: Principles and applications. *Springer Science & Business Media*.
- Kimiai K.* (2012). The Effects of Application of Correction-For-Guessing Formula on The Validity of The M-CtestsOf English Grammar Of EFL Students [Abstract]. *Frontiers of Language and Teaching*. 3: 228-238.
- Lord, F. M.* (1964). The effect of random guessing on test validity. *Educational and Psychological Measurement*, 24(4), 745-747.
- Lord, F. M.* (1974). Estimation of latent ability and item parameters when there are omitted responses. *Psychometrika*, 39(2), 247-264.
- Traub, R. E., Hambleton, R. K., & Singh, B.* (1969). Effects of promised reward and threatened penalty on performance of a multiple-choice vocabulary test. *Educational and Psychological Measurement*, 29(4), 847-861.
- Vyas R, Supe A.* (2008). Multiple choice questions: a literature review on the optimal number of options. *NatlMed J India*. 21(3): 130-3.