

## نقش نگرش خطرپذیری افراد، سطوح مختلف جریمه و دشواری آزمون بر همبستگی بین توانایی و نمره خام در آزمون‌های چندگزینه‌ای سراسری ورود به دانشگاه

حسین سلطانی ابقاء<sup>۱</sup>، بلال ایزانلو<sup>۲\*</sup>

تاریخ دریافت: ۹۷/۰۸/۱۵

تاریخ پذیرش: ۹۸/۰۲/۱۵

### چکیده

متداول‌ترین روش برای کاهش خطای ناشی از حدس، اعمال جریمه (نمره منفی) برای پاسخ‌های نادرست است. یکی از عوامل مؤثر بر میزان حدس، نگرش افراد نسبت به حدس زدن و خطر جریمه شدن است. هدف این پژوهش بررسی تأثیر سطوح مختلف جریمه و دشواری آزمون‌ها بر رابطه بین توانایی و نمره خام افراد و تعیین جریمه بهینه در سطوح مختلف توانایی با توجه به نگرش خطرپذیری آن‌ها است. برای بررسی تأثیر سطوح مختلف جریمه بر رابطه بین توانایی و نمره خام افراد، همبستگی توانایی و نمره خام افراد در سطوح مختلف جریمه برای سطوح مختلف توانایی بررسی و سطوحی از جریمه که موجب دستیابی به رابطه بیشینه بین توانایی و نمره خام افراد در هر سطح توانایی شد، به‌عنوان سطوح جریمه بهینه برای آن سطح توانایی انتخاب گردید. برای بررسی تأثیر دشواری آزمون‌ها بر رابطه بین توانایی و نمره خام افراد در سطوح جریمه بهینه، مقادیر بیشینه همبستگی در سطوح جریمه بهینه در هر سطح توانایی در سه آزمون (ریاضی، معارف و زبان انگلیسی) رشته ریاضی ۱۳۹۵ با توجه به میانگین ضریب دشواری آزمون‌ها مقایسه شد. نتایج نشان داد که هم عدم جریمه و هم اعمال جریمه بالا موجب کاهش همبستگی بین توانایی و نمره خام می‌شود. سطوح جریمه بهینه در افراد با توانایی بالا، بالاتر از افراد با توانایی کم بوده و به‌طور کلی جریمه بهینه برای کل افراد بالاتر از مقدار مرسوم آن (۰/۳۳) به دست آمد (برای کل افراد در آزمون ریاضی، معارف و زبان به ترتیب حداقل ۰/۴۰، ۰/۴۰ و ۰/۶۰). همچنین اگر میانگین ضریب دشواری سؤال‌های آزمون نزدیک به صفر باشد و سطوح جریمه بر اساس سطوح توانایی و با توجه به نگرش خطرپذیری بهینه باشد، رابطه بین توانایی و نمره افراد بیشینه می‌شود.

واژگان کلیدی: اصلاح حدس، نگرش خطرپذیری، جریمه بهینه.

۱. کارشناسی ارشد تحقیقات آموزشی، دانشگاه خوارزمی، تهران، ایران

۲. \*استادیار گروه روان‌شناسی، دانشگاه خوارزمی، تهران، ایران. izan.b@khu.ac.ir

## مقدمه

یکی از ابزارهایی که به‌طور گسترده برای ارزیابی سطوح دانش و یادگیری دانش‌آموزان استفاده می‌شود، آزمون‌های چندگزینه‌ای است. دلیلی که در بیشتر موارد آزمون‌های چندگزینه‌ای بر سایر اشکال آزمون‌ها ترجیح داده می‌شوند اجرای آسان، قابلیت استفاده در محتواهای مختلف، تضمین محتوای زیاد و نمره‌گذاری عینی آن‌ها است، به‌طوری‌که خطای اندازه‌گیری ناشی از نمره‌گذاری به حداقل ممکن می‌رسد (سنل، برکی پلون و آیتل<sup>۱</sup>، ۲۰۱۵). با توجه به اینکه هدف آزمونگر بیشینه کردن رابطه بین سطح دانش (توانایی) و نمره خام افراد در آزمون است، یک آزمون خوب (به‌خصوص از نوع چندگزینه‌ای) باید بیشینه‌کننده رابطه سطح دانش (توانایی) و نمره خام باشد (اسپینوزا و گاردیوزیل<sup>۲</sup>، ۲۰۱۰).

یکی از اشکالات مهم آزمون‌های چندگزینه‌ای احتمال یافتن پاسخ درست با استفاده از حدس است (سنل، برکی پلون و آیتل، ۲۰۱۵). در شرایط عدم جریمه برای پاسخ‌های نادرست، آزمون‌دهندگان انگیزه‌ای برای حدس زدن دارند. از این‌رو، ممکن است عامل حدس بر رابطه بین سطح دانش (توانایی) و نمره خام فرد تأثیرگذار باشد (اسپینوزا و گاردیوزیل، ۲۰۱۰). حدس زدن موجب افزایش خطای اندازه‌گیری و کاهش پایایی آزمون می‌شود (فراری، کراس و لوری<sup>۳</sup>، ۱۹۷۷). برای کاهش خطای اندازه‌گیری ناشی از حدس روش‌های مختلفی پیشنهاد شده است؛ مثلاً اجرای آزمون‌هایی با تعداد زیادی سؤال و یا افزایش تعداد گزینه‌های نادرست هر سؤال. در عمل به دلیل محدودیت زمانی رویکرد اول (استفاده از آزمونی با تعداد زیادی سؤال) امکان‌پذیر نیست. با توجه به اینکه نمی‌توان برای همه سؤال‌ها نیز تعداد زیادی گزینه نادرست طراحی کرد، رویکرد دوم نیز در عمل با دشواری همراه است. متداول‌ترین روش برای کاهش خطای اندازه‌گیری ناشی از حدس، استفاده از جریمه<sup>۴</sup> (نمره منفی) برای پاسخ‌های نادرست با استفاده از نمره‌دهی فرمولی<sup>۵</sup> جهت کاهش حدس است (اسپینوزا و گاردیوزیل، ۲۰۱۰).

- 
1. Şenel, Burcu Pehlivan & Alatl
  2. Espinosa and Gardeazabal
  3. Frary, Cross & Lowry
  4. penalize
  5. formula scoring

یکی از عوامل مهمی که بر حدس زدن دانش آموزان با توجه به میزان جریمه تأثیرگذار است، نگرش دانش آموزان نسبت به حدس و خطر جریمه شدن است. بعضی از افراد خیلی محتاط هستند، به طوری که اگر احتمال جریمه شدن کمی نیز وجود داشته باشد، فقط سؤال‌هایی را که مطمئن هستند را علامت می‌زنند. درحالی که در طرف دیگر افرادی هستند که بی‌پروا و صرف‌نظر از دستورالعمل جریمه شدن، همه سؤال‌ها را پاسخ می‌دهند (کوئرشی<sup>۱</sup>، ۱۹۷۴). این موضوع در نظریه تصمیم‌گیری<sup>۲</sup> (DT) در شرایط عدم اطمینان مورد بحث قرار گرفته است. بر طبق این نظریه افراد از نظر نگرش نسبت به حدس زدن و خطر جریمه شدن به سه دسته تقسیم می‌شوند. طبق این نظریه افرادی که سؤال با پاداش مورد انتظار مثبت را پاسخ نمی‌دهند، بیزار از خطر (خطرگریز)<sup>۳</sup> نامیده می‌شوند. افرادی که تمام سؤال‌های با پاداش مورد انتظار مثبت را پاسخ می‌دهند بی‌تفاوت نسبت به خطر<sup>۴</sup> و افرادی که سؤال‌های با پاداش مورد انتظار منفی را پاسخ می‌دهند خطرپذیر<sup>۵</sup> نامیده می‌شوند. تعیین جریمه زیاد باعث می‌شود تا افراد با اعتمادبه‌نفس کمتر و اضطراب زیاد سؤال‌هایی را که قادر به پاسخگویی آن هستند را پاسخ ندهند (شریفز و بومر<sup>۶</sup>، ۱۹۵۴). با توجه به نظریه تصمیم‌گیری در شرایط عدم اطمینان نیز تعیین یک جریمه خیلی بالا ممکن است سبب شود که حتی افراد با سطوح دانش بالا اما با نگرش خطرگریز سؤال‌هایی با پاداش مورد انتظار مثبت را پاسخ ندهند و نمره خام آزمون این افراد کاهش یابد که این موجب کاهش رابطه بین سطح دانش و نمره افراد خواهد شد و به نوبه خود بر پایایی آزمون تأثیر نامطلوبی خواهد داشت (اسپینوزا و گاردیوزبل، ۲۰۱۰). از این رو، به نظر می‌رسد که هم عدم جریمه و هم تعیین جریمه خیلی بالا برای حدس بر پایایی آزمون تأثیر نامطلوبی دارد. به همین دلیل بررسی تأثیر سطوح مختلف جریمه بر همبستگی بین توانایی و نمره خام افراد و تعیین اینکه مقدار جریمه بهینه چقدر باشد تا همبستگی بین توانایی و نمره خام افراد و پایایی آزمون افزایش یابد اهمیت زیادی پیدا می‌کند. همچنین میانگین دشواری آزمون‌ها نیز بر رابطه بین توانایی و نمره افراد تأثیرگذار است (اسپینوزا و گاردیوزبل، ۲۰۱۰).

- 
1. Quereshi
  2. Decision Theory
  3. risk aversion
  4. risk neutrality
  5. risk seeking
  6. Sherriffs and Boomer

در پیشینه ارزیابی آموزشی، این بحث قدیمی که آیا رویه نمره‌گذاری بهینه‌آزمون، بهتر است حدس‌ها و پاسخ‌های نادرست را به خاطر جلوگیری از حدس جریمه کند یا نه وجود دارد. با وجود استفاده گسترده از آزمون‌های چندگزینه‌ای، در مورد اینکه بهتر است پاسخ‌های نادرست جریمه شوند یا نه و یا اینکه میزان جریمه چقدر باشد، اتفاق‌نظری وجود ندارد. این موضوع که بهتر است از آزمودنی‌ها خواسته شود تا برای تمام سؤال‌ها در آزمون‌های چندگزینه‌ای و صحیح-غلط پاسخ‌ها را ثبت کنند و اگر لازم است حدس کورکورانه بزنند یا اینکه بهتر است با یک جریمه برای پاسخ‌های نادرست از حدس زدن دلسرد شوند، بحثی طولانی است (بورتون<sup>۱</sup>، ۲۰۰۲). بعضی از پژوهشگران استفاده از نمره-دهی فرمولی با استفاده از یک جریمه را پیشنهاد داده‌اند، اما برخی دیگر نمره‌گذاری تعداد پاسخ‌های درست را توصیه کرده‌اند. اِجینگِن<sup>۲</sup> (۱۹۶۵) و لیتل<sup>۳</sup> (۱۹۶۶) نمره‌گذاری تعداد پاسخ‌های درست را توصیه کرده‌اند، درحالی‌که کورتون<sup>۴</sup> (۱۹۶۶) و دیویس<sup>۵</sup> (۱۹۶۷) در طرفداری از نمره‌دهی فرمولی (استفاده از جریمه) بحث و استدلال کرده‌اند. بارهیلر، بودسکو و آتالی<sup>۶</sup> (۲۰۰۵) اظهار کرده‌اند که استفاده از نمره‌گذاری تعداد درست یک روش ممتاز نمره‌گذاری است درحالی‌که استفاده از نمره‌دهی فرمولی شکل نامعقولی از نقش آزمونگر است. کسانی که حدس زدن پاسخ درست را مشکل در نظر گرفته‌اند معتقدند که چاره آن دلسرد کردن پاسخ‌دهندگان از حدس زدن با استفاده از یک جریمه است و کسانی که مخالف استفاده از فرمول اصلاح حدس هستند، بر نرخ پایین موفقیت که ممکن است با شانس کسب شود اشاره می‌کنند و استدلال می‌کنند که احتمال کسب نمره بالا با شانس وجود ندارد (میرنز و لیمن<sup>۷</sup>، ۱۹۸۴). ایبل<sup>۸</sup> (۱۹۶۵) بیان می‌کند که مقایسه بین نمره‌های اصلاح‌شده و اصلاح‌نشده نشان می‌دهد که رابطه بین این نمره‌ها معمولاً خیلی بالا است. فراری<sup>۹</sup> (۱۹۸۸) نمره‌دهی فرمولی را برای بیشتر آزمون‌های کلاسی نامناسب ولی

- 
1. Burton
  2. Edgington
  3. Little
  4. Cureton
  5. Davis
  6. Bar-Hillel, Budescu, & Attali
  7. Mehrens, & Lehmann
  8. Ebel
  9. Frary

برای آزمون‌های سرعت و آزمون‌های دشوار با نمره قبولی پایین مطلوب می‌داند. برخی از پژوهشگران نیز گفته‌اند که در آزمون‌هایی که نمره‌ای به‌عنوان نقطه شکست-قبولی وجود ندارد و یا افراد برای پاسخ‌ها جریمه نمی‌شوند افراد با دانش کمتر تمایل دارند تا بیشتر سؤال‌ها را با حدس زدن پاسخ دهند. پس برای کاهش حدس یک جریمه لازم است (اسپینوزا و گاردیوزبل، ۲۰۱۰).

در مورد اثر حدس و اصلاح حدس بر روایی و پایایی<sup>۱</sup> آزمون نیز پژوهش‌های متفاوتی انجام شده است. کورتون (۱۹۶۶) بیان می‌کند زمانی که از فرمول اصلاح حدس استفاده می‌شود، پایایی آزمون افزایش می‌یابد. همچنین کورتون (۱۹۶۶) بیان می‌کند که نمره‌دهی فرمولی روایی آزمون را افزایش می‌دهد. دایموند و اوانز<sup>۲</sup> (۱۹۷۳) گزارش می‌کنند که مطالعات زیادی با اندازه‌های روایی افزایش یافته وجود دارد که نمره‌دهی فرمولی در آن به کار رفته است. لرد<sup>۳</sup> (۱۹۷۵) استدلال می‌کند که نمره‌دهی فرمولی همیشه پایایی را بهبود خواهد داد. این مزیت نمره‌دهی فرمولی در مطالعات اخیر فراوانی بحث شده و از نظر تجربی مورد حمایت قرار گرفته است. افزایش تعداد سؤال‌هایی که در آزمون‌های چندگزینه‌ای وجود دارد به‌طور بالقوه منجر به پایایی بیشتر می‌شود. آزمون‌هایی با تعداد پرسش‌های کم در معرض مسئله پایایی مربوط به سهمی از نمره نهایی که به حدس تصادفی نسبت داده می‌شود قرار دارند. پایایی آزمون‌های چندگزینه‌ای هم به تعداد کل سؤال‌ها و هم به تعداد گزینه‌های پاسخ برای هر سؤال وابسته است. در حقیقت پرسیدن سؤال‌های متفاوت از یک مفهوم یا سؤالات مشابه می‌تواند به‌طور ساختگی اندازه‌گیری پایایی آزمون را بالا ببرد (کمپل<sup>۴</sup>، ۲۰۱۵). به‌طور کلی به نظر می‌رسد که پایایی آزمون باید با روش نمره‌دهی فرمولی، بالاتر از نمره‌دهی تعداد درست-مخصوصاً در آزمون‌های کوتاه‌تر- باشد. نمره‌دهی فرمولی (با فراهم کردن اطلاعات پایاتر از دانش نادرست) به معلمان و (با ایجاد شاخص‌های معنادارتر تمیز و سهولت سؤال) به آزمونگرها سود می‌رساند (بورتون، ۲۰۰۲). کمپل (۲۰۱۵) نیز اشاره می‌کند که نمره‌دهی فرمولی به افزایش پایایی آزمون منجر می‌شود. وی همچنین نشان داد که نمره‌دهی فرمولی روایی<sup>۵</sup> را بهبود

1. Reliability And Validity
2. Diamond, Evans
3. Lord
4. Campbell
5. validity

می‌دهد. تأثیر مهم دیگر نمره‌دهی فرمولی این است که برای دانش نادرست مطمئناً نمره‌ها کم می‌شود که پایایی آزمون را افزایش می‌دهد. کوپین<sup>۱</sup> (۱۹۸۸) اشاره کرده است که اصلاح حدس زدن سه مسئله را در بر دارد: ۱) حدس زدن، عاملی تصادفی در نمره‌های آزمون را ارائه می‌کند که به شدت پایایی و روایی را کاهش می‌دهد، ۲) حدس‌های مورد انتظار درست برآورد توانایی‌های دانشجویان را بالا می‌برد و ۳) افزایش حدس زدن برای دانشجویانی که متناوباً حدس می‌زنند نسبت به دانشجویان دارای توانایی برابر که حدس نمی‌زنند مزیت ناعادلانه‌ای دارد. اعمال اصلاح برای حدس زدن به نفع دانشجویانی که متناوباً حدس می‌زنند را کاهش می‌دهد. دایموند و اوانز (۱۹۷۳) درباره نیاز به آموزش‌های خاص به دانشجویان درباره حدس زدن گزارش داده‌اند که اجازه می‌دهد آزمون‌های دارای تصحیح برای حدس زدن پایایی خود را حفظ کنند. دانشجویان باید بدانند که اصلاح حدس زدن به کار می‌رود و باید اثر حدس زدن را بدون دانستن یا حتی با دانش نسبی و نیز مزایای بالقوه دانش نسبی در نظر بگیرند.

با توجه به مطالب ذکر شده در فوق یک آزمون خوب چند گزینه‌ای موجب دستیابی به رابطه بیشینه بین توانایی و نمره خام افراد می‌شود. حدس زدن عاملی است که بر رابطه بین توانایی و نمره خام افراد تأثیر می‌گذارد. چون اعمال نمره منفی برای حدس (پاسخ‌های نادرست) باعث افزایش پایایی آزمون‌ها می‌گردد وجود جریمه ضروری به نظر می‌رسد؛ اما مسئله مهم این است که میزان جریمه چقدر باشد تا رابطه بیشینه بین توانایی آزمودنی‌ها و نمره خام کسب‌شده آن‌ها برقرار شود. از سوی دیگر با توجه به نظریه تصمیم‌گیری در شرایط عدم اطمینان، میزان جریمه بر حدس زدن افراد با توجه به نگرش خطرپذیری آن‌ها و در نتیجه نمره خام آن‌ها تأثیرگذار است. حال با توجه به نظریه تصمیم‌گیری در شرایط عدم اطمینان و نگرش خطرپذیری افراد، تعیین جریمه بهینه به طوری که رابطه بیشینه بین توانایی و نمره خام افراد به دست آید اهمیت زیادی پیدا می‌کند. همچنین مسئله دیگر بررسی تأثیر دشواری آزمون‌ها بر همبستگی بیشینه بین توانایی و نمره خام افراد است. هدف این پژوهش بررسی این موارد در آزمون‌های سراسری ورود به دانشگاه است.

## روش پژوهش

در این پژوهش داده‌های حاصل از پاسخنامه‌های داوطلبان رشته ریاضی-فیزیک در آزمون سراسری سال ۱۳۹۵ در دو آزمون عمومی معارف اسلامی و زبان انگلیسی هر کدام با ۲۵ سؤال و آزمون اختصاصی ریاضی با ۵۵ سؤال مورد تحلیل قرار گرفت. داده‌های این پژوهش قبلاً توسط سازمان سنجش آموزش کشور جمع‌آوری شده است و جزو داده‌های ثانویه محسوب می‌شوند. برای انجام تحلیل ابتدا نگرش خطرپذیری افراد براساس نظریه تصمیم‌گیری در شرایط عدم اطمینان به دست آمد. این کار براساس سه نوع نمره صورت گرفت که در ادامه توصیف شده‌اند، در این مقاله از نمره خطرپذیری که با روش دوم به دست آمده، استفاده شده است که دلیل آن در ادامه توضیح داده شده است.

روش اول: نمره خام خطرپذیری<sup>۱</sup> (RAS) افراد که از کم کردن پاسخ‌های سفید از پاسخ‌های نادرست به دست می‌آید. در صورت بیشتر بودن تعداد نادرست‌ها از تعداد بی-پاسخ‌ها فرد خطرپذیر و اگر تعداد نادرست‌ها کمتر از تعداد بی‌پاسخ‌ها باشد فرد خطرگریز و اگر تعداد نادرست‌ها برابر با تعداد بی‌پاسخ‌ها فرد بی‌تفاوت نسبت به خطر طبقه‌بندی می‌شود.

روش دوم: محاسبه نمره خام خطرپذیری با احتساب حدس تصادفی ۰/۳۳ (RAS0.33). در این روش ابتدا با استفاده از احتمال حدس تصادفی تأثیر حدس محاسبه شد. طبق احتمال حدس تصادفی، افرادی که در سؤال‌های چهارگزینه‌ای به‌طور تصادفی دست به انتخاب می‌زنند از هر ۴ سؤال، ۳ مورد نادرست و ۱ مورد درست خواهند داشت. پس افراد علاوه بر تعداد پاسخ‌های نادرستی که حدس زده‌اند، به اندازه یک‌سوم پاسخ‌های نادرست نیز درست حدس زده‌اند. براین اساس تعداد پاسخ‌های نادرست اصلاح‌شده افراد با عبارت زیر محاسبه می‌شود که در آن  $q$  تعداد پاسخ‌های نادرست و  $q_c$  تعداد پاسخ‌های نادرست اصلاح‌شده به خاطر حدس است:

$$q_c = q + \frac{1}{3} q$$

سپس براساس نظریه تصمیم‌گیری در شرایط عدم اطمینان با کم کردن تعداد پاسخ‌های سفید از تعداد پاسخ‌های نادرست اصلاح‌شده ( $q_c$ ) در عبارت فوق، نگرش‌های خطرپذیری افراد به‌صورت زیر مشخص می‌شود. اگر تعداد پاسخ‌های نادرست اصلاح‌شده بیشتر از

تعداد بی پاسخ‌ها باشد فرد خطرپذیر، اگر تعداد پاسخ‌های نادرست اصلاح شده کمتر از تعداد بی پاسخ‌ها باشد فرد خطرگریز و اگر تعداد پاسخ‌های نادرست اصلاح شده برابر با تعداد بی پاسخ‌ها باشد فرد بی تفاوت نسبت به خطر خواهد بود. جدول ۱ درصد افراد در هر نگرش خطرپذیری در سه آزمون را نشان می‌دهد. با توجه به این جدول درصد افراد بی تفاوت به خطر در هر سه آزمون خیلی کم می‌باشد و به این دلیل در این مقاله فقط دو گروه خطرپذیر و خطرگریز با یکدیگر مقایسه شده است.

جدول ۱. درصد افراد در هر نگرش خطرپذیری (با روش دوم) در سه آزمون

آزمون	نوع نگرش خطرپذیری	درصد
ریاضی	خطرپذیر	۴/۵۷
	بی تفاوت به خطر	۰/۰۰
	خطرگریز	۹۵/۴۳
معارف اسلامی	خطرپذیر	۳۳/۳۵
	بی تفاوت به خطر	۰/۰۲
	خطرگریز	۶۶/۶۳
زبان انگلیسی	خطرپذیر	۲۴/۱۲
	بی تفاوت به خطر	۰/۱۶
	خطرگریز	۷۵/۷۲

همان‌طور که قبلاً ذکر شد در این مقاله از نمره خطرپذیری با روش دوم استفاده شده است، به این دلیل که با استفاده از این روش علاوه بر تعداد پاسخ‌های نادرست، حتی تعداد پاسخ‌های درستی نیز که با حدس زدن پاسخ داده شده‌اند، در تعیین نگرش خطرپذیری افراد در نظر گرفته شده است که از نظر نویسنده با این روش نگرش خطرپذیری افراد دقیق‌تر مشخص می‌شود، چون که با این روش تمام حدس‌ها (نه فقط حدس‌های نادرست) محاسبه می‌شود.

روش سوم: محاسبه توانایی نگرش خطرپذیری<sup>۱</sup> (RAA) براساس نمره‌های توانایی (نه براساس نمره‌های خام). داده‌های مربوط به هر آزمون چندگزینه‌ای در قالب کدهای عددی مشخص می‌شوند. این کدها یا مربوط به پاسخ‌های افراد به چهار گزینه هر سؤال است که با کدهای ۱ تا ۴ در فایل داده‌ها نشان داده می‌شود که در هر سؤال یکی از آن‌ها پاسخ درست است و بقیه نادرست خواهند بود و یا مربوط به عدم پاسخ از سوی افراد است که

## 1. Risk Attitude Ability



در فایل داده‌ها با کد صفر مشخص می‌گردند (لازم به ذکر است که در داده‌های آزمون‌های سراسری ورود به دانشگاه افرادی که بیش از یک گزینه را به‌عنوان پاسخ انتخاب کنند، در هر سؤال با کد ۵ مشخص می‌شوند که تعداد آن‌ها خیلی کم است و همانند پژوهش حاضر می‌توان آن‌ها را به‌عنوان پاسخ نادرست در نظر گرفت). با توجه به این کدها و توصیف‌های مربوط به نمره نگرش خطرپذیری در فوق می‌توان داده‌ها را به چند صورت کدگذاری کرد: الف) اگر در داده‌های هر آزمون پاسخ‌های درست با عدد ۱ و پاسخ‌های نادرست و بی‌پاسخ‌ها با عدد صفر کدگذاری شوند و داده‌ها با مدل ۳ پارامتری لجستیک (3PL) تحلیل شوند، توانایی برآورده شده افراد با استفاده از این نوع کدگذاری، توانایی پاسخ‌های درست (p-Abi) است. ب) اگر در داده‌های هر آزمون پاسخ‌های نادرست با عدد ۱ و پاسخ‌های درست و بی‌پاسخ‌ها با عدد صفر کدگذاری شوند و داده‌های حاصل از این کدگذاری با مدل ۲ پارامتری لجستیک تحلیل شوند، توانایی برآورد شده افراد از این طریق، توانایی پاسخ نادرست (q-Abi) است. ج) اگر در داده‌های هر آزمون بی‌پاسخ‌ها با عدد ۱ و پاسخ‌های درست و نادرست با عدد صفر کدگذاری شود و داده‌های حاصل از این کدگذاری با مدل ۲ پارامتری لجستیک تحلیل شوند، توانایی برآورد شده افراد از این طریق، توانایی سفید گذاشتن<sup>۱</sup> افراد (o-Abi) است. د) اگر پاسخ‌های درست با کد یک، پاسخ‌های نادرست با کد صفر و پاسخ‌های سفید (بی‌پاسخ‌ها) با کد NA<sup>۲</sup> مشخص و داده‌ها با مدل ۳ پارامتری لجستیک شوند، توانایی افراد با احتساب داده‌های بی‌پاسخ برآورد می‌شود (Abi.ONA) که در ادامه تحت عنوان توانایی متداول مورد اشاره قرار گرفته است. بعد از برآورد انواع توانایی ذکر شده در فوق نگرش خطرپذیری افراد براساس کم کردن توانایی پاسخ‌های سفید از توانایی پاسخ‌های نادرست محاسبه و همبستگی بین آن‌ها بررسی و تحلیل‌های لازم انجام شد.

در بخشی از تحلیل‌ها لازم بود افراد از نظر توانایی برآورد شده به سه دسته ضعیف، متوسط و قوی تقسیم کنیم و سپس رابطه توانایی و خطرپذیری را در هر سه گروه بررسی نماییم. برای این کار از توانایی متداول (Abi.ONA) استفاده شد. به این صورت که در هر آزمون براساس میانگین و انحراف استاندارد توانایی متداول (جدول ۲) که تقریباً دارای

- 
1. Omit
  2. Not Available (NA)

توزیع نرمال بود، افراد طوری تقسیم شدند که ۳۳ درصد پایین به‌عنوان ضعیف، ۳۳ درصد وسط به‌عنوان متوسط و ۳۳ درصد بالا به‌عنوان قوی طبقه‌بندی شدند.

جدول ۲: میانگین و انحراف استاندارد توزیع توانایی در سه آزمون ریاضی، معارف و زبان انگلیسی

درس	میانگین	انحراف استاندارد
ریاضی	۰	۰/۷۷
معارف	۰	۰/۸۴
زبان انگلیسی	۰	۰/۷۹

سپس توانایی افراد با استفاده از نمره خام افراد برآورد شد و بعد با استفاده از تعداد پاسخ‌های درست و نادرست، نمره خام اصلاح‌شده افراد با استفاده از رابطه زیر در سطوح مختلف جریمه به دست آمد:

$$s = r - pw$$

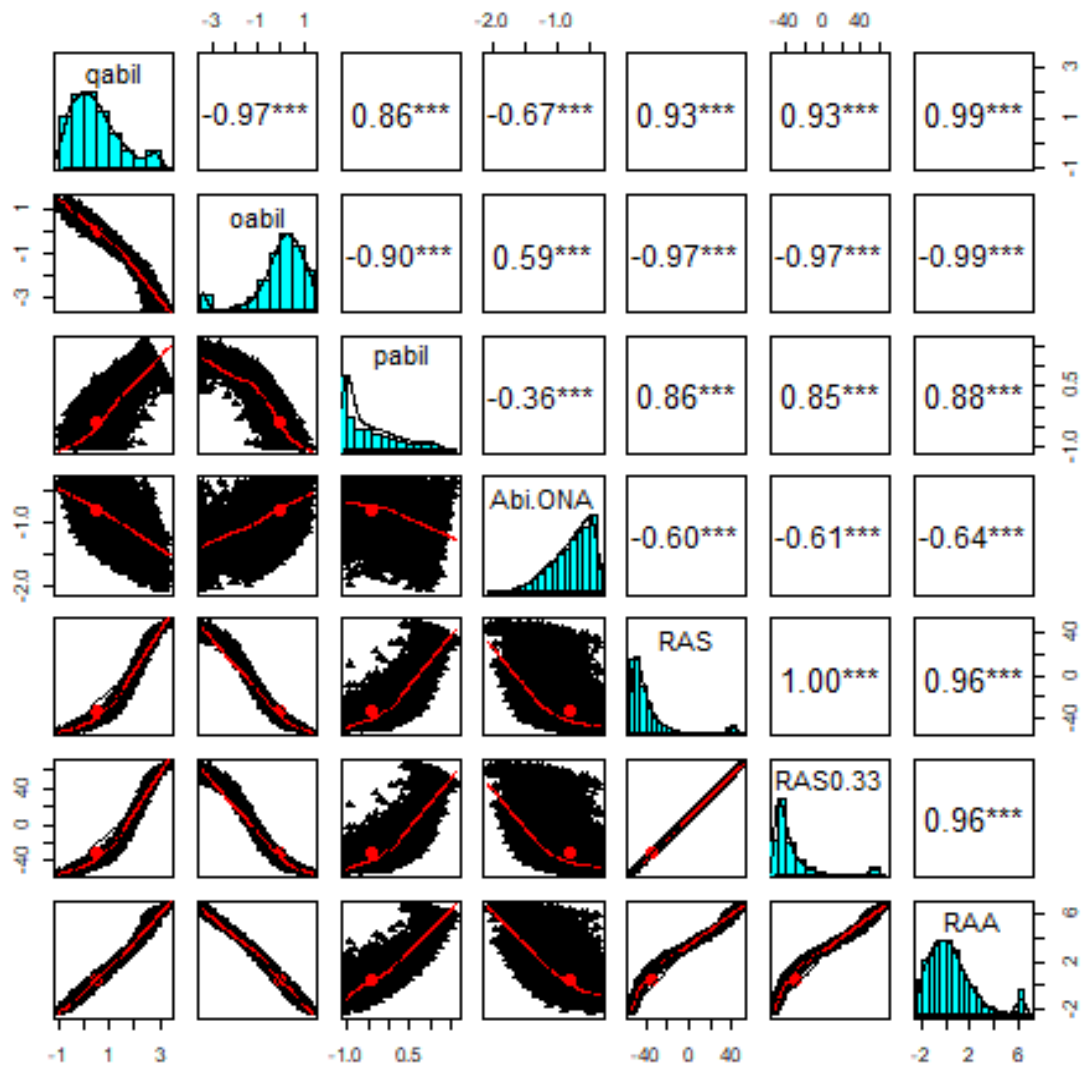
که در این رابطه  $S$  (نمره خام اصلاح‌شده)، تابعی از  $r$  (تعداد پاسخ درست)،  $w$  (تعداد پاسخ نادرست) و  $p$  (سطوح مختلف جریمه برای پاسخ‌های نادرست‌ها) است. سپس رابطه بین توانایی برآورد شده و نمره خام اصلاح‌شده افراد در سطوح مختلف جریمه در هر سه سطح توانایی در سه آزمون بررسی شد تا تأثیر سطوح مختلف جریمه بر مقدار همبستگی بین توانایی برآورد شده و نمره خام اصلاح‌شده به خاطر حدس آن‌ها و تعیین جریمه بهینه و همچنین تأثیر میزان دشواری آزمون بر میزان جریمه بهینه و مقدار همبستگی بین توانایی و نمره خام در سطوح جریمه بهینه مشخص شود. برای تحلیل از نرم‌افزار  $R$  (تیم هسته  $R^1$ ، ۲۰۱۸) و بسته‌های  $mirt1.28$  (چالمرز<sup>۲</sup>، ۲۰۱۲) و  $psych1.8.4$  (رول<sup>۳</sup>، ۲۰۱۸) استفاده شد.

## نتایج

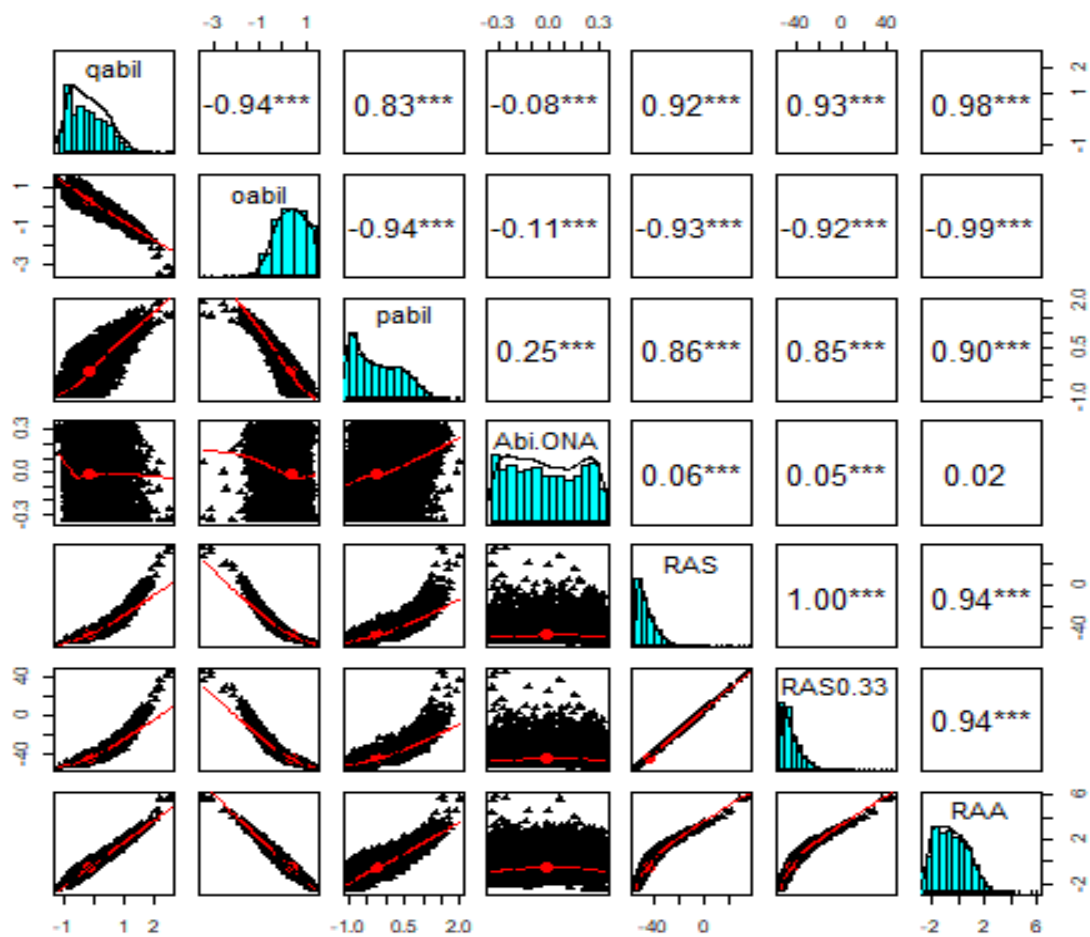
سؤال اول: آیا بین توانایی مبتنی بر پاسخ‌های نادرست، عدم پاسخ، پاسخ درست، توانایی متداول و نمره خام خطرپذیری (با روش دوم) در سطوح مختلف توانایی و نگرش‌های مختلف خطرپذیری رابطه وجود دارد؟

1. R Core Team
2. Chalmers
3. Revelle

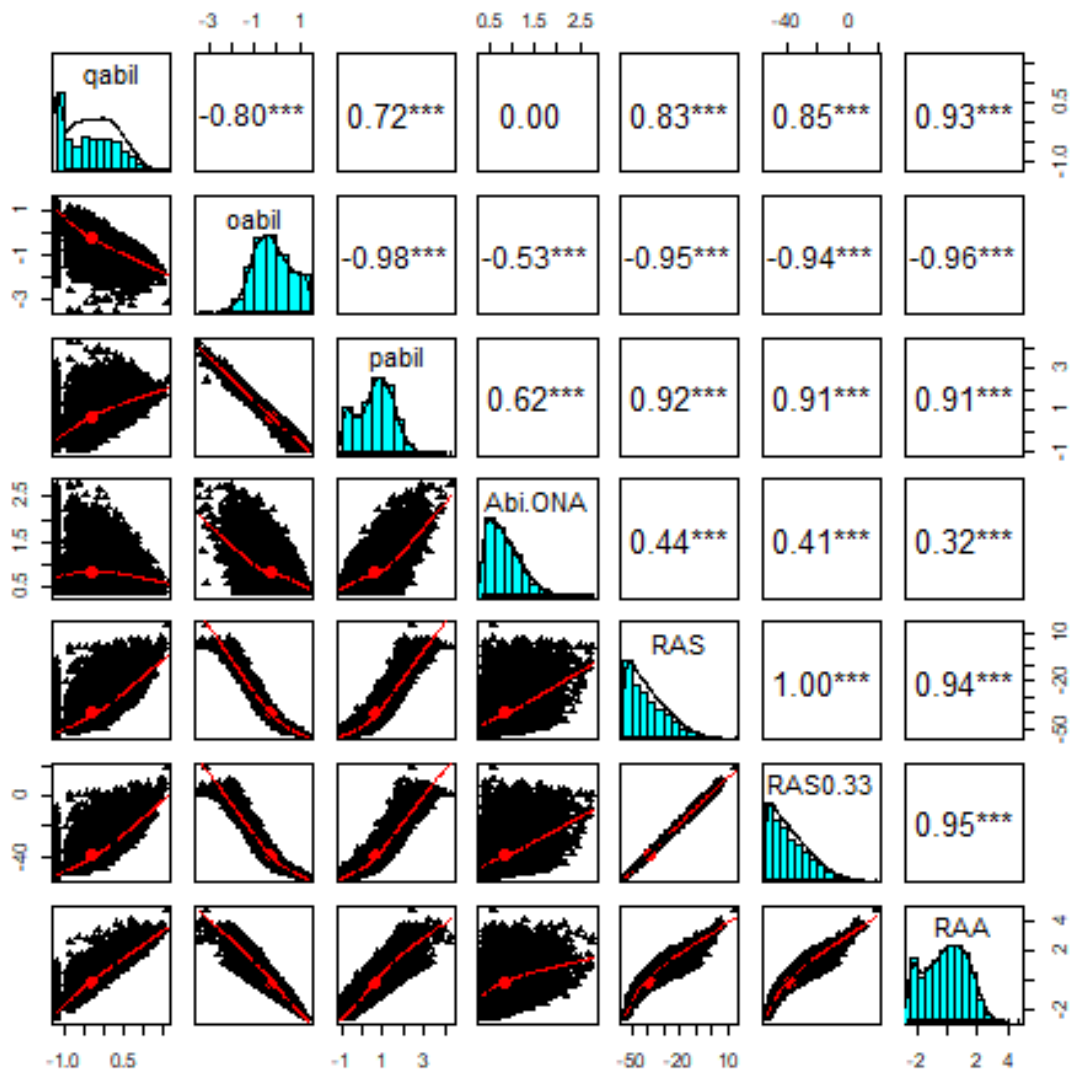
برای پاسخ به این سؤال همبستگی‌های بین توانایی‌های مبتنی بر پاسخ‌های درست، نادرست، عدم پاسخ، توانایی متداول و نمره خطرپذیری در گروه‌های مختلف (از نظر توانایی و نگرش خطرپذیری) در سه آزمون بررسی شد. نمودار ۱ همبستگی‌های متغیرهای فوق را در گروه دارای توانایی ضعیف برای آزمون ریاضی نشان می‌دهد. در این نمودار، قطر اصلی توزیع فراوانی متغیرها (توضیح داده شده در بخش روش پژوهش)، مثلث بالایی مقادیر همبستگی بین متغیرها و مثلث پایینی نمودار پراکندگی بین متغیرها را نشان می‌دهد. نمودارهای ۲ و ۳ نیز همبستگی‌های ذکر شده را برای گروه متوسط و قوی در آزمون ریاضی نشان می‌دهد. لازم به ذکر است که به دلیل محدودیت حجم مقاله امکان ارائه نمودارهای مشابه برای آزمون‌های معارف و زبان وجود نداشت.



نمودار ۱. رابطه توانایی‌های مبتنی بر انواع پاسخ با توانایی متداول، نمره‌های خام و توانایی خطرپذیری گروه ضعیف (آزمون ریاضی)



نمودار ۲. رابطه توانایی‌های مبتنی بر انواع پاسخ با توانایی متداول، نمره‌های خام و توانایی خطرپذیری گروه متوسط (ریاضی)



نمودار ۳. رابطه توانایی‌های مبتنی بر انواع پاسخ با توانایی متداول، نمره‌های خام و توانایی خطرپذیری گروه قوی (ریاضی)

همبستگی متغیرهای فوق در گروه‌های مختلف (از نظر توانایی و نگرش خطرپذیری) در سه آزمون بررسی و نتایج آن‌ها در جداول ۳ تا ۶ خلاصه شده است. جدول ۳ همبستگی توانایی مبتنی بر کدگذاری انواع پاسخ‌ها و توانایی متداول در سه آزمون در سطوح مختلف توانایی را نشان می‌دهد. با توجه به این همبستگی‌ها، رابطه توانایی مبتنی بر عدم پاسخ با پاسخ نادرست در همه سطوح توانایی در هر سه آزمون منفی و نسبتاً بالا است ولی با افزایش سطح توانایی شدت این همبستگی کاهش می‌یابد. رابطه توانایی مبتنی بر عدم پاسخ با توانایی مبتنی بر پاسخ‌های درست در همه سطوح توانایی هر سه آزمون منفی و بالا است و با افزایش سطح توانایی شدت این همبستگی افزایش می‌یابد که نشان می‌دهد با افزایش تعداد بی‌پاسخ‌ها تعداد پاسخ‌های نادرست و درست هر دو کاهش می‌یابند. رابطه توانایی مبتنی بر پاسخ‌های درست با نادرست در هر سه آزمون، در تمام سطوح توانایی مثبت است ولی شدت همبستگی آن در گروه متوسط بیشتر و در گروه قوی کمتر است که نشان می‌دهد در گروه متوسط تعداد پاسخ‌هایی که به‌طور حدسی پاسخ داده شده‌اند بیشتر و در گروه قوی کمتر است، چون هنگامی که پاسخ‌ها با حدس زدن باشند پاسخ‌های درست و نادرست به نسبت ۱ به ۳ افزایش می‌یابند. رابطه توانایی متداول با توانایی مبتنی بر پاسخ نادرست در تمام سطوح توانایی تقریباً منفی است ولی شدت همبستگی آن در گروه ضعیف بیشتر از دو گروه دیگر است که نشان می‌دهد با افزایش توانایی، تعداد پاسخ‌های نادرست کاهش می‌یابد. همبستگی توانایی متداول با توانایی عدم پاسخ در گروه ضعیف مثبت، در گروه متوسط تقریباً صفر ولی در گروه قوی منفی است که نشان دهنده این است که در گروه ضعیف به دلیل دانش کمتر و عدم توانایی حذف گزینه‌های نادرست، با افزایش توانایی، تعداد بی‌پاسخ‌ها بیشتر می‌شود ولی در گروه قوی به دلیل دانش بیشتر و توانایی حذف گزینه‌های نادرست، با افزایش توانایی، تعداد بی‌پاسخ‌ها کمتر می‌شود. همبستگی توانایی متداول با توانایی مبتنی بر پاسخ درست در تمام گروه‌ها تقریباً مثبت است، ولی شدت همبستگی آن در گروه قوی بیشتر از دو گروه دیگر است که نشان می‌دهد با افزایش توانایی، تعداد پاسخ‌های درست نیز افزایش می‌یابد.

جدول ۳: همبستگی توانایی مبتنی بر انواع کدگذاری در گروه ضعیف، متوسط و قوی برای هر سه آزمون

رابطه	آزمون	ضعیف	متوسط	قوی
توانایی عدم پاسخ با پاسخ نادرست	ریاضی	-۰/۹۷	-۰/۹۴	-۰/۸۰
	معارف	-۰/۹۲	-۰/۹۲	-۰/۵۴
	زبان	-۰/۹۴	-۰/۹۴	-۰/۶۵
توانایی عدم پاسخ با پاسخ درست	ریاضی	-۰/۹۰	-۰/۹۴	-۰/۹۸
	معارف	-۰/۷۸	-۰/۹۶	-۰/۹۵
	زبان	-۰/۸۰	-۰/۹۷	-۰/۹۷
توانایی پاسخ درست با پاسخ نادرست	ریاضی	۰/۸۶	۰/۸۳	۰/۷۲
	معارف	۰/۵۵	۰/۸۴	۰/۳۲
	زبان	۰/۶۳	۰/۸۸	۰/۵۱
توانایی متداول با توانایی پاسخ نادرست	ریاضی	-۰/۶۷	-۰/۰۸	۰/۰۰
	معارف	-۰/۵۳	-۰/۲۶	-۰/۵۲
	زبان	-۰/۵۳	-۰/۲۹	-۰/۱۸
توانایی متداول با توانایی عدم پاسخ	ریاضی	۰/۵۹	-۰/۱۱	-۰/۵۳
	معارف	۰/۲۲	۰/۰۵	-۰/۳۶
	زبان	۰/۳۱	۰/۱۱	-۰/۵۵
توانایی متداول با توانایی پاسخ درست	ریاضی	-۰/۳۶	۰/۲۵	۰/۶۲
	معارف	۰/۳۴	۰/۲۰	۰/۵۷
	زبان	۰/۲۰	۰/۰۰	۰/۶۹

همبستگی نمره خام خطرپذیری اصلاح شده به خاطر حدس (RASO.33) با توانایی مبتنی بر انواع پاسخها و توانایی متداول در سه آزمون در سطوح مختلف توانایی در جدول ۴ نشان داده شده است. با توجه به این جدول، رابطه نمره خام خطرپذیری اصلاح شده با توانایی مبتنی بر پاسخهای نادرست در همه سطوح توانایی مثبت و بالا است ولی با افزایش سطح توانایی شدت همبستگی آن کمتر شده است که نشان دهنده این است که با افزایش خطرپذیری تعداد پاسخهای نادرست نیز افزایش می یابد ولی تعداد پاسخهای نادرست در گروه قوی به دلیل دانش بالا کمتر است. همچنین رابطه نمره خام اصلاح شده خطرپذیری با توانایی مبتنی بر پاسخهای درست نیز در تمام سطوح توانایی مثبت و بالا است ولی شدت همبستگی آن در گروه ضعیف کمتر از دو گروه دیگر است که نشان می دهد با افزایش



خطرپذیری تعداد پاسخ‌های درست نیز افزایش می‌یابد ولی در گروه ضعیف به دلیل دانش کمتر و عدم توانایی حذف گزینه‌های نادرست، تعداد پاسخ‌های درست کمتر افزایش می‌یابد. رابطه نمره خام خطرپذیری با توانایی مبتنی بر عدم پاسخ در تمام سطوح توانایی منفی است ولی شدت آن در گروه ضعیف بیشتر از دو گروه دیگر است که نشان می‌دهد با افزایش خطرپذیری تعداد بی‌پاسخ‌ها کاهش می‌یابد ولی در گروه ضعیف به دلیل دانش کمتر و عدم توانایی حذف گزینه‌های نادرست تعداد بی‌پاسخ‌ها بیشتر است. رابطه نمره خام خطرپذیری با توانایی متداول در گروه ضعیف منفی، در گروه متوسط تقریباً صفر و در گروه قوی مثبت است که نشان می‌دهد در گروه ضعیف با افزایش توانایی به دلیل دانش کمتر و عدم حذف گزینه‌های نادرست، خطرپذیری کمتر ولی در گروه قوی با افزایش توانایی به دلیل دانش بیشتر و توانایی حذف گزینه‌های نادرست، خطرپذیری بیشتر می‌شود.

جدول ۴. همبستگی نمره خام خطرپذیری اصلاح‌شده به خاطر حدس (RAS0.33) با توانایی مبتنی بر انواع کدگذاری و توانایی متداول در سه آزمون برای سطوح مختلف توانایی

رابطه	آزمون	ضعیف	متوسط	قوی
نمره خطرپذیری با توانایی پاسخ نادرست	ریاضی	۰/۹۳	۰/۹۳	۰/۸۵
	معارف	۰/۹۷	۰/۹۶	۰/۷۵
	زبان	۰/۹۷	۰/۹۷	۰/۸۱
نمره خطرپذیری با توانایی عدم پاسخ	ریاضی	-۰/۹۷	-۰/۹۲	-۰/۹۴
	معارف	-۰/۹۷	-۰/۹۷	-۰/۹۴
	زبان	-۰/۹۷	-۰/۹۷	-۰/۹۶
نمره خطرپذیری با توانایی پاسخ درست	ریاضی	۰/۸۵	۰/۸۵	۰/۹۱
	معارف	۰/۶۷	۰/۹۱	۰/۸۵
	زبان	۰/۷۱	۰/۹۳	۰/۹۰
نمره خطرپذیری با توانایی متداول	ریاضی	-۰/۶۱	۰/۰۵	۰/۴۱
	معارف	-۰/۳۷	-۰/۰۵	۰/۱۳
	زبان	-۰/۴۱	-۰/۱۴	۰/۳۸

همبستگی توانایی مبتنی بر انواع پاسخ‌ها و توانایی متداول با یکدیگر در سه آزمون در هر نگرش خطرپذیری در جدول ۵ نشان داده شده است. با توجه به این جدول، رابطه

توانایی مبتنی بر عدم پاسخ با نادرست به جز آزمون معارف و زبان در گروه خطرپذیر در بقیه موارد منفی است ولی در گروه خطرگریز شدت رابطه منفی بیشتر است که نشان دهنده این است در گروه خطرگریز با کاهش تعداد پاسخ‌های نادرست تعداد بی‌پاسخ‌ها افزایش می‌یابد. رابطه توانایی مبتنی بر عدم پاسخ با درست نیز در هر دو نگرش منفی است ولی شدت همبستگی آن در گروه خطرگریز بیشتر است که نشان می‌دهد در هر دو نگرش با کاهش تعداد پاسخ‌های درست تعداد بی‌پاسخ‌ها افزایش می‌یابد، ولی در گروه خطرگریز با کاهش تعداد پاسخ‌های درست، تعداد بی‌پاسخ‌ها بیشتر افزایش می‌یابد.

رابطه توانایی مبتنی بر پاسخ‌های درست با نادرست نیز منفی ولی شدت همبستگی آن در گروه خطرپذیر بیشتر است که نشان می‌دهد با افزایش تعداد پاسخ‌های درست تعداد پاسخ‌های نادرست کاهش می‌یابد. رابطه توانایی متداول با توانایی مبتنی بر پاسخ‌های نادرست نیز منفی ولی شدت همبستگی آن در گروه خطرپذیر بیشتر است که نشان دهنده این است که با افزایش توانایی متداول تعداد پاسخ‌های نادرست کاهش می‌یابد. رابطه توانایی متداول با توانایی مبتنی بر عدم پاسخ نیز منفی است و شدت همبستگی آن به جز آزمون ریاضی که در گروه خطرپذیر شدیدتر است، در دو آزمون دیگر در هر دو نگرش تقریباً برابر است که نشان می‌دهد با افزایش توانایی متداول تعداد بی‌پاسخ‌ها نیز کاهش می‌یابد. رابطه توانایی متداول با توانایی مبتنی بر پاسخ‌های درست در هر دو نگرش مثبت است ولی شدت همبستگی آن در گروه خطرپذیر بیشتر است که نشان می‌دهد با افزایش توانایی متداول تعداد پاسخ‌های درست نیز در هر دو نگرش افزایش می‌یابد، ولی در گروه خطرپذیر با افزایش توانایی به دلیل حذف گزینه‌های نادرست بیشتر و خطرپذیری آن‌ها در حدس زدن تعداد درست‌ها بیشتر افزایش می‌یابد.

جدول ۵: همبستگی توانایی مبتنی بر انواع پاسخ‌ها و توانایی متداول با یکدیگر در سه آزمون به تفکیک نگرش خطرپذیری (با روش دوم)

رابطه	آزمون	خطرپذیر	خطرگریز
توانایی عدم پاسخ با توانایی پاسخ نادرست	ریاضی	-۰/۶۴	-۰/۷۰
	معارف	۰/۱۹	-/۲۰
	زبان	۰/۱۹	-۰/۵۲
توانایی عدم پاسخ با توانایی پاسخ درست	ریاضی	-۰/۳۲	-۰/۸۹
	معارف	-۰/۶۷	-۰/۹۳
	زبان	-۰/۶۱	-۰/۹۲
توانایی پاسخ درست با توانایی پاسخ نادرست	ریاضی	-۰/۴۶	۰/۳۵
	معارف	-۰/۸۳	-۰/۱۶
	زبان	-۰/۸۷	۰/۱۸
توانایی متداول با توانایی پاسخ نادرست	ریاضی	-۰/۷۷	-۰/۴۰
	معارف	-۰/۹۳	-۰/۷۱
	زبان	-۰/۹۴	-۰/۵۴
توانایی متداول با توانایی عدم پاسخ	ریاضی	-۰/۹۲	-۰/۳۲
	معارف	-۰/۴۷	-۰/۵۰
	زبان	-۰/۴۴	-۰/۳۵
توانایی متداول با توانایی پاسخ درست	ریاضی	۰/۸۱	۰/۶۶
	معارف	۰/۹۶	۰/۷۷
	زبان	۰/۹۷	۰/۶۶

همبستگی نمره خام خطرپذیری اصلاح شده به خاطر حدس (RAS0.33) با توانایی مبتنی بر انواع پاسخ‌ها و توانایی متداول در سه آزمون برای نگرش‌های خطرپذیری در جدول ۶ خلاصه شده است. با توجه به این جدول، رابطه خطرپذیری با توانایی مبتنی بر پاسخ‌های نادرست در هر دو نگرش مثبت ولی شدت آن در گروه خطرپذیر بیشتر است که نشان می‌دهد در هر دو نگرش با افزایش خطرپذیری تعداد پاسخ‌های نادرست افزایش می‌یابد ولی در گروه خطرپذیر به دلیل خطرپذیری بالا در حدس زدن تعداد نادرست‌ها بیشتر افزایش می‌یابد. رابطه خطرپذیری با توانایی مبتنی بر عدم پاسخ در هر دو نگرش منفی ولی شدت آن در گروه خطرگریز بیشتر است که نشان می‌دهد با افزایش خطرپذیری

در هر دو نگرش تعداد بی‌پاسخ‌ها کمتر می‌شود. رابطه خطرپذیری با توانایی مبتنی بر پاسخ درست در گروه خطرپذیر منفی و در گروه خطرگریز مثبت است که نشان می‌دهد در گروه خطرپذیر با افزایش خطرپذیری تعداد پاسخ‌های درست کمتر ولی در گروه خطرگریز با افزایش خطرپذیری تعداد پاسخ‌های درست بیشتر می‌شود. رابطه خطرپذیری با توانایی متداول در گروه خطرپذیر منفی و در گروه خطرگریز تقریباً صفر است که نشان می‌دهد در گروه خطرپذیر با افزایش خطرپذیری توانایی متداول کاهش می‌یابد.

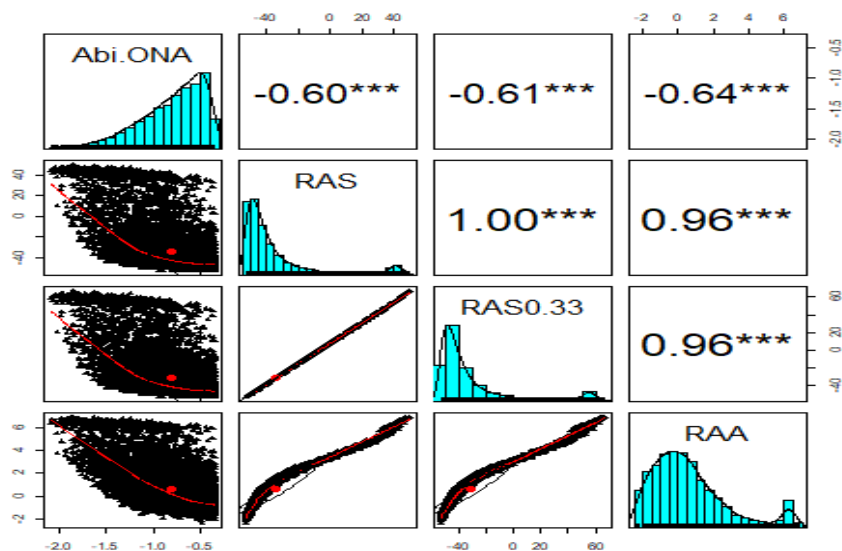
جدول ۶: همبستگی نمره خام خطرپذیری اصلاح‌شده به خاطر حدس (RAS0.33) با توانایی

مبتنی بر انواع پاسخ‌ها و توانایی متداول در سه آزمون در نگرش‌های خطرپذیری

رابطه	آزمون	خطرپذیر	خطرگریز
نمره خطرپذیری با توانایی پاسخ نادرست	ریاضی	۰/۸۹	۰/۸۱
	معارف	۰/۷۷	۰/۵۹
	زبان	۰/۸۲	۰/۷۵
نمره خطرپذیری با توانایی عدم پاسخ	ریاضی	-۰/۹۰	-۰/۹۲
	معارف	-۰/۴۴	-۰/۸۸
	زبان	-۰/۳۷	-۰/۹۳
نمره خطرپذیری با توانایی پاسخ درست	ریاضی	-۰/۰۴	۰/۷۳
	معارف	-۰/۳۳	۰/۶۷
	زبان	-۰/۴۷	۰/۷۵
نمره خطرپذیری با توانایی متداول	ریاضی	-۰/۴۷	۰/۰۹
	معارف	-۰/۵۴	۰/۱۱
	زبان	-۰/۶۳	۰/۰۸

**سؤال دوم: آیا رابطه بین توانایی و خطرپذیری در گروه‌های ضعیف، متوسط و قوی متفاوت است؟**

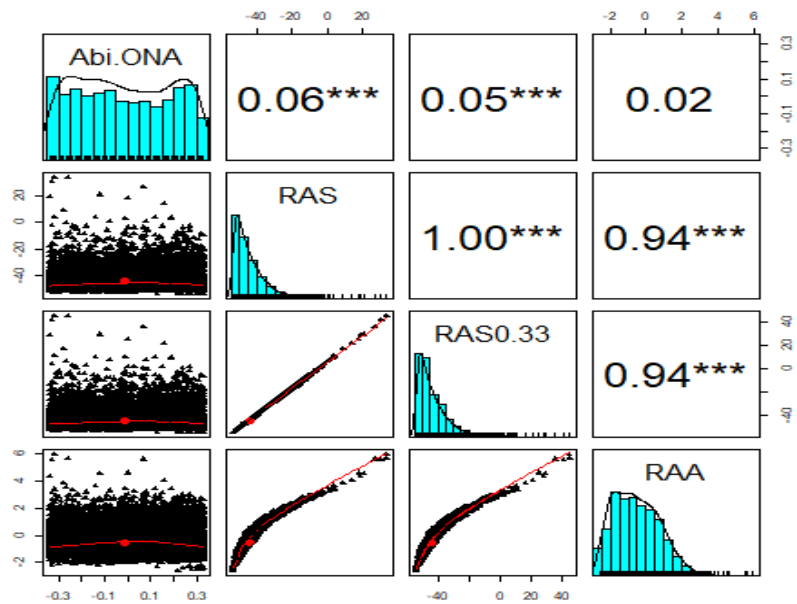
برای پاسخ به این سؤال همبستگی توانایی متداول با نمره‌های خطرپذیری در سه سطح توانایی در سه آزمون بررسی شد. برای مثال نمودار ۴ همبستگی توانایی متداول با سه نوع نمره عادی خطرپذیری، نمره اصلاح‌شده خطرپذیری به خاطر حدس و توانایی خطرپذیری را برای گروه ضعیف در آزمون ریاضی نشان می‌دهد.



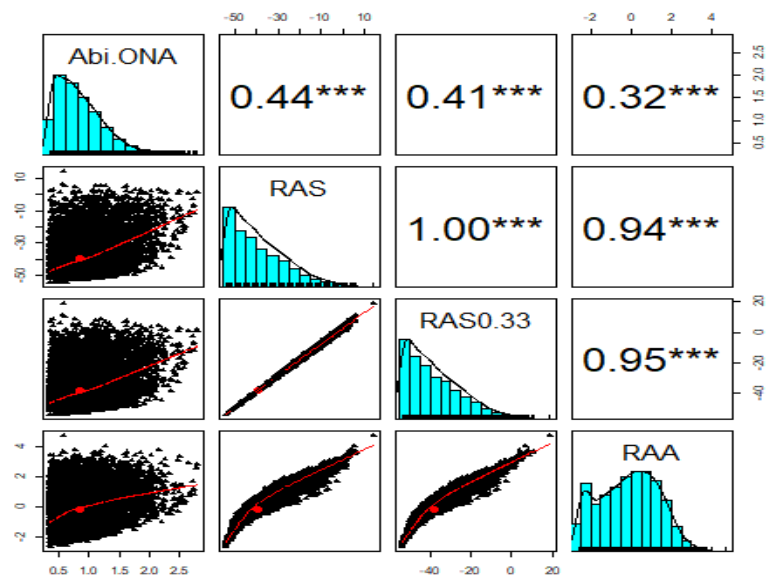
نمودار ۴. همبستگی نمره‌های خطرپذیری با توانایی متداول برای افراد ضعیف (ریاضی)

نمودارهای ۵ و ۶ نیز همبستگی‌های ذکر شده در فوق را به ترتیب برای گروه‌های متوسط و قوی در آزمون ریاضی نشان می‌دهد.

جدول ۷ همبستگی بین توانایی متداول و نمره خام خطرپذیری اصلاح شده به خاطر حدس (RAS0.33) را در هر سطح توانایی در سه آزمون را می‌دهد. با توجه به این جدول، مقادیر همبستگی در گروه متوسط خیلی قابل توجه نیست و تقریباً صفر است. این مقادیر همبستگی نشان می‌دهد که رابطه بین نگرش خطرپذیری افراد با سطح توانایی افراد در گروه با توانایی متوسط اگرچه معنادار است ولی قابل توجه نیست. در سطوح توانایی کم، رابطه توانایی با خطرپذیری منفی و در سطوح توانایی بالا، رابطه توانایی با خطرپذیری مثبت است؛ یعنی در سطوح توانایی کم با افزایش توانایی، به دلیل دانش کمتر و عدم توانایی حذف گزینه‌های نادرست، خطرپذیری کم و خطرگریزی بیشتر می‌شود ولی در سطوح توانایی بالا، به دلیل دانش بیشتر و توانایی حذف گزینه‌های نادرست بیشتر، با افزایش توانایی، خطرپذیری بیشتر و خطرگریزی کمتر می‌گردد. با توجه به این نتایج می‌توان نتیجه گرفت که افراد با توانایی بالا، با داشتن دانش بیشتر و حذف کردن گزینه‌های نادرست بیشتر، خطرپذیرتر و افراد با توانایی پایین، با دانش کمتر و عدم توانایی حذف گزینه‌های نادرست، خطرگریزتر هستند.



نمودار ۵. همبستگی نمره‌های خطرپذیری با توانایی متداول برای افراد متوسط (ریاضی)



نمودار ۶. همبستگی نمره‌های خطرپذیری با توانایی متداول برای افراد قوی (ریاضی)

جدول ۷: مقادیر همبستگی‌های بین توانایی و خطرپذیری در سطوح مختلف توانایی در سه آزمون

آزمون	گروه ضعیف	متوسط	قوی
ریاضی	-۰/۶۱	۰/۰۵	۰/۴۱
معارف	-۰/۳۷	-۰/۰۵	۰/۱۳
زبان	-۰/۴۱	-۰/۱۴	۰/۳۸

همچنین برای بررسی این سؤال میانگین نمره افراد در هر سطح توانایی به تفکیک هر نگرش خطرپذیری محاسبه شد. جدول ۸ این میانگین‌ها را در هر سه آزمون به تفکیک سطوح توانایی و نگرش خطرپذیری نشان می‌دهد. با توجه به این جدول، در گروه ضعیف در هر سه آزمون میانگین توانایی متداول افراد خطرپذیر کمتر از میانگین توانایی متداول افراد خطرگریز است. در گروه متوسط در هر سه آزمون میانگین توانایی افراد خطرپذیر و میانگین توانایی افراد خطرگریز تفاوت چندانی با هم ندارند. در گروه قوی در دو آزمون ریاضی و زبان میانگین توانایی افراد خطرپذیر بیشتر از میانگین توانایی افراد خطرگریز است.

این نتایج نشان می‌دهد که خطرپذیری در افراد ضعیف باعث کاهش نمره آن‌ها و در افراد قوی باعث افزایش نمره آن‌ها می‌شود. باز هم می‌توان به همان نتایج قسمت قبل رسید که افراد با توانایی بالا بعد از حذف تعدادی گزینه نادرست، از بین گزینه‌های باقی مانده بیشتر حدس می‌زنند و این ریسک در میان گزینه‌های کمتر منجر به افزایش نمره آن‌ها می‌شود ولی افراد با توانایی کمتر چون نمی‌توانند گزینه‌ای را حذف کنند احتمال درست کمتری دارند و خطرگریزی افراد ضعیف باعث افزایش نمره افراد ضعیف نسبت به نمره افراد ضعیف با نگرش خطرپذیر می‌شود. چون افراد ضعیفی که خطرپذیر هستند نمره آن‌ها به دلیل حدس در بین گزینه‌های بیشتر، کاهش می‌یابد.

براساس نتایج پژوهش اسپینوزا و گاردیوزبل (۲۰۱۰) بیزاری از خطر موجب می‌شود تا تعداد عدم پاسخ‌های افراد بیشتر شده و نمره افراد کاهش یابد، البته تفاوت افراد بیزار از خطر و خطرپذیر در سطوح توانایی پایین قابل توجه ولی در سطوح توانایی بالا ناچیز بوده است. براساس نتایج پژوهش حاضر نیز در گروه قوی بیزاری از خطر موجب کاهش نمره و خطرپذیری باعث افزایش نمره آن‌ها می‌شود ولی در گروه ضعیف نتیجه برعکس می‌باشد

یعنی بیزاری از خطر افراد ضعیف باعث افزایش نمره آن‌ها نسبت به افراد ضعیف خطرپذیر می‌شود.

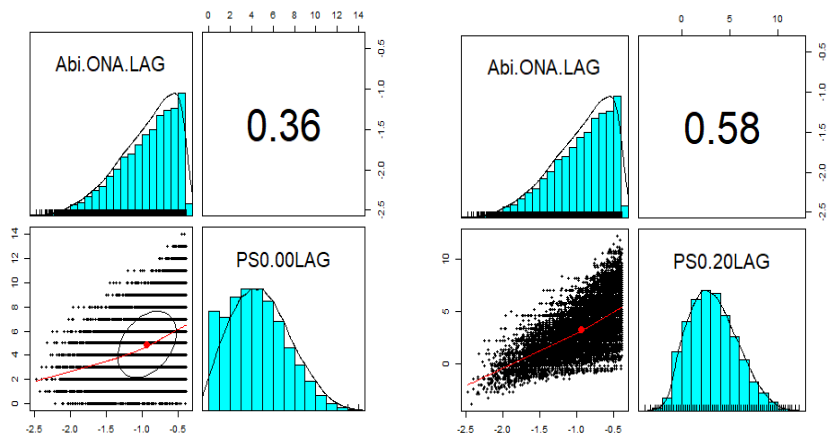
جدول ۸: میانگین توانایی متداول هر نگرش خطرپذیری در هر سطح توانایی در سه آزمون

سطوح توانایی	آزمون‌ها	خطرپذیرها	خطرگریزها	تفاوت میانگین‌ها
ضعیف	ریاضی	-۱/۲۷	-۰/۷۴	۰/۵۳
	معارف	-۱/۰۶	-۰/۸۵	۰/۲۱
	زبان	-۱/۰۸	-۰/۸۰	۰/۲۸
متوسط	ریاضی	-۰/۰۷	-۰/۰۱	۰/۰۶
	معارف	-۰/۰۱	۰/۰۰	۰/۰۱
	زبان	۰/۰۱	۰/۰۲	۰/۰۱
قوی	ریاضی	۱/۲۵	۰/۸۴	۰/۴۱
	معارف	۰/۹۲	۰/۹۳	۰/۰۱
	زبان	۱/۰۲	۰/۷۶	۰/۲۶

سؤال سوم: تأثیر سطوح مختلف جریمه بر همبستگی بین توانایی و نمره خام در سطوح مختلف توانایی چیست؟

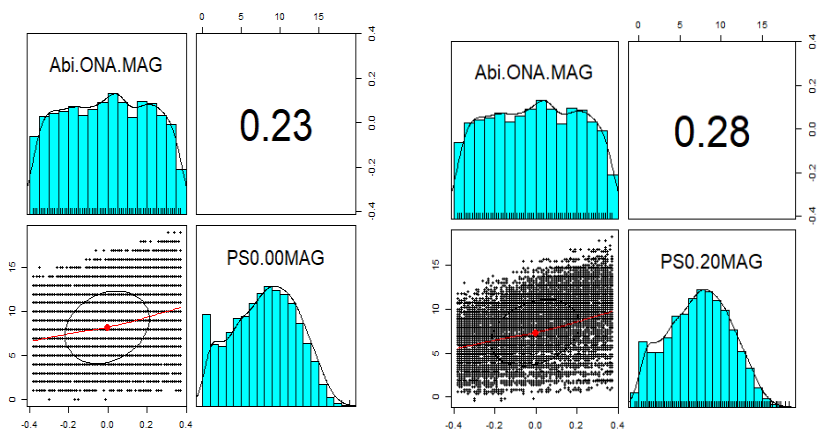
به منظور بررسی تأثیر سطوح مختلف جریمه بر رابطه بین توانایی و نمره خام افراد، همبستگی بین نمره به دست آمده و توانایی متداول در سطوح مختلف جریمه (۰ تا ۲) در هر سه سطح توانایی پایین، متوسط و بالا بررسی شد. برای مثال نمودار ۷ همبستگی بین توانایی متداول برآورد شده و نمره خام اصلاح شده افراد در سطوح جریمه صفر (بدون جریمه) و ۰/۲۰ را در آزمون معارف برای گروه ضعیف نشان می‌دهد.



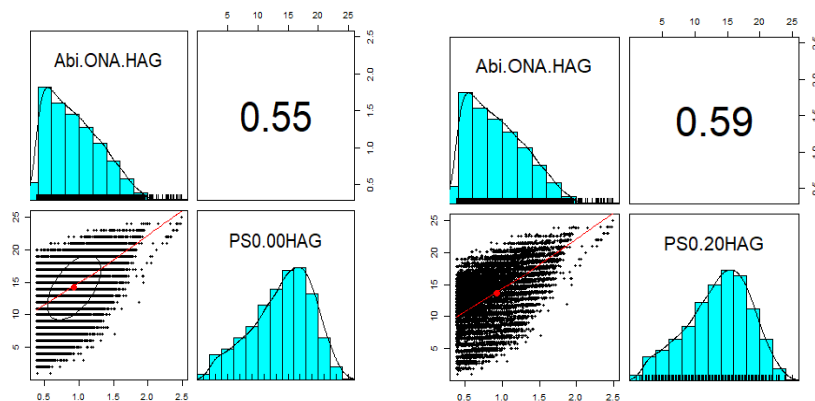


نمودار ۷. نمودارهای همبستگی توانایی متداول با نمره خام گروه ضعیف در سطوح جریمه صفر (راست) و ۰/۲۰ (چپ) (آزمون معارف)

نمودارهای ۸ و ۹ نیز همبستگی توانایی متداول با نمره خام اصلاح شده را در سطوح جریمه صفر و ۰/۲۰ برای آزمون معارف در دو گروه متوسط و قوی نشان می دهد.



نمودار ۸. نمودارهای همبستگی توانایی متداول با نمره خام گروه متوسط در سطوح جریمه صفر (راست) و ۰/۲۰ (چپ) (آزمون معارف)



نمودار ۹. نمودارهای همبستگی توانایی متداول با نمره خام گروه قوی در سطوح جریمه صفر (راست) و ۰/۲۰ (چپ) (آزمون معارف)

به علاوه، همبستگی‌های بین توانایی متداول برآورد شده و نمره خام اصلاح شده افراد در سطوح مختلف جریمه برای هر سه سطح توانایی در هر سه آزمون بررسی شد و نتایج آن‌ها در جدول ۹ نشان داده شده است. با توجه به این مقادیر همبستگی، عدم اعمال جریمه و یا اعمال جریمه کم که باعث تشویق به حدس زدن بیشتر می‌شود موجب کاهش رابطه بین توانایی و نمره خام در همه گروه‌ها می‌گردد. اعمال جریمه خیلی بالا (برای اصلاح حدس) نیز موجب کاهش رابطه بین توانایی و نمره خام می‌شود؛ ولی مقدار جریمه بالایی که موجب کاهش رابطه بین توانایی و نمره خام شود برای گروه قوی بیشتر از مقدار آن برای گروه‌های دیگر است. براساس نتایج پژوهش‌های گذشته نیز حدس زدن (شرایط عدم جریمه) موجب خطای اندازه‌گیری در نمره‌ها می‌شود و نمره افراد براساس دانش واقعی آن‌ها برآورد نمی‌شود و به این ترتیب موجب کاهش پایایی آزمون می‌شود (فراری ۱۹۸۸؛ بورتون ۲۰۰۱ و کمپل ۲۰۱۵).

همچنین براساس نتایج پژوهش‌های گذشته اصلاح حدس موجب کاهش خطای اندازه‌گیری ناشی از حدس و افزایش پایایی آزمون می‌شود. البته اصلاح حدس استفاده شده در پژوهش‌های گذشته به مقدار مرسوم آن (جریمه یک سوم برای سؤال‌های چهارگزینه‌ای) بوده است (لرد<sup>۱</sup> ۱۹۶۴؛ بورتون، ۲۰۰۲ و کمپل ۲۰۱۵).

1. Lord

نقش نگرش خطرپذیری افراد، سطوح مختلف جریمه... / ۴۵

جدول ۹: مقادیر همبستگی بین توانایی و نمره خام اصلاح شده در سطوح مختلف جریمه در سطوح مختلف توانایی برای سه آزمون

گروهها	جریمه	۰/۲۰	۰/۳۳	۰/۴۰	۰/۶۰	۰/۸۰	۱/۰۰	۱/۲۰	۱/۴۰	۱/۶۰	۱/۸۰	۲/۰۰
ضعیف	ریاضی	-۰/۳۵	۰/۰۲	۰/۵۰	۰/۶۸	۰/۷۸	۰/۷۵	۰/۷۳	۰/۷۱	-	-	-
	معارف	۰/۳۶	۰/۵۸	۰/۷۱	۰/۷۶	۰/۸۵	۰/۸۶	۰/۸۵	۰/۸۲	-	-	-
	زبان	۰/۲۲	۰/۴۵	۰/۶۰	۰/۶۶	۰/۷۸	۰/۸۱	۰/۸۰	۰/۷۸	-	-	-
متوسط	ریاضی	۰/۲۵	۰/۳۱	۰/۳۵	۰/۳۸	۰/۴۶	۰/۵۲	۰/۵۵	۰/۵۱	۰/۴۵	-	-
	معارف	۰/۲۳	۰/۲۸	۰/۳۲	۰/۳۵	۰/۴۲	۰/۴۹	۰/۵۶	۰/۶۲	۰/۶۶	-	-
	زبان	۰/۰۴	۰/۰۸	۰/۱۱	۰/۱۳	۰/۱۹	۰/۲۶	۰/۳۴	۰/۴۳	۰/۵۱	-	-
قوی	ریاضی	۰/۶۵	۰/۶۷	۰/۶۹	۰/۷۰	۰/۷۲	۰/۷۴	۰/۷۶	۰/۷۸	۰/۸۰	۰/۸۱	۰/۸۲

۰/۸۵ ۰/۸۴ ۰/۸۲ ۰/۸۰ ۰/۷۷ ۰/۷۴ ۰/۷۱ ۰/۶۷ ۰/۶۳ ۰/۶۲ ۰/۵۹ ۰/۵۵

ب.۴

۰/۹۰ ۰/۸۹ ۰/۸۸ ۰/۸۶ ۰/۸۴ ۰/۸۱ ۰/۷۸ ۰/۷۶ ۰/۷۳ ۰/۷۲ ۰/۷۰ ۰/۶۷

ب.۵

### سؤال چهارم: جریمه بهینه برای اصلاح حدس برای سطوح مختلف توانایی چه میزان است؟

برای پاسخ به این پرسش، همبستگی بین توانایی متداول برآورد شده و نمره خام اصلاح شده در سطوح مختلف جریمه در هر سه سطح توانایی پایین، متوسط و بالا بررسی شد و اولین سطوحی از جریمه که رابطه بین توانایی و نمره خام را بیشینه می‌سازد به‌عنوان سطوح جریمه بهینه در هر سطح توانایی انتخاب شد. جدول ۱۰ مقادیر بیشینه همبستگی بین توانایی و نمره خام اصلاح شده در سطوح جریمه بهینه را برای هر یک از سه گروه توانایی نشان می‌دهد. در این جدول در سطوحی از جریمه که مقادیر همبستگی بین توانایی متداول و نمره خام اصلاح شده ثابت مانده و یا کاهش پیدا کرده است، آن مقادیر همبستگی ارائه نشده است (به‌صورت جا خالی است). با توجه به این جدول میزان جریمه بهینه برای هر آزمون با دشواری متفاوت و با توجه به سطوح مختلف توانایی متداول افراد متفاوت است؛ با بررسی رابطه‌های بیشینه در سطوح مختلف توانایی در هر سه آزمون متوجه می‌شویم که میزان جریمه بهینه برای افراد ضعیف کمتر و برای افراد قوی بالاتر است. علت آن را می‌توان با توجه به همبستگی نمره توانایی متداول این افراد با نمره خطرپذیری (با روش دوم) که در سؤال دوم بررسی شد شرح داد. با توجه به نتایجی که در سؤال دوم به دست آمد افراد قوی به دلیل دانش بالا و حذف تعدادی از گزینه‌های نادرست خطرپذیرتر و افراد ضعیف به دلیل دانش کمتر و عدم حذف گزینه‌های نادرست خطرگریزتر هستند.

حال با توجه به این نتیجه می‌توان علت اینکه جریمه بهینه برای افراد ضعیف کمتر و برای افراد قوی بالاتر است را شرح داد؛ افراد قوی خطرپذیرتر هستند و این موجب حدس

بیشتر در تعداد گزینه‌های کمتر شده و در نتیجه خطرپذیری این افراد موجب می‌شود تا نمره این افراد بیشتر از توانایی واقعیشان به دست آید.

پس باید جریمه افراد قوی که خطرپذیری بالایی دارند بیشتر باشد تا نمره به دست آمده آن‌ها مطابق با دانش واقعیشان باشد. افراد ضعیف خطرگریزتر هستند و اگر میزان جریمه برای آن‌ها خیلی بالا باشد خطرگریزی آن‌ها موجب می‌شود تا سؤال‌های بیشتری را پاسخ ندهند و در نتیجه نمره این افراد نسبت به دانش واقعیشان بیشتر کاهش یافته و رابطه بین نمره و توانایی کاهش می‌یابد.

همچنین میزان جریمه نباید خیلی هم کم (کمتر از ۰/۳۳) باشد، چون در این صورت با توجه به جریمه کم خطرگریزی این افراد کمتر شده و بیشتر حدس می‌زنند که ممکن است موجب شود تا نمره این افراد بیشتر از توانایی واقعیشان به دست بیاید. به‌طور کلی میزان جریمه بهینه برای کل افراد در این سه آزمون بالاتر از میزان مرسوم آن (یک سوم در سؤال‌های چهارگزینه‌ای) ۰/۴۰ است.

به بیان دیگر، آنچه از جدول ۱۰ استنباط می‌شود این است جریمه بهینه برای همه سطوح توانایی بالاتر از میزان مرسوم آن ( $P = \frac{1}{K-1}$ ) است. مثلاً در سؤال‌های چهارگزینه‌ای که جریمه معمول ۰/۳۳ است، نتایج این پژوهش نشان می‌دهد که جریمه بهینه برای به دست آوردن رابطه بیشینه بین نمره خام و توانایی بالاتر از ۰/۳۳ است. برای مثال برای کل افراد در هر سه آزمون جریمه بهینه حدود ۰/۴۰ تا ۰/۸۰ است. این نتایج با نتایج اسپینوزا و گاردیوزیل (۲۰۱۰) که میزان جریمه بهینه را بالاتر از  $\frac{1}{K-1}$  (۰/۳۳ تا ۱/۰۰) عنوان کرده‌اند، سازگار است.

جدول ۱۰: مقادیر همبستگی بین توانایی متداول و نمره خام اصلاح شده (RAS0.33) در سطوح جریمه بهینه در سطوح مختلف توانایی در سه آزمون

گروه	آزمون	سطح جریمه								
		۰/۴۰	۰/۶۰	۰/۸۰	۱/۰۰	۱/۲۰	۱/۴۰	۱/۶۰	۱/۸۰	۲/۰۰
کل افراد	ریاضی	۰/۷۶	۰/۸۱	۰/۸۳	-	-	-	-	-	-
	معارف	۰/۸۶	۰/۹۰	۰/۹۲	۰/۹۳	۰/۹۴	-	-	-	-
	زبان	-	۰/۸۵	۰/۸۸	۰/۹۰	۰/۹۱	۰/۹۲	-	-	-
ضعیف	ریاضی	-	۰/۷۸	-	-	-	-	-	-	-
	معارف	-	-	۰/۷۱	-	-	-	-	-	-
	زبان	-	-	۰/۷۱	-	-	-	-	-	-
متوسط	ریاضی	-	-	۰/۵۳	۰/۵۵	-	-	-	-	-



معارف از دو آزمون دیگر بیشتر است. بر این اساس با توجه به میانگین ضریب دشواری آزمون‌ها، می‌توان نتیجه گرفت که هر چه میانگین ضریب دشواری آزمون نزدیک به صفر (متوسط) باشد، همبستگی خطی بین توانایی و نمره خام بیشینه می‌گردد و هر چه میانگین ضریب دشواری از صفر فاصله داشته باشد، از مقدار همبستگی خطی بین توانایی و نمره خام کاسته می‌شود. اسپینوزا و گاردیوزبل (۲۰۱۰) نیز به این نتیجه رسیدند که با میانگین دشواری صفر رابطه بین نمره و دانش افراد بیشینه می‌شود.

### بحث و نتیجه‌گیری

براساس همبستگی‌های بررسی شده رابطه بین توانایی و خطرپذیری در گروه ضعیف منفی ولی در گروه قوی مثبت است که نشان می‌دهد افراد ضعیف خطرگریزتر و افراد قوی خطرپذیرتر هستند. علت احتمالی آن است که افراد با توانایی پایین معمولاً نمی‌توانند گزینه‌های نادرست زیادی را حذف کنند و در نتیجه کمتر حدس می‌زنند ولی افراد با توانایی بالا معمولاً با استفاده از دانشی که دارند تعدادی از گزینه‌های نادرست در هر سؤال را حذف کرده و در نتیجه در میان گزینه‌های باقی مانده بیشتر حدس می‌زنند. همچنین با توجه به میانگین توانایی متداول هر نگرش خطرپذیری در هر سطح توانایی مشخص می‌شود که در گروه ضعیف میانگین نمره افراد خطرگریز و در گروه قوی میانگین نمره افراد خطرپذیر بیشتر است؛ یعنی در افراد ضعیف خطرگریزی و در افراد قوی خطرپذیری باعث افزایش نمره آن‌ها می‌شود. براساس نتایج پژوهش اسپینوزا و گاردیوزبل (۲۰۱۰) بیزاری از خطر موجب می‌شود تا تعداد عدم پاسخ‌های افراد بیشتر شده و نمره افراد کاهش می‌یابد.

همچنین بررسی تأثیر سطوح مختلف جریمه بر رابطه بین توانایی و نمره افراد نشان می‌دهد که عدم جریمه یا جریمه خیلی کم که باعث حدس زدن بیشتر می‌شود موجب کاهش همبستگی بین توانایی و نمره افراد می‌شود و اعمال جریمه خیلی بالا نیز موجب کاهش همبستگی بین توانایی و نمره خام افراد می‌شود ولی میزان این جریمه بالا در هر گروه توانایی با توجه به نگرش خطرپذیری آن‌ها متفاوت است. براساس نتایج پژوهش‌های گذشته نیز حدس زدن (شرایط عدم جریمه) موجب خطای اندازه‌گیری در نمره‌ها می‌شود و نمره افراد براساس دانش واقعی آن‌ها برآورد نمی‌شود و به این ترتیب موجب کاهش پایایی آزمون می‌شود (فراری ۱۹۸۸؛ بورتون ۲۰۰۱ و کمپل ۲۰۱۵). همچنین براساس نتایج



پژوهش‌های گذشته اصلاح حدس موجب کاهش خطای اندازه‌گیری ناشی از حدس و افزایش پایایی آزمون می‌شود. البته اصلاح حدس استفاده شده در پژوهش‌های گذشته به مقدار مرسوم آن (جریمه یک‌سوم برای سؤال‌های چهارگزینه‌ای) بوده است (لرد، ۱۹۶۴؛ بورتون، ۲۰۰۲ و کمپل، ۲۰۱۵).

با توجه به نتایج این پژوهش برای دستیابی به رابطه بیشینه بین توانایی و نمره خام افراد میزان جریمه افراد ضعیفی که نگرش خطرگریزی دارند کمتر و میزان جریمه افراد قوی که نگرش خطرپذیری دارند باید بیشتر باشد. اسپینوزا و گاردیوزیل (۲۰۱۰) میزان جریمه بهینه را بالاتر از  $\frac{1}{K-1}$  (۰/۳۳ تا ۱/۰۰) بیان کرده‌اند. همچنین با بررسی تأثیر دشواری آزمون‌ها بر میزان جریمه بهینه و مقدار همبستگی بین توانایی و نمره خام در سه آزمون با دشواری متفاوت مشخص شد که هر چه دشواری آزمون بیشتر باشد میزان جریمه بهینه باید کمتر باشد و بالعکس. همچنین مقادیر همبستگی به دست آمده بین توانایی متداول با نمره خام اصلاح شده در سطوح مختلف جریمه در این سه آزمون در آزمون معارف که دارای دشواری متوسط بود بیشینه بود. با توجه به این نتایج برای به دست آمدن همبستگی بیشینه بین توانایی و نمره خام باید دشواری آزمون متوسط (نزدیک صفر) باشد. اسپینوزا و گاردیوزیل (۲۰۱۰) نیز به این نتیجه رسیدند که با میانگین دشواری صفر رابطه بین نمره و دانش افراد بیشینه می‌شود. در این پژوهش به دلیل محدودیت زمانی امکان اجرای تحلیل‌ها بر روی آزمون‌های گروه‌های تجربی و انسانی و نیز تحلیل بر روی آزمون‌های بیشتر وجود نداشت و به همین دلیل در تعمیم یافته‌ها باید احتیاط کرد. با توجه به نتایج این پژوهش، برای به دست آوردن رابطه بیشینه بین توانایی و نمره خام افراد در آزمون‌هایی همانند آزمون سراسری پیشنهادی زیر توصیه می‌شود با توجه به نتایج جدول ۱۰، به‌طور کلی میزان جریمه باید بیشتر از ۰/۳۳ (حداقل ۰/۴۰) باشد. همچنین میزان جریمه باید با توجه به سطوح توانایی و همچنین نگرش خطرپذیری افراد متفاوت باشد. برای این منظور می‌توان قبل از اجرای آزمون‌ها نگرش خطرپذیری افراد را با استفاده از پرسشنامه‌های نگرش سنجی و توانایی افراد را با توجه به میانگین نمره‌های آزمون‌های استاندارد دوران تحصیلی مشخص کرد و برای هر سطح توانایی جریمه بهینه متناسب با نگرش خطرپذیری اعمال کرد. برای بیشینه‌سازی رابطه بین توانایی و نمره در آزمون‌ها، میانگین دشواری آزمون‌ها باید نزدیک به صفر باشد، درحالی‌که در آزمون‌های سراسری معمولاً میزان دشواری

سؤال‌ها بیشتر از توانایی افراد است. همچنین می‌توان با استفاده از داده‌های آزمون‌های پیشرفت تحصیلی که در آموزش و پرورش اجرا می‌شود، این پژوهش را اجرا کرد و نتایج آن را با نتایج پژوهش حاصل از آزمون‌های کنکور سراسری مقایسه شود. تا بتوان تفاوت‌های موجود در نمرات افراد در دو نوع آزمون و میزان جریمه‌های بهینه را با یکدیگر مقایسه کرد.

### منابع

- Bar-Hillel, M., Budescu, D., & Attali, Y. (2005). «Scoring and keying multiple choice tests: A case study in irrationality», *Mind and Society*, 4(1), 3-12.
- Burton, R. F. (2002). «Misinformation, partial knowledge and guessing in true/false tests», *Blackwell Science Ltd medical education*, 36(9), 805-811.
- Campbell, M. L. (2015). «Multiple-Choice Exams and Guessing: Results from a One-Year Study of General Chemistry Tests Designed To Discourage Guessing», *Journal of Chemical Education*, 92(7), 1194-1200.
- Chalmers, R. P. (2012). «Mirt: A multidimensional item response theory package for the R environment», *Journal of Statistical Software*, 48(6), 1-29.
- Choppin BH. (1988). *Correction for guessing. In: Keeves J. P. (ed) Educational research, methodology, and measurement: an international handbook.* (384-6). Oxford: Pergamon Press.
- Cureton, E. E. (1966). «The Correction for Guessing», *Journal of Experimental Education*, 34(4), 44-47.
- Davis, F. B. (1967). «A note on the correction for chance success», *Journal of Experimental Education*, 35(3), 42-47.
- Diamond, J., & Evans, W. (1973). «The correction for guessing», *Review of Educational Research*, 43(2), 181-191.
- Ebel, R. (1965). *Measuring Educational Achievement*, New Jersey, Prentice Hall.
- Edgington, E. S. (1965). «Scoring formulas that correct for guessing», *Journal of Experimental Education*, 33(4), 345-346.
- Espinosa, M. P., Gardeazabal, J. (2010). «Optimal correction for guessing in multiple-choice tests», *Journal of Mathematical Psychology*, 54(5), 415-425.
- Frary, R. B. (1988). «Formula Scoring of Multiple-Choice Tests (Correction for Guessing)», *Educational Measurement: Issues and Practice*, 7(2), 33-38.
- Frary, R. B., Cross L. H. & Lowry, S. R. (1977). «Random Guessing, Correction for Guessing, and Reliability of Multiple-Choice Test Scores», *The Journal of Experimental Education*, 46(1), 11-15.

- Little, E. B. (1966). «Overcorrection and undercorrection in multiple-choice test scoring», *Journal of Experimental Education*, 35(1), 44-47.
- Lord, F. M. (1964). «The Effect of Random Guessing on Test Validity», *Educational and Psychological Measurement*, 24(4), 745-747
- Lord, F. M. (1975). «Formula scoring and number-right scoring», *Journal of Educational Measurement*, 12(1), 7-11.
- Mehrens, W. A. & Lehmann, i. J. (1984). *Measurement and Evaluation in Education and Psychology*, 3 Edition, New York, CBS College Publishing.
- Quereshi, M. Y. (1974). «Performance on Multiple Choice Tests and Penalty for Guessing», *The Journal of Experimental Education*, 42(3), 74-77.
- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Revelle, W. (2018). psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA.
- Senel, S., Pehlivan, E. B. and Alatl B. (2015). «Effect of Correction-for-Guessing Formula on Psychometric Characteristics of Test», *Procedia - Social and Behavioral Sciences* 191, 925 – 929.
- Sherriffs, A. C. and Boomer, D. S. (1954). «Who Is Penalized by the Penalty for Guessing?», *The Journal of Educational Psychology*, 45(2), 81-90.