

## طراحی و ساخت آزمون پیشرفت تحصیلی درس مطالعات اجتماعی پایه‌ی اول دبیرستان بر اساس مدل لجستیک پرسش - پاسخ

عبدالمجید اورنگی\*<sup>۱</sup> و کمال درانی<sup>۲</sup>

### چکیده

هدف از این ساخت، مقیاس پیشرفت تحصیلی درس مطالعات اجتماعی پایه‌ی اول دبیرستان با استفاده از نظریه‌ی پرسش- پاسخ<sup>۳</sup> بود. نمونه‌ی مورد مطالعه در این پژوهش ۳۲۱ نفر از دانش‌آموزان دختر و پسر پایه‌ی اول دبیرستان شهر تهران بودند که از روش نمونه‌گیری خوشه‌ای چند مرحله‌ای تصادفی انتخاب شدند. در تهیه‌ی این آزمون، ابتدا دو فرم همتای آزمون از نوع چند گزینه‌ای با استفاده از جدول مشخصات که یک بعد آن را هدف‌های آموزشی (رفتاری) و بعد دیگر را محتوای درسی پوشش می‌داد، ساخته شد. سپس هر دو فرم در دو مرحله‌ی تجربی (برای تعیین ضریب دشواری پرسش، تعیین ضریب تمیز پرسش، همبستگی پرسش با نمره‌ی کل آزمون، همبستگی میان پرسش‌های آزمون، تحلیل گزینه‌های انحرافی) و نهایی در مورد گروه نمونه با فاصله‌ی زمانی یک هفته از یکدیگر اجرا شدند. نتایج بدست آمده از این پژوهش حاکی از این بود که: آزمون‌های ساخته شده‌ی فرم الف و ب از روایی و پایایی قابل قبولی برخوردارند و با مدل سه پارامتری نظریه‌ی پرسش - پاسخ برازندگی دارند.

واژه‌های کلیدی: ضریب دشواری، تمیز، روایی، پایایی، مدل سه پارامتری نظریه‌ی پرسش - پاسخ.

۱- دکترای مدیریت آموزشی و عضو هیئت علمی گروه پژوهشی فناوری آموزشی راهبر owmajid@gmail.com

۲- عضو هیئت علمی دانشگاه تهران

## مقدمه

شیوه‌های مورد استفاده در سنجش و اندازه‌گیری مشخصه‌های انسانی، مبتنی بر دو نظریه اساسی "روان‌سنجی"<sup>۱</sup> و "طرح و هیئت کلی"<sup>۲</sup> هستند. نظریه‌ی روان‌سنجی درصد برآورد جنبه‌های خاصی از ویژگی‌های افراد از قبیل، استعداد، توانایی و پیشرفت آدمی براساس مقادیر کمی و عددی است. این نظریه که ملهم از باورهای ثرن‌دایک (شرتزر، استون<sup>۳</sup>، ۱۹۷۱، ترجمه‌ی شریفی، ۱۳۷۳) می‌باشد، بر این باور است که "اگر چیزی وجود داشته باشد دارای کمیت است و آنچه دارای کمیت است، قابل اندازه‌گیری است". نظریه‌ی طرح و هیئت کلی مترصد بدست دادن تصویری جامع و توصیف کننده از خصایص فرد، برای دست‌یابی به شناختی پویا از آدمی است. این نظریه که در پی جستجوی "کلیت و وحدت" آدمی است، بیش‌تر بر روش‌های مشاهده‌ی رفتار، داده‌های توصیفی و گزارش‌های شخصی مبتنی است. در حالی که روان‌سنجی بر روش‌های مشخص و سازمان یافته استوار است (همان).

پیدایش آزمون‌های میزان شده برای اندازه‌گیری نتایج فعالیت‌های تحصیلی با الهام از نظریه‌ی روان‌سنجی و به پیروی از کوشش‌های اولیه‌ی ال ثرن‌دایک<sup>۴</sup> به نخستین سال‌های قرن بیستم بر می‌گردد. در این آزمون‌ها اصول اندازه‌گیری حاصل از پژوهش‌های آزمایشگاهی مورد استفاده قرار گرفت. بعدها مجموعه آزمون‌های پیشرفت تحصیلی<sup>۵</sup> به میان آمد که آزمون پیشرفت تحصیلی استفورد در سال (۱۹۲۳) نخستین چاپ آن‌ها بود. به همان نسبت که کاربرد ماده‌های نوع جدید در آزمون‌های پیشرفت تحصیلی گسترش می‌یافت، در شکل این ماده‌ها نیز تغییراتی داده شد تا سنجش میزان فهم و کاربرد داده‌های درسی و سایر هدف‌های وسیع تربیتی با آن‌ها امکان‌پذیر شود (آناستازی، ۱۹۷۶).

این نوع از آزمون‌ها که برای اندازه‌گیری بازده فعالیت‌های آموزشی و یا آموخته‌های کلاسی دانش‌آموزان در دروس گوناگون (سپاسی، ۱۳۷۵) یا سنجش عملکرد یادگیرندگان و مقایسه‌ی نتایج بدست آمده با هدف‌های آموزشی از پیش تعیین شده، بمنظور تصمیم‌گیری درباره‌ی میزان مطلوبیت نتایج فعالیت‌های آموزشی معلم و کوشش‌های یادگیری دانش‌آموزان (سیف، ۱۳۷۵) بکار می‌روند، شاخصی از آموخته‌های گذشته و وسیله‌ای برای پیش‌بینی موفقیت آینده‌اند.

1 -Sychometric

2 -Holistic

3- Shertzer e

4 E.L.Thorndike

5 -Achivement Test

حدود تأثیرگذاری آزمون‌های پیشرفت تحصیلی بر بهبود یادگیری و تدریس، اساساً ناشی از اصول حاکم بر تهیه، توسعه و کاربرد این آزمون‌هاست. بدین معنا که می‌توانند توجه شاگرد را به هدف‌های تدریس جلب کنند یا از آن دور سازند، دانش‌آموزان را ترغیب کنند که توجهشان به جنبه محدودی از محتوای درس و یا به تمام حوزه‌ی مهم آن معطوف شود. به یادگیری سطحی و یا به فهم عمیق مطالب بپردازند. داده‌هایی مناسب برای تصمیم‌گیری در امر تدریس یا اطلاعاتی مخدوش و دارای سوگیری ارایه نمایند (گران‌لند<sup>۱</sup>، ۱۹۸۵). بنابراین، در ساخت این‌گونه آزمون‌ها رعایت اصولی مانند: اندازه‌گیری نتایج مشخص یادگیری، نمونه‌ای معرف از تکالیف یادگیری، انتخاب مناسب‌ترین نوع پرسش، تدوین متناسب با کاربردهای ویژه برای مقاصد گوناگون، برخورداری از بیش‌ترین پایایی و انعکاس هدف‌های تدریس از ضروریات اساسی تهیه‌ی این‌گونه از آزمون‌هاست (لطف‌آبادی، ۱۳۷۵).

یکی از انواع آزمون‌های پیشرفت تحصیلی آزمون‌های دروس ویژه می‌باشند. این نوع آزمون‌ها از قبیل آزمون‌های علوم تجربی، ریاضیات، زبان و ... که برای اندازه‌گیری آموخته‌های دانش‌آموزان در یک درس و یا برنامه‌ی آموزشی ویژه تهیه و استاندارد می‌شوند، ابزار مفیدی برای ارزشیابی سطح معلومات و نقاط قوت و ضعف دانش‌آموزان کلاس در مقایسه با سطح محلی، منطقه‌ای یا ملی بشمار می‌آیند. تحلیل پرسش‌های این‌گونه آزمون‌ها معلمان را کمک می‌کند تا هدف‌های تدریس خود را بهتر بشناسند و یا آن دسته از هدف‌هایی را که در عمل مورد توجه آن‌ها نبوده است، مدنظر قرار دهند، کار گذشته را اصلاح نمایند، انتظارات خود را از دانش‌آموزان روشن نمایند، یادگیری دانش‌آموزان را بهبود بخشند و توجه خود را به برنامه درسی متمرکز کنند (دفتر همکاری‌های علمی بین‌المللی وزارت آموزش و پرورش، ۱۳۷۹). این نوع آزمون‌ها به عنوان مهم‌ترین وسیله‌ی اندازه‌گیری پیشرفت تحصیلی می‌توان به دو دسته عینی<sup>۲</sup> و غیر عینی<sup>۳</sup> یا ذهنی تقسیم نمود. آزمون‌های چند گزینه‌ای<sup>۴</sup> یکی از متداول‌ترین آزمون‌های عینی هستند که هم در سنجش توانایی‌های گوناگون و هم در سنجش یادگیری در زمینه‌های گوناگون تحصیلی مورد استفاده واقع می‌شوند. این نوع از آزمون‌ها از لحاظ یکنواختی پرسش‌ها، حساسیت کم در مقابل حدس کورکورانه و سهولت در تصحیح پاسخنامه‌ها بهترین نوع آزمون‌های عینی هستند (سیف، ۱۳۷۵).

1- Gronlund

2 - Objective

3 - Subjective

۴- این پژوهش نیز ناظر بر ساخت این نوع از آزمون‌ها می‌باشد.

در تجزیه، تحلیل و تفسیر نمره‌های این‌گونه از آزمون‌ها دو نظریه مورد استفاده قرار می‌گیرد. یکی نظریه "کلاسیک" اندازه‌گیری و دیگری نظریه "پرسش- پاسخ" می‌باشد. نظریه کلاسیک اندازه‌گیری با ارایه‌ی یک مدل مفهومی و آماری بمنظور تجزیه، تحلیل و تفسیر نمره‌های آزمون‌ها تلاش می‌کند تا استنباط و برداشت آزمونگر را از نمره‌ی آزمودنی توسعه دهد و این بینش را به آزمونگر لقا نماید که ورای نمره‌ی مشاهده شده آزمون و جدای از توان فرد برای پاسخگویی به پرسش‌های آن، عوامل دیگری را مدنظر قرار داده و در تعبیر، تفسیر و تصمیم‌گیری خود، پیرامون سطح صفت آزمودنی دخالت دهد. بر اساس تئوری کلاسیک هر نمره‌ی مشاهده شده‌ای از دو جزء جمع‌پذیر تشکیل شده است که عبارتند از: نمره‌ی حقیقی و نمره‌ی خطا. به بیان دیگر، با اجرای یک آزمون در مورد یک فرد یا گروهی از افراد، نمره‌هایی برای آن‌ها بدست می‌آید که اگرچه مطلوب آزمونگران است و آنان این نمرات را معرف میزان حقیقی صفت یا توانایی مورد سنجش می‌دانند، اما به علت یک سری عوامل دخیل و گریز ناپذیر نمره‌ی بدست آمده از اجرای آزمون به خودی خود، گویای میزان حقیقی صفت یا بیانگر جایگاه حقیقی فرد در پیوستار خصیصه نبوده و نمره‌ی مشاهده شده‌ای بیش نیست (مهرنز و لهما، ۱۹۸۶). این نظریه با وجود قابلیت‌هایی که دارد، ضعف‌هایی نیز بر آن مترتب است که اهم آن‌ها عبارتند از: به دلیل ضعف در برخورداری از یک بنای نظری محکم، برآورد پارامترهای جامعه از روی شاخص‌های پرسش دقت لازم را ندارد. سطح اندازه‌گیری هیچ‌گاه به مقیاس فاصله‌ای نمی‌رسد، یعنی واحدها یکسان نیست و در صورتی به واحدهای یکسان می‌رسیم که پرسش‌های محتوایی نسبتاً یکسانی را اندازه بگیرند و از دشواری یکسانی برخوردار باشند - مفروضه‌های این تئوری (نمره‌ی حقیقی و خطا) قابل مشاهده و محاسبه نیستند. مثلاً واریانس خطاهای همه‌ی اعداد یا توانایی‌های متفاوت، برای همه یکسان در نظر گرفت می‌شوند. از راه کاربرد این نظریه نمی‌توان فهمید که چه توانایی‌هایی در عملکرد فرد در پاسخ دادن به پرسش تأثیر داشته است. نظر به این‌که این تئوری از مدل ریاضی خاصی گرفته نشده، طبعاً ملاکی برای ارزشیابی ندارد (سلیمی‌زاده، ۱۳۷۸؛ سپاسی، ۱۳۷۵).

نظریه‌ی پرسش - پاسخ که به وسیله‌ی لرد برای تشخیص طبقه‌ای از مدل‌های صفت مکنون در ارتباط با پرسش‌های دو ارزشی ارایه گردیده (هومن، ۱۳۷۲)، با الهام از اندیشه‌ی منطقی آلفرد بینه (لین و گرانلاند، ۱۹۹۰) بر مرتبط بودن، افزایش نسبت کودکانی که پاسخ درست یک پرسش را می‌دانند با افزایش سن، از واژه‌ی توانایی استفاده نموده و بر این باور است که با افزایش توانایی، احتمال پاسخ درست دادن به پرسش نیز افزایش می‌یابد. با این وصف می‌توان گفت که نظریه‌ی پرسش - پاسخ بر نظریه یا الگوی صفت مکنون استوار است (آلن، ۱۹۸۹). در این نظریه به جای تأکید بر نمره‌های کل آزمون، بر پاسخ‌های داده شده به تک تک پرسش‌های آزمون توجه می‌شود و

با بهره‌گیری از مدل‌های پیچیده‌ی ریاضی می‌توان به معادله‌ای دست یافت که از آن راه می‌توان احتمال پاسخ درست به یک پرسش آزمون را تابعی از توانایی آزمون شونده تلقی نمود. ضمن آن‌که پیش‌بینی برخی از ویژگی‌های پرسش جهت داوری در بازه‌ی پرسش، طراحی آزمون برای مقاصد ویژه و حل برخی از مسایل اندازه‌گیری آموزشی که نظریه‌ی کلاسیک ناتوان از حل آن‌هاست را نیز موجب می‌گردد (گرانلاندولین، ۱۹۸۵).

ویژگی‌های اصلی این نظریه عبارتست از: عدم وابستگی به پارامترهای پرسش (دشواری و تمیز) با گروه مورد آزمون، عدم وابستگی نمره‌هایی که بیانگر توانایی آزمودنی‌ها هستند با نمره‌ی کل آزمون و آزمون‌های هم‌تا برای برآورد پایایی، توجه به ویژگی‌های تک تک پرسش‌های آزمون و رای‌ی یک الگوی دقیق برای برآورد توانایی آزمودنی (همیلتون و سوامی ناتان، ۱۹۸۵).

در بررسی پیشینه‌ی مربوط به آزمون‌های پیشرفت تحصیلی با استفاده از نظریه‌ی پرسش - پاسخ در ایران می‌توان به پژوهش مشتاقیان (۱۳۷۸) جهت ساخت و هنجاریابی آزمون پیشرفت تحصیلی علوم تجربی، آزمون‌ی محتوی ۹۰ پرسش در دو فرم الف و ب (هرکدام حاوی ۴۵ پرسش) را در مورد گروهی از دانش‌آموزان پایه‌ی سوم راهنمایی اجرا نمود. نتایج تحلیل آزمون ذکر شده بر اساس مدل پرسش - پاسخ نشان داد که پرسش‌ها و هم‌چنین کل آزمون با مدل لجستیکی سه پارامتری در سطح ( $P < .001$ ) برازش دارند.

همان‌گونه که مشاهده می‌شود، با وجود محدودیت‌هایی که به نظریه‌ی کلاسیک وارد است، تعداد محدودی از پژوهش‌های انجام شده در داخل کشور از نظریه‌ی پرسش - پاسخ استفاده نموده‌اند. به همین منظور، در این مطالعه سعی شده است فرایند استفاده از نظریه‌ی پرسش - پاسخ در ساخت آزمون پیشرفت تحصیلی درس مطالعات اجتماعی برای دانش‌آموزان پایه‌ی اول دبیرستان تبیین شود.

### روش پژوهش

گروه نمونه‌ی مورد مطالعه در این پژوهش ۳۲۴ نفر از دانش‌آموزان دختر و پسر (به نسبت مساوی) پایه‌ی اول دبیرستان‌های مدارس دولتی روزانه‌ی شهر تهران بودند که به صورت نمونه‌گیری خوشه‌ای چند مرحله‌ای انتخاب شدند. برای ساخت آزمون پیشرفت تحصیلی، ابتدا هدف‌های کلی محتوایی و هدف‌های آموزشی (رفتاری) کلیه‌ی فصول درس مطالعات اجتماعی پایه‌ی اول دبیرستان با استفاده از جدول مشخصات استخراج و طبقه‌بندی شدند. سپس با استفاده از آن‌ها دو فرم آزمون هم‌تا با استفاده از حجم مطالب، زمان صرف شده برای رای‌ی هر مطلب و تعداد صفحات اختصاص داده شده به هر مطلب، ساخته شد (فرم الف و ب، هر کدام دارای ۱۱۴

پرسش). پس از بررسی دو فرم آزمون ساخته شده از نظر: محتوا و هدف‌های آموزشی، ابهام‌های احتمالی در متن پرسش و گزینه‌ها، میزان شباهت با موضوع مورد اندازه‌گیری و وحدت در اندازه‌گیری اهداف آموزشی با یکدیگر، برای هر دو فرم یک راهکار واحد تهیه گردید. آزمون‌های ساخته شده از راه نمونه‌گیری خوشه‌ای چند مرحله‌ای از کلیه دانش‌آموزان دختر و پسر پایه‌ی اول دبیرستان‌های مدارس دولتی روزانه‌ی شهر تهران، در دو مرحله‌ی تجربی و نهایی (بر روی گروه نمونه)، به فاصله‌ی زمانی یک هفته به اجرا درآمد. سپس نتایج بدست آمده از این مطالعه با استفاده از نظریه‌ی کلاسیک (شامل ضریب دشواری، ضریب تمیز، همبستگی پرسش‌ها، تحلیل گزینه‌های انحرافی و بدست آوردن پایایی و اعتبار آزمون) و مدل لجستیک پرسش - پاسخ و منحنی‌های ویژه و آگاهی پرسش مورد تجزیه و تحلیل قرار گرفت. برای تحلیل داده‌ها از نرم افزار پارامتری پل<sup>۱</sup> استفاده شده است.<sup>۲</sup>

## یافته‌ها

نتایج بدست آمده از این مطالعه نشان داد که: دو فرم آزمون ساخته شده‌ی الف و ب با توجه به نظرات دبیران با تجربه‌ی درس مطالعات از نظر میزان معرف محتوا بودن و تناسب با هدف‌های رفتاری تدوین شده با توجه به حوزه‌ی محتوایی درس مطالعات اجتماعی از روایی محتوایی نسبتاً مطلوبی برخوردار است. در این رابطه روی هم رفته، پرسش‌های تدوین شده‌ی هر دو فرم<sup>۳</sup> مورد تأیید قرار گرفتند. سپس آزمون‌های تهیه شده جهت آگاهی از ابهام در نگارش، درک اصطلاحات، رسا و گویا بودن راهنمای آزمون و ... در مورد ۳۰ نفر از دانش‌آموزان دختر (۱۵ نفر) و پسر (۱۵ نفر) اجرا و اصلاحات لازم در آن‌ها انجام گرفت. در اجرای فرم تجربی، آزمون الف با ۱۱۴ پرسش دامنه‌ای بین ۰/۶۹ + تا ۰/۱۴ - را از نظر میزان همبستگی هر پرسش با نمره‌ی کل آزمون، به خود اختصاص داد که در اصلاح نهایی با توجه به هدف‌های رفتاری تدوین شده و حیطه‌های شناختی، کم‌ترین ضریب همبستگی ۰/۴۱ برای ساخت فرم نهایی آزمون الف انتخاب گردید.<sup>۴</sup> آزمون فرم ب نیز مشتمل بر ۱۱۴ پرسش بود که ضرایب محاسبه شده از نظر میزان همبستگی هر پرسش با

1 - PARAM-3PL

2 - PARAM-3PL is a public domain, freeware tool for calibrating items and individuals using the 3 parameter logistic item response theory model. Parameterization uses Newton-Raphson maximum likelihood estimation as outlined in Lord (1980).

۳- به جز چند پرسش که از آزمون حذف شدند.

۴- پرسش‌های هر دو فرم آزمون در اجرای نهایی به ۵۵ پرسش کاهش یافت.

نمره‌ی کل آزمون برای این فرم نیز دامنه‌ای بین ۰/۶۵ تا ۰/۱۷- را به خود اختصاص می‌داد که در اصلاح نهایی پرسش‌های تدوین شده با ضریب همبستگی کمینه‌ی ۰/۳۶ انتخاب شدند.<sup>۱</sup>

در تأمین روایی ملاکی آزمون، هر دو فرم الف و ب تقریباً با فاصله‌ی ده روز در مورد یک گروه نمونه‌ی مورد مطالعه به اجرا درآمد. سپس همبستگی نمره‌های بدست آمده از این دو فرم با نمره‌های درس مطالعات اجتماعی خردادماه دانش‌آموزان شرکت کننده در آزمون محاسبه گردید. نتایج بدست آمده از این همبستگی برای فرم الف (۰/۴۶) و برای فرم ب (۰/۴۸) بدست آمد که هر دو در سطح کم‌تر از یک هزارم ( $p < .001$ ) معنی دار بودند.

جدول ۱- میزان همبستگی آزمون‌های فرم الف و ب با نمره‌های درس مطالعات اجتماعی دانش‌آموزان در خردادماه

نمونه‌ی فرم ب		نمونه‌ی فرم الف		رابطه‌ی آزمون‌ها و درس اجتماعی
میزان همبستگی	کوواریانس	میزان همبستگی	کوواریانس	
۰/۴۸	۱۳/۰۴	۰/۴۶	۱۳/۳۰	نمره‌ی درس اجتماعی
۰/۷۳	۱۳۲/۷	---	۲۰۶/۰۱	نمونه‌ی فرم الف
---	۱۷۳/۴۲	---	---	نمونه فرم ب

( $p < .001$ ) N=321

برای محاسبه‌ی پایایی آزمون‌ها، فرم‌های تهیه شده در دو بخش الف و ب (هر کدام دارای یکصد و چهارده پرسش) با فاصله‌ی زمانی ده روز بر روی یک گروه از دانش‌آموزان اجرا گردید. نتایج رابطه ( $r = .73$ ) را بین دو آزمون فرم الف و ب نشان داد که در سطح کم‌تر از یک هزارم معنی دار بود.

در تحلیل پرسش‌های آزمون فرم الف با مدل نظری لجستیکی سه پارامتری (قدرت تشخیص، سطح دشواری و عامل حدس) نتایج بیانگر آن است که تقریباً همه‌ی پرسش‌های این فرم (به جز پرسش‌های ۲۴، ۳۷ و ۵۵)<sup>۲</sup>، به ترتیب با درجه‌ی آزادی (۹، ۱۰ و ۱۴) و خی دو محاسبه شده ۱۷/۵، ۱۸/۷ و ۲۶/۶ در سطح کم‌تر از ( $p < .05$ ) معنی دار بوده و با الگوی نظریه‌ی پرسش - پاسخ برازندگی دارند.

<sup>۱</sup> - کمینه‌های در نظر گرفته شده بر هر یک از دو فرم آزمون در سطح یک هزارم معنی دار بودند.

<sup>۲</sup> - این سه پرسش به دلیل این‌که از نظر هدف آموزشی سطوح ادراک و دانش را مورد سنجش قرار می‌دهند و چنانچه از فهرست مجموعه آزمون حذف شوند، از نظر محتوایی به آزمون آسیب وارد می‌شود، بنابراین از حذف آن‌ها خودداری شده است. هم‌چنین، تحلیل نظریه‌ی کلاسیک در مورد این پرسش‌ها آن‌ها را تأیید می‌کند.

اندازه‌ی پارامتر ضریب تشخیص (a) در فرم الف، در همه‌ی موارد بالاتر از یک نشان داده شده است. دامنه‌ی آن از  $1/01$  برای پرسش ۲ با  $6\%$  خطای استاندارد به عنوان پایین‌ترین ضریب تشخیص تا  $3/31$  برای پرسش ۴۸ با  $0/18$  خطای استاندارد و به عنوان بالاترین ضریب تشخیص گسترده شده است. این بدان معناست که پرسش ۴۸ بیش‌ترین و پرسش ۲ کم‌ترین قدرت شیب را داراست. بر اساس پرسش‌های این فرم از آزمون، پیشرفت تحصیلی درس مطالعات اجتماعی از نظر میزان تشخیص با مدل الگوی پرسش - پاسخ در سطح مطلوبی قرار دارند. هم‌چنین، بر اساس نظریه‌ی پرسش - پاسخ تمامی پرسش‌های آزمون از سطح دشواری (b) متوسطی برخوردارند. به بیان دیگر، هیچ یک از پرسش‌ها نه آنچنان دشوار و سخت هستند که کسی نتوانسته باشد به آن پاسخ دهد و نه آن‌گونه ساده که همگی توانسته باشند به آن پاسخ درست بدهند. در این تحلیل، پرسش ۴۲ با انتخاب گزینه‌ی صحیح از سوی ۱۱۲ نفر از ۳۲۱ نفر به عنوان دشوارترین پرسش‌ها ( $b=1/099$ ) پرسش ۴ با انتخاب گزینه‌ی صحیح به وسیله‌ی ۲۵۷ نفر از ۳۲۱ نفر به عنوان ساده‌ترین پرسش آزمون ( $b=-1/154$ ) قلمداد می‌شوند که در مجموع مقایسه‌ی ضرایب بدست آمده برای پرسش‌های فرم الف آزمون با شاخص ضریب دشواری تعیین شده برای پرسش‌های چهارگزینه‌ای نشان می‌دهد که پرسش‌های این فرم از آزمون در مطلوب‌ترین سطح دشواری قرار دارند.

در تحلیل پارامتر شاخص حدس (c)، فرم الف آزمون نتایج نشان می‌دهد که پرسش ۴۸ با شاخص حدس ( $0/28$ ) در مقایسه با سایر پرسش‌ها احتمال انتخاب تصادفی گزینه‌ی صحیح آن از سوی افراد گروه ضعیف بیش‌تر است و پرسش ۲۶ با شاخص حدس ( $0/10$ ) در مقایسه با سایر پرسش‌های آزمون کم‌تر امکان دارد که پاسخ صحیح آن از روی حدس و تصادف به وسیله‌ی افراد گروه ضعیف انتخاب شود. بنابراین، می‌توان این‌گونه قضاوت کرد که تقریباً همه پرسش‌های فرم الف آزمون درس مطالعات اجتماعی براساس نظریه‌ی پرسش - پاسخ با الگوی IRT برازندگی دارد. این نتایج در جدول ۲ ارایه شده است.



جدول ۲- تجزیه و تحلیل پرسش‌های فرم الف بر اساس الگوی نظریه‌ی پرسش-پاسخ<sup>۱</sup>

احتمال برازش	پارامترهای پرسش‌های آزمون			سؤال‌ها
	c	b	a	
۰/۴۶۳	۰/۲۴۳	۱/۱۳۳	۱/۱۱۱	۱
۰/۴۰۶	۰/۲۴۱	۰/۵۲۵	۱/۰۱۰	۲
۰/۲۶۵	۰/۲۲۴	۰/۸۳۱	۱/۵۳۸	۳
۰/۷۹۷	۰/۱۹۶	۱/۱۵۴	۱/۶۰۵	۴
۰/۲۷۸	۰/۱۵۶	۰/۶۸۰	۱/۲۲۵	۵
۰/۷۰۳	۰/۱۶۹	۰/۰۳۳	۱/۵۱	۶
۰/۲۵۱	۰/۱۶۸	۰/۵۳۱	۱/۲۶۴	۷
۰/۱۲۵	۰/۱۵۱	۰/۴۹۴	۱/۴۶۸	۸
۰/۳۵۲	۰/۱۷۰	۰/۱۶۰	۲/۰۹۰	۹
۰/۳۳۰	۰/۱۷۰	۰/۲۰۹	۱/۳۷۱	۱۰

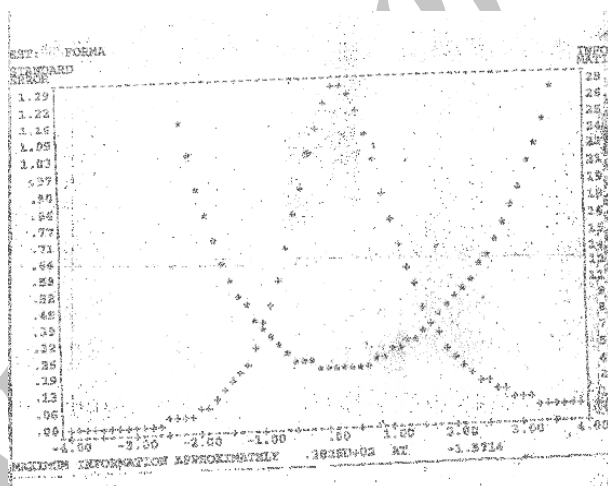
N = 321

تحلیل‌های انجام شده در رابطه با فرم ب آزمون نیز در خصوص محاسبه‌ی پارامترهای شیب (a)، دشواری (b) و حدس (c) نیز نشان می‌دهد که تقریباً همه‌ی پرسش‌های این فرم به جز پرسش‌های ۱۱، ۳۶ و ۵۴، با الگوی نظریه‌ی پرسش-پاسخ برازندگی دارند. در حالی که این سه پرسش، براساس نظریه‌ی کلاسیک از نظر همسانی درونی به ترتیب با ۰/۴۶، ۰/۴۹ و ۰/۴۴، ضریب دشواری با ۰/۶۱، ۰/۴۸ و ۰/۳۵ و همچنین، تحلیل گزینه‌های انحرافی در سطح نسبتاً مطلوبی قرار دارند. پارامتر شیب (a) یا ضریب تشخیص برآورد این فرم بیانگر این است که تمامی پرسش‌ها از ضریب تشخیص مناسبی برخوردارند و دامنه‌ی آن برای پرسش ۶ با ۰/۸۸ و ۰/۵، خطای استاندارد که در مجموعه آزمون به عنوان پایین‌ترین ضریب تشخیص و پرسش ۴۵ با ۳/۱۸ شیب و ۰/۲۰ خطای استاندارد به عنوان بالاترین ضریب تشخیص با تفاضل ۲/۳۱ گسترده شده است. پارامتر سطح دشواری (b)، نشان می‌دهد که تمامی پرسش‌های آزمون از سطح دشواری قابل قبولی برخوردارند. پرسش ۵۴ و ۳۸ به ترتیب با انتخاب گزینه‌های صحیح به وسیله‌ی ۱۱۵ و ۱۳۴ نفر در هر پرسش و ضرایب دشواری (۱/۲۰۶) به عنوان سخت‌ترین پرسش‌های آزمون و پرسش ۲ با انتخاب گزینه‌ی صحیح به وسیله‌ی ۲۵۸ نفر و ضریب دشواری (b = -۱/۲۳۳) به عنوان ساده‌ترین پرسش آزمون در فرم ب تعیین می‌شوند. پارامتر حدس (c)، در فرم ب آزمون، نشان می‌دهد که

<sup>۱</sup>- در این مقاله به دلیل حجم زیاد جدول، نتیجه‌ی تحلیل ۱۰ پرسش ارائه شده است. همچنین، جدول فرم ب نیز ارائه نشده است.

پرسش ۲۴ با شاخص حدس ۰/۳۳ در مقایسه با سایر پرسش‌ها، ممکن است بیش‌تر پاسخ صحیح آن از سوی افراد گروه ضعیف به گونه‌ی تصادفی انتخاب شود و پرسش ۵۳ با شاخص حدس (۰/۸۸) در مقایسه با دیگر پرسش‌های آزمون ممکن است کم‌تر پاسخ صحیح آن از سوی افراد گروه ضعیف به گونه‌ی تصادفی انتخاب شود.

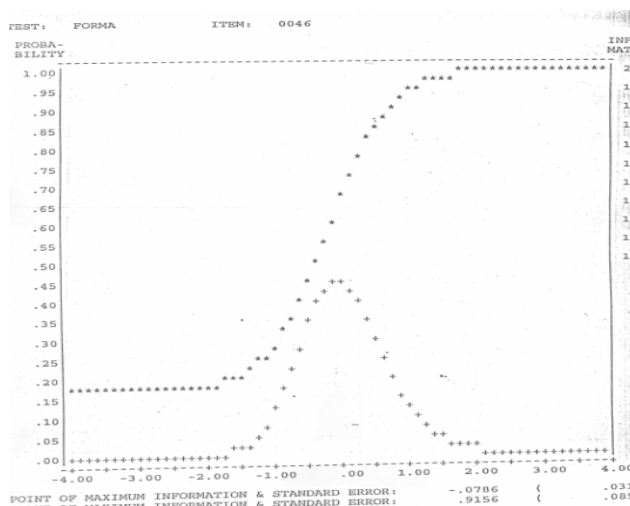
منحنی تابع آگاهی و خطای استاندارد توانایی آزمودنی‌های فرم الف و ب<sup>۱</sup>، همان‌گونه که در نمودار نشان داده شده است، در محدوده‌ی یک انحراف استاندارد از دو طرف، آگاهی مناسبی را فراهم می‌آورد، ولی میزان آگاهی‌دهندگی آن در کرانه‌های مقیاس توانایی ناچیز است. بیش‌ترین میزان آگاهی‌دهندگی در  $\theta = 0$  می‌باشد. در این شکل نمودار خطای اندازه‌گیری نیز همزمان ارایه شده است که با مقایسه‌ی نمودار آگاهی می‌توان دریافت که بین میزان آگاهی و خطای اندازه‌گیری رابطه‌ی معکوس وجود دارد. به گونه‌ای که با افزایش آگاهی میزان خطای اندازه‌گیری کاهش می‌یابد. به بیان دیگر، تقریباً مجموعه پرسش‌های آزمون (برای فرم الف ۵۰ پرسش و برای فرم ب ۴۵ پرسش) بیش‌ترین آگاهی را برای دانش‌آموزان با سطوح متوسط توانایی فراهم می‌سازند.



منحنی ۱- تابع آگاهی و خطای استاندارد توانایی آزمودنی‌های فرم الف

۱- به دلیل تشابه بسیار زیاد هر دو منحنی ترسیم شده برای فرم آزمون الف و ب، فقط منحنی فرم الف ارایه شده است.

برای نمایش رابطه‌ی بین پارامترهای پرسش در هر یک از سطوح توانایی، در این مقاله برای تنها یکی از پرسش‌های آزمون فرم الف منحنی ویژه و آگاهی پرسش ارائه شده است.



### منحنی ۲ - منحنی ویژه و آگاهی پرسش ۴۶ از فرم الف

همان‌گونه که مشاهده می‌شود، در تبیین پارامتر شیب (a) منحنی ۲ پرسش ۴۶ از فرم الف، شیب نسبتاً ملایمی دارد. به بیان دیگر، با افزایش سطح توانایی، احتمال پاسخ درست به این پرسش نیز به نسبت افزایش می‌یابد. در توضیح این مطلب، در  $\theta = -1/5$  احتمال پاسخ صحیح برابر تقریباً  $0/21$  و در  $\theta = +2$  همین احتمال برابر تقریباً  $0/98$  می‌باشد، یعنی افزایش  $3/5$  واحد در مقیاس توانایی،  $0/77$  افزایش در احتمال پاسخ درست می‌شود. هم‌چنین، این پرسش تا  $\theta = -1/6$  قادر به تمیز افراد قوی از ضعیف نیست، چرا که در این سطوح از توانایی ( $-4$  تا  $-1/6$ ) احتمال پاسخ صحیح تغییر نمی‌کند زیرا منحنی مربوط در این محدوده تخت می‌باشد، اما برای سطوح بالاتر به تدریج قدرت تمیز این پرسش افزایش می‌یابد. هم‌چنین، در توضیح پارامتر سطح دشواری (b) پرسش یاد شده مشاهده می‌شود که سطح دشواری پرسش مورد نظر برابر با شیب خط مماس بر بیشینه‌ی منحنی است که در این‌جا برای این پرسش در تتای برابر با  $-0/15$  قرار می‌گیرد.

احتمال موفقیت فردی با توانایی تتای کم‌تر از  $(\theta < -4)$  (پاسخ از روی حدس یا پارامتر حدس (c)) مطابق با منحنی یاد شده برابر  $c = 0/17$  می‌باشد. منحنی آگاهی پرسش ۴۶ فرم الف به

میزان ۰/۹ در تتای  $\theta = -0/15$  (که در آن شیب منحنی ویژه‌ی پرسش مربوط به بیشینه است)، بیش‌ترین آگاهی را بدست می‌دهد.

### بحث و نتیجه‌گیری

همان‌گونه که در بخش یافته‌ها نشان داده شد، در برخی شاخص‌ها مانند همبستگی پرسش‌های دو فرم آزمون از نظریه‌ی کلاسیک نیز استفاده شده است. بر اساس این نظریه، هر دو فرم آزمون در شاخص‌های مورد نظر از وضعیتی مناسب برخوردار بودند. در استفاده از مدل لجستیک پرسش- پاسخ کلیه‌ی پرسش‌های هر دو فرم آزمون (به جز سه پرسش، در فرم الف (۵۵، ۳۷ و ۲۴) و در فرم ب (۵۴ و ۳۶، ۱۱)) با این الگو برازندگی داشتند. منحنی توانایی هر دو فرم آزمون نشان داد که در محدوده‌ی یک انحراف استاندارد از دو طرف  $+1$  تا  $-1$  پرسش‌های آزمون، آگاهی مناسبی را فراهم می‌آورند که بیش‌ترین میزان آگاهی‌دهندگی در  $\theta = 0$  است.

روی هم رفته، هر دو فرم آزمون پیشرفت تحصیلی ساخته شده در درس مطالعات اجتماعی برای پایه‌ی اول دبیرستان از بیش‌ترین روایی و پایایی برخوردار بوده و با مدل لجستیکی سه پارامتری برازندگی دارند. بنابراین، دو فرم آزمون ساخته شده برای درس مطالعات اجتماعی پایه‌ی اول دبیرستان، به احتمال زیاد می‌توانند ملاکی مناسب برای اندازه‌گیری سطوح گوناگون یادگیری، پیشرفت تحصیلی و مقایسه‌ی ابعاد شناختی دانش آموزان در درس مطالعات اجتماعی بشمار می‌روند شوند. این آزمون‌ها در مقایسه با سایر آزمون‌های پیشرفت تحصیلی ساخته شده، بویژه در ایران به احتمال زیاد، از چند جهت از برتری نسبی برخوردار است.

- نخست این‌که این آزمون در رابطه با یکی از دروس علوم انسانی، صرف نظر از درس مطالعات اجتماعی، تدوین و ساخته شده است که در ایران به جز آزمون پیشرفت تحصیلی زبان فارسی در دوره‌ی ابتدایی و راهنمایی (شریفی، ۱۳۷۷) که در رابطه با دروس علوم انسانی ساخته شده است، در نوع خود کم‌نظیر است.

- دوم آن‌که در تجزیه و تحلیل این آزمون از الگوی نظریه‌ی پرسش - پاسخ استفاده شده است. این الگو که بر پایه‌ی صفت مکتون استوار است (لرد، ۱۹۸۰)، از نظر مبانی نظری و جنبه‌های عملی متفاوت و پیشرفته‌تر از روش کلاسیک است و همان‌گونه که در منحنی‌های خصیصه‌ی پرسش ترسیم شده، برای برخی از پرسش‌ها نشان داده شده است، پارامترهای شیب، دشواری و حدس مورد تجزیه و تحلیل قرار گرفته‌اند.

- از جمله ویژگی‌های دیگر این آزمون در مقایسه با سایر آزمون‌های پیشرفت تحصیلی ساخته شده (بویژه در ایران)، روایی و پایایی بسیار بالایی است که این آزمون (فرم الف و ب) بدست آورده است.

## منابع

- ۱- آموزش و پرورش، ک. د. (۱۳۷۹). مجموعه گفتارهای ارزشیابی در آموزش. تهران، ایران: دفتر همکاری‌های علمی بین‌المللی وزارت آموزش و پرورش چاپ اول.
- ۲- حسین، م (۱۳۷۸). ساخت و هنجاریابی آزمون پیشرفت تحصیلی علوم تجربی پایه سوم راهنمایی. پایان‌نامه کارشناسی ارشد رشته تحقیقات آموزشی. تهران، ایران. دانشکده روان‌شناسی و علوم تربیتی دانشگاه تهران.
- ۳- حیدرعلی، ه. (۱۳۷۲). اندازه‌گیری روانی و فن تهیه تست. تهران: زیبا.
- ۴- سپاسی، ح. (۱۳۷۵). نظریه سؤال-پاسخ در آزمون‌های روانی - تربیتی. مجله علوم تربیتی و روان‌شناسی دانشگاه چمران اهواز، شماره‌های سوم و چهارم.
- ۵- سرمد، بازرگان، حجازی، ز. ع. (۱۳۷۶). روش‌های تحقیق در علوم رفتاری. تهران. آگاه.
- ۶- سلیمی زاده، م. ک. (۱۳۷۸). مبانی نظری آزمون‌های روانی - تربیتی. جزوه درسی کارشناسی ارشد رشته سنجش و اندازه‌گیری. تهران، ایران. دانشگاه علامه طباطبایی.
- ۷- سیف، ع. ا. (۱۳۷۵). روش‌های اندازه‌گیری و ارزشیابی آموزشی. تهران، آگاه.
- ۸- شریفی، ح. پ. (۱۳۷۳). اصول روان‌سنجی و روان‌آزمایی. تهران: رشد.
- ۹- لطف آبادی، ح. (۱۳۷۵). سنجش و اندازه‌گیری در علوم تربیتی و روان‌شناسی. تهران: سمت.
- 10- Allen, J.M. (1989). *Introduction to measurement theory*. California . California : wadsworth.
- 11- Anastasi, A. (1976). *Psychological Testing*. (4 rd ed) NEW YORK: Mac Millan.
- 12- Grounlund, N.E. (1985). *How to construct achievement test*. Englewood cliffs , New Jersey : Prentic – Hill.
- 13- Hambleton, R.K. (1985). *Item Response Theory, principals and application*. Boston: Nighff publications.
- 14- Linn, R.L. (1990). *Measurment and Evaluation in Teaching (6th . ed)*. New York: Macmillan.
- 15- Lord, F. (1980). *Applications of item response theory to practical testing programs*. Hill sdale, nj: lawrence erbaun, Associates.
- 16- Mehrens, W.A. (1986). *Using standardized tests in education(4th ed)*. NEW YORK: Longman.