

سنجش مقایسه پذیری سازه و تحلیل کارکرد افتراقی سؤال‌ها (DIF) و بلوک‌های (DTF) آزمون علوم پایه هشتم تیمز ۲۰۰۷ در بین دانش‌آموزان ایران و آمریکا

اصغر مینایی^۱

چکیده

در سنجش‌های بین‌المللی مانند تیمز و پرلز لازم است تا آزمون‌های مورد استفاده از هم‌ارزی سازه که به مقایسه‌پذیری سازه معروف است برخوردار بوده و سؤال‌ها تا حد ممکن برای تمام کشورهای و گروه‌های مشارکت‌کننده، یکسان عمل نمایند. در پژوهش حاضر تلاش شد تا مقایسه‌پذیری سازه‌ای آزمون علوم پایه هشتم (سوم راهنمایی) تیمز ۲۰۰۷ و کارکرد افتراقی سؤال‌های آن در بین دانش‌آموزان ایران و آمریکا و همچنین میزان تأثیر سؤال‌های دارای کارکرد افتراقی بر عملکرد دانش‌آموزان ایرانی مورد بررسی قرار گیرد. برای تحلیل داده‌ها و پاسخگویی به سؤال‌های پژوهش از ترکیبی از رویکردهای تحلیل عاملی تأییدی و نظریه سؤال - پاسخ استفاده شد. نتایج حاصل از تحلیل عاملی تأییدی نشان می‌دهد که آزمون علوم از مقایسه‌پذیری سازه در بین دو گروه مورد مطالعه برخوردار است. این یافته بدان معنا است که دانش‌آموزان دو کشور ایران و آمریکا برای دادن پاسخ به سؤال‌های علوم از چارچوب مفهومی یکسانی استفاده می‌کنند. با این حال، نتایج حاصل از تحلیل کارکرد افتراقی سؤال‌ها نشان می‌دهد ۶۲٪ سؤال‌های درس علوم تیمز ۲۰۰۷ علیه دانش‌آموزان ایران از کارکرد افتراقی برخوردارند. به طور خلاصه، به نظر می‌رسد که عملکرد ضعیف دانش‌آموزان ایران در درس علوم تیمز ۲۰۰۷ را نمی‌توان به کارکرد افتراقی سؤال‌ها نسبت داد و بایستی علل آن را جایی دیگر جستجو کرد. مسئولان وزارت آموزش و پرورش و متولیان امر آموزش در کشور بایستی تلاش کنند تا با تهیه محتواهای آموزشی مناسب، مفاهیم مهم هر حوزه را به صورت زنجیره‌وار و در ارتباط با یکدیگر آموزش دهند و تلاش کنند تا توانایی و مهارت تفکر واگرا و چندبُعدی را در دانش‌آموزان پرورش دهند. همچنین بایستی تلاش شود تا از روش سنتی آموزش که به شکل سخنرانی و عمدتاً نظری صورت می‌گیرد فاصله گرفت و دانش‌آموزان را درگیر فعالیتهای عملی و آزمایشگاهی نمود.

واژگان کلیدی: تحلیل عاملی تأییدی، تغییرناپذیری عاملی، تیمز، کارکرد افتراقی سؤال، نظریه سؤال - پاسخ.

۱. هیأت علمی پژوهشگاه آموزش و پرورش. Email: as_minaei@yahoo.com

تاریخ پذیرش: ۱۳۹۲/۳/۲۵

تاریخ دریافت: ۱۳۹۲/۱/۱۸

مقدمه

سنجش‌های بین‌المللی، مانند TIMSS و PIRLS، به طور فزاینده‌ای مورد توجه وزارتخانه‌ها و سیاست‌گذاران آموزشی کشورهای مختلف قرار دارند. چنین سنجش‌هایی این امکان را فراهم می‌آورد تا دروندادها، فرایندها و پیشرفت تحصیلی کشورهای مختلف را مقایسه و چشم‌انداز وسیعی را برای ارزیابی و بهبود آموزش فراهم نمود (ارسی‌کان^۱، ۱۹۹۸). نمره‌های حاصل از این سنجش‌ها، هسته اصلی مقایسه عملکردهای کشورها، تصمیم‌گیری‌های مربوط به سیاست‌های آموزشی و پژوهش در زمینه نتایج ارزیابی را تشکیل می‌دهند.

معادل‌بودن یا هم‌ارزی^۲ سازه در میان گروه‌های مورد مطالعه، که به مقایسه‌پذیری سازه معروف است، شرط اصلی برای مطالعه تفاوت‌های گروهی است و روایی استنباط‌هایی که بر اساس نتایج سنجش‌های بین‌المللی به عمل می‌آید اساساً به مقایسه‌پذیری نمره‌ها در میان کشورها بستگی دارد.

آزمون تیمز یکی از مهم‌ترین سنجش‌های بین‌المللی است که توسط انجمن بین‌المللی ارزشیابی پیشرفت تحصیلی^۳ (IEA) برگزار می‌گردد. کشور جمهوری اسلامی ایران به منظور ارزیابی و بهبود نظام آموزشی خود از سال ۱۳۷۰ برابر با ۱۹۹۱ میلادی رسماً همکاری خود را با انجمن IEA آغاز کرد و تا کنون در ۵ مطالعه در فاصله سال‌های ۱۹۹۵، ۱۹۹۹، ۲۰۰۳، ۲۰۰۷، ۲۰۱۱ و تیمز پیشرفته ۲۰۰۸ شرکت کرده است (کریمی، ۱۳۸۸).

در سنجش‌های بین‌المللی، ایده آل این است که صرفاً از یک فرم مشترک و یکسان آزمون استفاده شود. با این حال، از آنجا که آزمون در کشورهای مختلف اجرا می‌شود، لازم است به زبان کشورهای مشارکت‌کننده ترجمه شود. اگر ترجمه، به طور نامناسبی صورت گیرد ویژگی‌های روان‌سنجی آزمون‌ها ممکن است تحت تأثیر قرار گیرد (ارسی‌کان، ۱۹۹۸).

در خصوص کارکرد افتراقی سؤال‌های تیمز مطالعات متعددی صورت گرفته است. از

1. Ercikan

2. Equivalence

3. International Association for the Evaluation of Educational Achievement

جمله این مطالعات می‌توان به ارسی کان (۱۹۹۸)، ارسی کان و کاه (۲۰۰۵)، ارسی کان و مک کریس (۲۰۰۲)، بیتون (۱۹۹۸)، وو، لی، و زامبو^۱ (۲۰۰۷) اشاره کرد. به عنوان مثال، ارسی کان (۱۹۹۸) دریافت که تعداد ۱۸ سؤال (۲۶٪) از ۷۰ سؤال مورد مطالعه در درس علوم TIMSS سال ۱۹۹۵ دارای DIF هستند. از این ۱۸ سؤال تعداد ۱۰ سؤال به نفع گروه مرجع^۲، که دانش‌آموزان انگلیسی زبان بودند، و تعداد ۸ سؤال به نفع گروه هدف^۳، که دانش‌آموزان فرانسوی زبان کشور کانادا بودند، عمل می‌کنند. پژوهش ارسی کان و کاه (۲۰۰۵) نیز نتایج مشابهی به دست داد. آن‌ها دریافتند که سؤال‌های هر دو بخش ریاضی و علوم TIMSS بر علیه دانش‌آموزان فرانسوی زبان عمل می‌کنند.

ارسی کان و مک کریس (۲۰۰۲) نیز دریافتند که در درس ریاضی تیمز ۱۹۹۵، از ۱۵۶ سؤال مورد مطالعه، ۲۲ سؤال در بین دانش‌آموزان فرانسوی و انگلیسی زبان کشور کانادا، ۶۱ سؤال در بین دانش‌آموزان کشور فرانسه و انگلستان، و از بین ۱۵۴ سؤال، ۹۱ سؤال در بین دانش‌آموزان کشور فرانسه و آمریکا دارای DIF هستند. تقریباً ۴۵٪ سؤال‌های دارای DIF به نفع دانش‌آموزان فرانسوی زبان عمل می‌کردند. در درس علوم در مقایسه با درس ریاضی نسبت بیشتری از سؤال‌ها دارای DIF بودند. از ۱۴۰ سؤال مورد مطالعه، تعداد ۵۲ سؤال در بین دانش‌آموزان فرانسوی و انگلیسی زبان کانادا، تعداد ۵۴ سؤال در بین دانش‌آموزان کشور فرانسه و انگلستان، و ۹۱ سؤال در بین دانش‌آموزان کشور فرانسه و آمریکا دارای DIF بودند. در واقع، بین ۵۴ تا ۶۵ درصد سؤال‌های دارای DIF به نفع دانش‌آموزان فرانسوی زبان عمل می‌کردند. این پژوهشگران دریافتند که در ۶۴ سؤال از ۲۹۶ سؤال تیمز، مشکلات مرتبط با ترجمه و انطباق وجود دارد. ارسی کان و مک کریس به این نتیجه رسیدند که در حدود ۳۵ تا ۴۰ درصد سؤال‌های دارای DIF ناشی از مشکلات ترجمه و انطباق سؤال‌ها و تفاوت‌ها در برنامه‌های درسی کشورها است.

1. Wu, Li & Zumbo
2. reference group
3. focal group

وو، لی، و زامبو (۲۰۰۷) تغییرناپذیری اندازه‌گیری یکی از دفترچه‌های آزمون ریاضی پایه هشتم تیمز ۱۹۹۹ را در میان ۷ کشور استرالیا، نیوزیلند، آمریکا، کانادا، کره، ژاپن، و تایوان مورد بررسی قرار دادند. یافته‌های حاصل از پژوهش آن‌ها نشان داد که در تمام ۲۱ مقایسه جفتی که بین کشورها صورت گرفت، آزمون ریاضیات از تغییرناپذیری شکلی^۱ که ابتدایی‌ترین حالت و اولین گام در اثبات تغییرناپذیری اندازه‌گیری^۲ یک ابزار است (وان دن برگ و لانس^۳، ۲۰۰۰) برخوردار است. پژوهشگران مذکور، به طور کلی به این نتیجه رسیدند که در مقایسه‌های بین فرهنگی، تغییرناپذیری اندازه‌گیری تنها در سطح ضعیف (تغییرناپذیری متریک^۴) برقرار است. این نتیجه حاکی از این است که در هیچ یک از مقایسه‌های بین فرهنگی، تغییرناپذیری عرض از مبدا (تغییرناپذیری قوی^۵)، که شرط کافی برای تغییرناپذیری اندازه‌گیری و مقایسه‌پذیری نمره‌ها محسوب می‌شود، برقرار نیست و از اینرو آزمون ریاضی به صورت یک کل، همواره بر علیه یکی از کشورهای مورد مقایسه، دارای سوگیری است.

با در نظر گرفتن مطالبی که در خصوص کاربردهای نتایج مطالعات TIMSS و لزوم معادل‌بودن و مقایسه‌پذیری سؤال‌ها و نتایج آزمون‌ها گفته شد و با توجه به تفاوت‌های موجود بین نظام آموزشی ایران با نظام آموزشی کشورهای مبدا این پروژه، و با در نظر گرفتن تأثیر ترجمه بر ویژگی مورد اندازه‌گیری و کارکرد سؤال، پژوهش حاضر تلاش می‌کند تا به سؤال‌های زیر پاسخ دهد:

۱- آیا ساختار عاملی / سازه‌ای آزمون علوم تیمز ۲۰۰۷ در بین دانش‌آموزان ایران و آمریکا یکسان و قابل مقایسه است؟

۲- آیا سؤال‌های آزمون علوم تیمز ۲۰۰۷ بر له یا علیه دانش‌آموزان ایرانی دارای DIF

1. configural invariance
2. measurement invariance
3. Vandenberg & Lance
4. metric invariance
5. strong invariance

است و به چه میزان؟

۳- تأثیر سئوال‌های دارای DIF بر عملکرد دانش‌آموزان ایرانی در آزمون علوم پایه هشتم تیمز ۲۰۰۷، چگونه و به چه میزان است؟

روش

جامعه آماری و گروه نمونه پژوهش حاضر همان جامعه آماری و گروه نمونه پایه هشتم (جمعیت ۲) در آزمون علوم تیمز ۲۰۰۷ است که در سال تحصیلی ۱۳۸۵-۱۳۸۶ به اجرا درآمد. حجم نمونه دانش‌آموزان ایرانی ۳۹۸۱ نفر و حجم نمونه دانش‌آموزان آمریکایی ۷۳۷۷ نفر است (السن، مارتین و مولیس، ۲۰۰۸).

چارچوب نمونه‌گیری در تیمز ۲۰۰۷، از طرح نمونه‌گیری خوشه‌ای طبقه‌ای دو مرحله‌ای متناسب با حجم نمونه استفاده شده است. در گام اول، مدارس و در گام دوم، کلاس در پایه هدف، نمونه‌گیری شده‌اند. طرح به روش «احتمال متناسب با حجم (PPS)» سهم هر یک از خوشه‌های نمونه را متناسب با حجم آن در جامعه تعیین می‌کند و با بهره‌گیری از وزن‌های نمونه‌گیری، شامل وزن کلاس، مدرسه، دانش‌آموز، وزن کل و وزن نهایی، اطمینان حاصل می‌شود که شاخص‌های آماری به دست آمده از نمونه مورد مطالعه، معرف جامعه است (السن، مارتین و مولیس، ۲۰۰۸). استفاده از این روش باعث می‌شود تا حجم مدارس، عاملی در افزایش احتمال انتخاب مدرسه باشد. بطوریکه احتمال انتخاب شدن مدرسه‌ای با ۴۰ دانش‌آموز، ۲ برابر مدرسه‌ای با ۲۰ دانش‌آموز است.

ابزار مورد استفاده در این پژوهش، آزمون علوم پایه هشتم (سوم راهنمایی) تیمز ۲۰۰۷ است. این آزمون دارای ۲۱۴ سئوال اصلی و ۱۴ زیر سئوال است که مجموعاً ۲۲۸ سئوال را تشکیل می‌دهند که در ۱۴ دفترچه ارائه شده‌اند. این دفترچه‌ها به صورت تصادفی در بین دانش‌آموزان توزیع شدند. از بین ۲۲۸ سئوال، تعداد نیمی از آن‌ها دو یا چند گزینه‌ای است که به صورت دو ارزشی ۰ یا ۱ نمره‌گذاری می‌شوند و نیم دیگر، سئوال‌های باز پاسخ

هستند که به صورت چند ارزشی از ۰ تا حداکثر ۳ نمره‌گذاری می‌شوند. در پژوهش حاضر به منظور ایجاد سهولت در تحلیل داده‌ها و به پیروی از لی، پارک و تایلان (۲۰۱۱) سؤال‌های چند ارزشی از طریق کدگذاری مجدد به سؤال‌های دو ارزشی تبدیل گردید. برای این منظور، اگر فردی در یک سؤال چند ارزشی بالاترین نمره را بدست آورده بود، نمره ۱ و چنانچه نمره کامل را نگرفته بود نمره صفر تعلق گرفت. همچنین به پیروی از ارسی‌کان (۱۹۹۸) سؤال‌هایی که فرد آن‌ها را خالی گذاشته بود، که در چارچوب تیمز آن‌ها را سؤال‌های بدون پاسخ^۱ می‌نامند، و همچنین سؤال‌هایی که بدلیل کمبود زمان فرصت پاسخ‌دهی به آن‌ها نبوده است، که در چارچوب تیمز آن‌ها را سؤال‌های جامانده^۲ می‌نامند، نمره صفر داده شد.

همانطور که ذکر شد آزمون علوم تیمز ۲۰۰۷ دارای ۱۴ دفترچه است که از ۱ تا ۱۴ شماره‌گذاری شده‌اند. سؤال‌های هر دفترچه در قالب دو بلوک تقسیم‌بندی شده است و یکی از بلوک‌ها با دفترچه ماقبل و بلوک دیگر با دفترچه بعد، مشترک و یکسان است. به عنوان مثال، دفترچه ۱ شامل بلوک ۱ و ۲، دفترچه ۲ شامل بلوک ۲ و ۳، دفترچه ۳ شامل بلوک ۳ و ۴ است و الی‌آخر. به منظور افزایش حجم نمونه و برآورد دقیق پارامترها بلوک‌های مشترک بین دفترچه‌ها با یکدیگر ترکیب شدند و ۱۴ بلوک سؤال شکل گرفت و تحلیل‌ها بر اساس بلوک‌ها انجام شد.

روش تحلیل داده‌ها

در پژوهش‌های مربوط به هم‌ارزی/تغییرناپذیری اندازه‌گیری آزمون‌های تیمز به صورتهای مختلفی عمل شده است. ارسی‌کان (۱۹۹۸) از روش χ^2 منقل-هانسزل استفاده کرد، در حالیکه ارسی‌کان و کاه (۲۰۰۵) از ترکیبی از روش‌های CFA و IRT استفاده نمودند. آن‌ها در تحلیل‌های خود از داده‌های سطح سؤال استفاده کردند. به این معنا که در هر دو رویکرد

1. omitted
2. not reached

IRT و CFA تحلیل‌ها در سطح سؤال صورت گرفته است و سؤال‌ها به صورت مستقیم مدل‌پردازی شدند. وو، لی، و زامبو (۲۰۰۷) از تحلیل عاملی تأییدی چند گروهی (MG-CFA) استفاده کردند. در پژوهش حاضر بنا به ۲ دلیل زیر از ترکیبی از روش‌های مبتنی بر IRT و CFA استفاده شد. ۱- محدودیت کاربرد مدل‌های CFA برای داده‌های گسسته (مانند داده‌های مربوط به سؤال‌های آزمون‌های تیمز)؛ و ۲- روش‌های تحلیل DIF بر مفروضه‌هایی استوارند که در موقعیت‌هایی که در آن‌ها آزمونی به زبانهای مختلف ترجمه و اجرا می‌شود ممکن است برقرار نباشند.... در چنین مواردی رویکردهای دیگری که کمتر، سؤال-محور^۱ هستند ارجحیت دارند (دورانز و کولیک، ۲۰۰۶). به عبارت دقیق‌تر، روش‌های تحلیل DIF (از جمله روش‌های مبتنی بر IRT) بر این فرض استوارند که گروه‌های مورد مطالعه به یک جامعه مربوط می‌شوند، یعنی اینکه آزمون، در گروه‌های مورد مطالعه، سازه یا سازه‌های یکسانی را می‌سنجد. بنابراین لازم است قبل از کاربرد این روش‌ها با استفاده از روش‌های مناسب نشان داد که ساختار عاملی آزمون در گروه‌های مورد مطالعه یکسان است.

تریزی (۲۰۰۶) معتقد است که یک رویکرد عملی برای استفاده ترکیبی از روش‌های CFA و IRT این است که ابتدا اولین سطح تغییرناپذیری عاملی، یعنی تغییرناپذیری شکلی با استفاده از رویکرد CFA مورد آزمون قرار گیرد و سپس با استفاده از روش‌های تحلیل DIF، وجود، نوع و بزرگی DIF مورد مطالعه قرار گیرد.

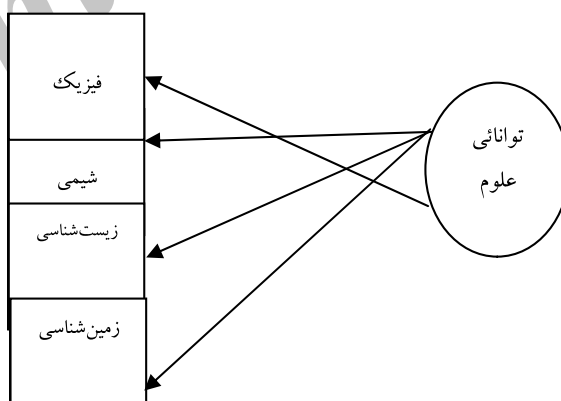
به همین دلیل در پژوهش حاضر برای پاسخگویی به سؤال‌های پژوهش، از ترکیبی از روش‌های مبتنی بر IRT و CFA استفاده گردید.

مجموعه تحلیل‌های اول (تحلیل عاملی تأییدی چند گروهی): با توجه به مطالبی که در خصوص کاربرد ترکیبی مدل‌های CFA و IRT گفته شد، در مجموعه تحلیل‌های اول، اولین سطح هم‌ارزی/تغییرناپذیری عاملی، یعنی تغییرناپذیری شکلی در سطح حیطه‌های محتوایی مورد بررسی قرار گرفت. همانطور که ذکر شد، به منظور انجام تحلیل‌ها بلوک‌های

1. item-centered

سؤال شکل گرفت که هر کدام از ۴ حیطه محتوایی زیست‌شناسی، فیزیک، شیمی و زمین‌شناسی تشکیل یافته است. در پژوهش حاضر هر یک از بلوک‌ها به صورت جداگانه یکبار به داده‌های هر یک از دو کشور ایران و آمریکا و یکبار به صورت همزمان به داده‌های هر دو کشور برازش داده شدند. به عبارت دقیق‌تر، در مجموع ۲۸ تحلیل عاملی تأییدی تک‌گروهی و ۱۴ تحلیل عاملی تأییدی دو‌گروهی انجام پذیرفت. برای انجام تحلیل‌های مربوط به CFA از نسخه ۶/۱۱ نرم‌افزار Mplus (موتن و موتن، ۲۰۱۰) استفاده شد.

ابتدا برای هر یک از بلوک‌ها یک مدل تک‌عاملی که هر یک از حیطه‌های محتوایی به عنوان نشانگرهای آن عمل می‌کنند، به صورت جداگانه به داده‌های هر یک از دو گروه دانش‌آموزان ایرانی و آمریکایی برازش داده شد. این مدل که مدل پایه اولیه نامیده می‌شود و در شکل زیر نشان داده شده است مدل ۱ را تشکیل می‌دهد. دلیل استفاده از حیطه‌های محتوایی به عنوان نشانگر این است که زمانی که آزمون، طولانی است به گونه‌ای که تعداد ۵ الی ۸ سؤال روی هر عامل بار می‌شود غیر ممکن است که مدل مورد نظر برازش رضایت‌بخشی با داده‌ها داشته باشد و مورد تأیید قرار گیرد. در اینگونه موارد، استفاده از بسته‌های سؤال، مناسب خواهد بود (فلوید، و ویدامن، ۱۹۹۵).



شکل ۱: مدل عاملی مفروض

1. Item Parcel
2. Floyd & Widaman

چنانچه نتایج آماره‌ها و شاخص‌های برازش نشان می‌داد که مدل ارائه شده در شکل ۱ از برازش مطلوبی با داده‌ها برخوردار است این مدل به عنوان مدل پایه در نظر گرفته می‌شد و تغییرناپذیری شکلی آن مورد آزمون قرار می‌گرفت. لکن چنانچه یافته‌ها نشان می‌داد که مدل با داده‌ها برازش ندارد، در اینصورت با استفاده از آزمون‌های لاگرانژ و شاخص‌های اصلاح، تغییرات لازم در مدل صورت می‌گرفت تا مدلی با بهترین برازش برای هر یک از گروه‌ها بدست آید. بدین صورت، مدل پایه نهایی بدست می‌آمد. پس از تدوین و فرمولبندی مدل پایه نهایی، تغییرناپذیری شکلی آن در بین دو گروه مورد آزمون قرار می‌گرفت.

مجموعه تحلیل‌های دوم (تحلیل‌های مبتنی بر IRT): هدف دومین مجموعه تحلیل‌ها بررسی کارکرد افتراقی سؤال‌های هر یک از بلوک‌ها است. در این بخش علاوه بر تحلیل‌های مربوط به مفروضه ابعاد و برآورد پارامترهای سؤال‌ها، به تعداد سؤال‌های آزمون علوم (یعنی ۲۲۸ سؤال) تحلیل‌های مربوط به DIF تکرار گردید. همچنین برای هر یک از سؤال‌ها و بلوک‌ها شاخص اندازه اثر برآورد گردید. در ادامه نحوه انجام این تحلیل‌ها توضیح داده می‌شود.

۱- بررسی مفروضه تک‌بعدی بودن سؤال‌های هر یک از حوزه‌های محتوایی و تعیین مدل یا مدل‌هایی که بهترین برازش را با داده‌ها دارد. جهت بررسی تک‌بعدی بودن سؤال‌های هر یک از بلوک‌ها از روش NOHARM (فریزر، ۱۹۸۸) استفاده شد.

۲- شناسایی سؤال‌های لنگر. پس از اتمام مرحله فوق، ابتدا یک مدل که در آن، کلیه پارامترهای تمام سؤال‌های بلوک مورد نظر بین دو گروه (گروه مرجع^۱ و گروه هدف^۲) محدود^۳ شده بودند، فرمولبندی^۴ و مورد آزمون قرار گرفت.

۳- شناسایی نهایی سؤال‌های دارای DIF. پس از تعیین مجموعه سؤال‌های نهایی لنگر، بقیه سؤال‌های بلوک مورد نظر از لحاظ دارا بودن DIF مورد آزمون قرار گرفتند.

1. reference
2. focal
3. constrained
4. specification

در این مرحله سؤال‌های لنگر به عنوان متغیر هم‌تاسازی بکار رفتند و دیگر از نظر DIF مورد مطالعه قرار نگرفتند.

۴- برآورد نهایی پارامترهای سؤال و توانایی. پس از شناسایی سؤال‌های دارای DIF و تعیین نوع DIF (یکنواخت یا غیریکنواخت)، یکبار دیگر، پارامترهای کلیه سؤال‌ها براساس یک مدل دو گروهی برآورد شد. در این مدل، سؤال‌های لنگر و همچنین آن دسته از سؤال‌هایی که فاقد DIF تشخیص داده شدند، بین دو گروه محدود شدند و برای سایر سؤال‌هایی که واجد DIF تشخیص داده شدند، آن پارامتری که باعث شده بود تا سؤال دارای DIF تشخیص داده شود بین گروه‌ها آزاد گذاشته می‌شد و پارامتر دیگر سؤال محدود می‌شد. به عبارت دقیق‌تر، اگر سؤال براساس پارامتر a دارای DIF بود هر دو پارامتر a و b بین دو گروه آزاد گذاشته می‌شد و برای هر گروه به صورت جداگانه برآورد می‌شدند، اما اگر سؤال دارای DIF یکنواخت بود صرفاً پارامتر b بین دو گروه آزاد گذاشته می‌شد و به صورت جداگانه برآورد می‌گردد.

۵- ارزشیابی بزرگی DIF. بزرگی DIF به میزان تفاوت در عملکرد افراد گروه مرجع و هدف در سؤال اشاره دارد که از لحاظ توانایی هم‌تاساز شده‌اند (تریزی و همکاران، ۲۰۰۷). به عبارت دیگر، بزرگی DIF به شدت یا درجه کارکرد افتراقی سؤال اشاره می‌کند. از نمره‌های مورد انتظار سؤال به عنوان شاخص بزرگی DIF استفاده شد. در سؤال‌های دو ارزشی، نمره مورد انتظار سؤال، همان احتمال پاسخ درست به سؤال است.

یک راه برای کمی کردن تفاوت‌های بین متوسط نمره مورد انتظار گروه‌ها در سؤال، شاخص NCDIF است که توسط کمیلی و شپارد (۱۹۹۴)، نقل از اورلاندو-ادلن و همکاران، (۲۰۰۶) برای سؤال‌های دو ارزشی ارائه گردید و راجو و همکارانش (۱۹۹۵) آن را برای سؤال‌های چند ارزشی بسط و گسترش دادند. برای بدست آوردن این شاخص، نمره مورد انتظار تک تک افراد گروه هدف یکبار به عنوان عضوی از همان گروه و یکبار به عنوان عضو گروه مرجع به دست آمد. میانگین مجذورات این تفاوت‌ها برای کل گروه

هدف بدست آمد که به عنوان شاخص بزرگی DIF تفسیر می‌شود.

۶- ارزشیابی تأثیر DIF. تأثیر^۱ به اثرگذاری DIF روی نمره کل آزمون اشاره دارد. در پژوهش حاضر منظور از نمره کل آزمون، نمره کل در بلوک است. برای ارزشیابی تأثیر DIF، بسته به روش مورد استفاده برای شناسایی DIF، روش‌های گوناگونی ارائه شده است. در بافت IRTLRLR تأثیر DIF از طریق تابع‌های پاسخ - آزمون^۲ (TRF) مورد بررسی قرار می‌گیرد. در واقع TRF، یک منحنی شبیه منحنی ویژگی سؤال^۳ است که نمره مورد انتظار در کل آزمون را برای هر سطح از توانایی نشان می‌دهد. برای بدست آوردن نمره مورد انتظار کل آزمون (بلوک)، نمره‌های مورد انتظار سؤال‌ها برای هر سطح از توانایی، جمع شدند. پس از بدست آوردن نمره‌های مورد انتظار کل آزمون، TRF‌های هر دو گروه هدف و مرجع روی یک نمودار رسم گردید. تفاوت‌های موجود بین این دو منحنی، تأثیر نسبی DIF را نشان می‌دهد. برای برآورد و ارزشیابی تأثیر DIF در سطح بلوک‌ها از نرم‌افزار DFIT5 (راجو، ۱۹۹۹) استفاده شد.

برای تحلیل داده‌ها از نسخه ۶/۱۱ نرم‌افزار Mplus (موتن و موتن، ۲۰۱۰)، نسخه ۷/۰۳ نرم‌افزار MULTILOG (تیسن، چن، و باک، ۲۰۰۳)، نرم‌افزار DFIT5 (راجو، ۱۹۹۹)، نرم‌افزار IRTLRLR (تیسن، ۲۰۰۱)، SPSS15 و برای رسم نمودارها از Excel استفاده گردید.

نتایج

جدول ۱ اطلاعات جمعیت‌شناختی دانش‌آموزان شرکت‌کننده در تیمز ۲۰۰۷ را به تفکیک کشور و بلوک نشان می‌دهد. در ستون مربوط به جنسیت، اعداد داخل پرانتز، درصد و اعداد خارج پرانتز، تعداد دانش‌آموزان است. در ستون مربوط به سن نیز اعداد خارج پرانتز، بیانگر میانگین و اعداد داخل پرانتز، بیانگر انحراف استاندارد است.

1. DIF impact
2. Test- response function
3. Item-response function

جدول ۱: اطلاعات جمعیت شناختی نمونه

بلوک	ایران				آمریکا			
	سن	جنسیت		تعداد کل	سن	جنسیت		تعداد کل
		دختر	پسر			دختر	پسر	
۱	۱۴/۱۴ (۰/۷۳)	۲۵۶ (۴۴/۸)	۳۱۶ (۵۵/۲)	۵۷۲	۱۴/۲۸ (۰/۷۹)	۵۵۶ (۵۱/۸)	۵۱۸ (۴۸/۲)	۱۰۷۴
۲	۱۴/۱۰ (۰/۶۸)	۲۶۳ (۴۵/۹)	۳۱۰ (۵۴/۱)	۵۷۳	۱۴/۲۷ (۰/۴۹)	۵۳۹ (۵۰/۶)	۵۲۶ (۴۹/۴)	۱۰۶۹
۳	۱۴/۰۹ (۰/۶۴)	۲۵۹ (۴۵/۸)	۳۰۶ (۵۴/۲)	۵۶۵	۱۴/۲۶ (۰/۵۰)	۵۲۴ (۴۹/۹)	۵۲۶ (۵۰/۱)	۱۰۵۰
۴	۱۴/۱۳ (۰/۶۴)	۲۵۶ (۴۴/۹)	۳۱۴ (۵۵/۱)	۵۷۰	۱۴/۲۶ (۰/۵۱)	۵۴۷ (۵۱/۵)	۵۱۶ (۴۸/۵)	۱۰۶۳
۵	۱۴/۱۹ (۰/۶۹)	۲۵۰ (۴۴/۱)	۳۱۷ (۵۵/۹)	۵۶۷	۱۴/۲۸ (۰/۴۹)	۵۵۱ (۵۲/۱)	۵۰۶ (۴۷/۹)	۱۰۵۷
۶	۱۴/۱۸ (۰/۷۱)	۲۴۵ (۴۳/۷)	۳۱۶ (۵۶/۳)	۵۶۱	۱۴/۲۶ (۰/۴۶)	۵۱۷ (۴۹/۳)	۵۳۲ (۵۰/۷)	۱۰۴۹
۷	۱۴/۱۶ (۰/۶۵)	۲۴۷ (۴۴/۳)	۳۱۱ (۵۵/۷)	۵۵۸	۱۴/۲۵ (۰/۴۷)	۵۰۸ (۴۸/۴)	۵۴۱ (۵۱/۶)	۱۰۴۹
۸	۱۴/۱۷ (۰/۶۷)	۲۵۵ (۴۴/۸)	۳۱۴ (۵۵/۲)	۵۶۹	۱۴/۲۸ (۰/۴۸)	۵۱۴ (۴۹/۶)	۵۳۵ (۵۰/۱)	۱۰۴۹
۹	۱۴/۱۷ (۰/۶۸)	۲۵۸ (۴۴/۷)	۳۱۹ (۵۵/۳)	۵۷۷	۱۴/۲۷ (۰/۴۷)	۵۱۱ (۴۹/۴)	۵۲۴ (۵۰/۶)	۱۰۳۹
۱۰	۱۴/۱۷ (۰/۷۲)	۲۵۸ (۴۴/۹)	۳۱۷ (۵۵/۱)	۵۷۵	۱۴/۲۷ (۰/۴۶)	۵۱۵ (۴۹/۶)	۵۲۴ (۵۰/۴)	۱۰۳۹
۱۱	۱۴/۱۶ (۰/۷۲)	۲۵۹ (۴۴/۷)	۳۲۱ (۵۵/۳)	۵۸۰	۱۴/۲۷ (۰/۴۸)	۵۲۵ (۴۹/۷)	۵۳۲ (۵۰/۳)	۱۰۵۷
۱۲	۱۴/۱۵ (۰/۶۹)	۲۵۱ (۴۴/۶)	۳۱۲ (۵۵/۴)	۵۶۳	۱۴/۲۶ (۰/۴۸)	۵۲۲ (۴۹/۸)	۵۲۶ (۵۰/۲)	۱۰۴۸
۱۳	۱۴/۱۸ (۰/۷۰)	۲۵۷ (۴۵/۸)	۳۰۵ (۵۴/۳)	۵۶۲	۱۴/۲۶ (۰/۴۸)	۵۴۶ (۵۱/۸)	۵۰۹ (۴۸/۲)	۱۰۵۹
۱۴	۱۴/۱۸ (۰/۷۲)	۲۵۸ (۴۵/۳)	۳۱۲ (۵۴/۷)	۵۷۰	۱۴/۲۸ (۰/۴۸)	۵۶۷ (۵۳/۳)	۴۹۷ (۴۶/۷)	۱۰۶۴

در ادامه نتایج حاصل از تحلیل های مربوط به تغییرناپذیری شکلی و کارکرد افتراقی سؤال (DIF) و کارکرد افتراقی آزمون (DTF) در دو بخش جداگانه ارائه می شود.

نتایج مربوط به تغییرناپذیری شکلی: انجام آزمون تغییرناپذیری شکلی و برآورد پارامترها در مدل CFA مستلزم انتخاب یک تابع یا برآورد کننده مناسب است. تابع بیشینه درست نمایی (ML) پرکاربردترین روش برای برآورد پارامترها و آزمون برازش مدل با داده ها است. یکی از مفروضه های این تابع، نرمال بودن چند متغیری توزیع داده های نشانگرها است. بررسی شاخص های چولگی و کشیدگی نشان داد که نمره های هیچ یک از حیطه های محتوایی بلوک ها دارای توزیع نرمال نیست. علاوه بر این، دامنه آماره Q_3 (اسمال، ۱۹۸۰، نقل از دی کارلو، ۱۹۹۷)، که نرمال بودن چندمتغیری را آزمون می کند، برای داده های کشور ایران از ۱۰/۱/۶۶ برای بلوک شماره ۱۲ تا ۴۰۹/۳۳ برای بلوک شماره

۱ است. دامنه این آماره برای داده‌های کشور آمریکا از $380/05$ برای بلوک شماره ۲ تا $2651/80$ برای بلوک شماره ۵ متغیر است. آماره Q_3 بر توزیع χ^2 استوار است و برای آزمون معناداری از مقادیر این توزیع استفاده می‌شود. نتایج بدست آمده نشان داد که کلیه مقادیر مشاهده شده آماره Q_3 در سطح آلفای $0/01$ به لحاظ آماری معنادار است. معنای این مطلب آن است توزیع داده‌های هیچ کدام از بلوک‌ها نرمال چندمتغیری نیست. از این رو برای برآورد پارامترها و برازش مدل با داده‌ها از برآوردکننده بیشینه درست‌نمایی مقاوم یا MLM (براون، ۲۰۰۶) استفاده گردید. در این برآوردکننده، پارامترهای مدل به شیوه ML برآورد می‌شوند، اما برای آزمون معناداری پارامترها و آزمون برازش مدل با داده‌ها از خطاهای استاندارد مقاوم و آماره χ^2 اصلاح شده که به آماره χ^2 ساتورا-بنتلر ($SB \chi^2$) نیز معروف است استفاده می‌شود (کلاین، ۲۰۱۱).

برای سنجش برازش مدل، معمولاً از چندین شاخص استفاده می‌شود. آماره χ^2 بدلیل حساسیتی که به حجم نمونه دارد، در گذشته بر استفاده از این آماره برای بررسی برازش مدل با داده‌ها تأکید زیادی نمی‌شد. اما امروزه پژوهشگران زیادی از جمله (بارت، ۲۰۰۷) معتقدند که برای بررسی برازش مدل با داده‌ها بایستی بر این آماره توجه زیادی شود.

در پژوهش حاضر برای سنجش برازش مدل، از آماره‌های χ^2 ساتورا-بنتلر، ریشه دوم برآورد واریانس خطای تقریب (RMSEA)، شاخص برازش تطبیقی (CFI)^{۱۷} و ریشه دوم میانگین مجذورات پس‌مانده‌های استاندارد شده^{۱۸} (SRMR) و شاخص تاکر-لویز^{۱۹} (TLI) استفاده گردید.

در تفسیر شاخص‌های برازندگی، از قواعد پیشنهادی زیر استفاده شد. براونی و کاداک (۱۹۹۳) معتقدند که مقادیر بالاتر از $0/90$ برای شاخص‌های TLI و CFI حاکی از برازش خوب مدل است. در مورد شاخص RMSEA نیز از ملاک پیشنهادی هیو و بنتلر (۱۹۹۹) استفاده شد. آن‌ها معتقدند اگر مقدار این شاخص کمتر از $0/06$ باشد می‌توان نتیجه گرفت که برازش مدل با داده‌ها خوب است. در ارتباط با شاخص SRMR نیز مقادیر کمتر از $0/10$

به‌طور کلی به‌صورت مقادیر مطلوب در نظر گرفته می‌شود (کلاین، ۲۰۰۵).

برای آزمون تغییرناپذیری شکلی، ابتدا برای هریک از بلوک‌ها یک مدل تک عاملی که هر یک از حیطه‌های محتوایی به عنوان نشانگرهای آن عمل می‌کنند، به صورت جداگانه به داده‌های دو گروه دانش‌آموزان ایرانی و آمریکایی برازش داده شد. این مدل که مدل پایه اولیه نامیده می‌شود در شکل ۱ نشان داده شده است.

نتایج حاصل از تحلیل‌ها نشان داد که مدل پایه اولیه با داده‌های برخی از بلوک‌ها از برازش مطلوبی برخوردار نیست. بررسی شاخص‌های اصلاح^۱ (MI) نشان داد که واریانس یا عبارت خطای نشانگرها با یکدیگر همبسته هستند. لذا پس از اعمال تغییرات لازم در مدل و همبسته کردن عبارت‌های خطای نشانگرها، مدل پایه نهایی، به صورت جداگانه به داده‌های دانش‌آموزان ایران و آمریکا برازش داده شد و شاخص‌های برازش مورد بررسی قرار گرفت. در مرحله بعد، مدل پایه نهایی، به طور همزمان به داده‌های دو گروه برازش داده شد و میزان برازش مدل با داده‌ها با استفاده از آماره‌های برازش مورد بررسی قرار گرفت.

تحلیل عاملی تأییدی و تغییرناپذیری شکلی برای تمام ۱۴ بلوک صورت گرفت. لکن به دلیل محدودیت جا از ارائه مفصل نتایج مربوط به این بخش از تحلیل‌ها خودداری و صرفاً نتیجه مربوط به بلوک ۷ به عنوان نمونه ارائه می‌شود.

جدول ۳ نتایج مربوط به برازش مدل و تغییرناپذیری شکلی را برای بلوک ۷ نشان می‌دهد. نتایج آماره χ^2 و سایر شاخص‌های برازش نشان داد که مدل پایه اولیه (مدل ۱) با داده‌های هیچ یک از دو گروه از برازش مناسبی برخوردار نیست. بررسی شاخص‌های اصلاح در گروه ایران نشان داد که کوواریانس بین عبارت‌های خطای دو درس زمین‌شناسی و فیزیک به لحاظ آماری معنادار است ($MI = 26/45$). بنابراین کوواریانس بین عبارت‌های خطای این دو درس، آزاد گذاشته شد و برازش مدل با داده‌های ایران، مجدداً مورد آزمون قرار گرفت. نتایج آماره χ^2 و سایر شاخص‌ها نشان داد که مدل از

1. modification index

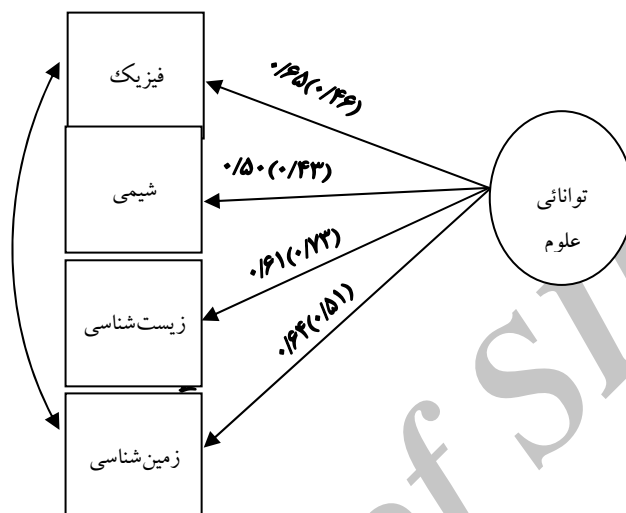
برازش مطلوبی با داده‌ها برخوردار است. این مدل (مدل ۲) به عنوان مدل پایه نهایی برای ایران در نظر گرفته شد.

بررسی شاخص‌های اصلاح در گروه آمریکا نیز نشان داد که کوواریانس بین عبارت خطای درس زمین‌شناسی و فیزیک به لحاظ آماری معنادار است ($MI = 7/16$). بنابراین، کوواریانس بین عبارت‌های خطای این دو درس، آزاد گذاشته شد و مدل مجدداً به داده‌های آمریکا برازش داده شد. همانطور که در جدول ۳ مشاهده می‌کنید، مدل از برازش مطلوبی با داده‌های آمریکا برخوردار است. این مدل (مدل ۲) به عنوان مدل پایه نهایی برای کشور آمریکا در نظر گرفته شد. در گام بعد، مدل‌های پایه نهایی (مدل‌های ۲) به صورت همزمان به داده‌های دو گروه ایران و آمریکا برازش داده شد. هدف این مرحله، آزمون تغییرناپذیری شکلی است. نتایج مربوط به این مرحله نشان می‌دهد که مدل از تغییرناپذیری شکلی مطلوبی بین دو گروه برخوردار است.

جدول ۳: آزمون برازش مدل و تغییرناپذیری شکلی بلوک ۷ در بین دو گروه ایران و آمریکا

کشور	مدل	df	p	CFI	TLI	SRMR	RMSEA	RMSEA حد پایین	فاصله اطمینان ۹۰٪ برای
ایران	مدل ۱	۲۵/۶۰۳	۲	۰/۰۰۰	۰/۹۳۶	۰/۸۰۹	۰/۰۴۲	۰/۱۴۵	۰/۱۹۸
	مدل ۲	۰/۰۳۳	۱	۰/۸۵۶	۱/۰۱۶	۰/۰۰۱	۰/۰۰۱	۰/۰۰۰	۰/۰۶۲
آمریکا	مدل ۱	۸/۱۴۱	۲	۰/۰۱۷	۰/۹۹۳	۰/۹۷۹	۰/۰۱۷	۰/۰۵۴	۰/۰۹۵
	مدل ۲	۱/۰۶۶	۱	۰/۳۰۲	۱	۱	۰/۰۰۵	۰۰۸	۰/۰۸۳
تغییرناپذیری شکلی		۱/۰۹۸	۲	۰/۵۷۷	۱/۰۰۴	۱/۰۰۴	۰/۰۰۴	۰/۰۰۰	۰/۰۵۹

شکل ۲ ضرایب استاندارد شده پارامترهای مدل (بارهای عاملی) را برای دو کشور ایران و آمریکا نشان می‌دهد. ضرایب داخل پرانتز مربوط به ایران و ضرایب خارج پرانتز مربوط به آمریکا است. همه این ضرایب در سطح آلفای ۰/۰۱ معنادار است.



شکل ۲: نمودار مدل عاملی بلوک ۷ همراه با برآوردهای استانداردهای شده پارامترها

به طور کلی نتایج حاصل از تحلیل عاملی تأییدی نشان داد که تمام ۱۴ بلوک، از تغییرناپذیری شکلی در بین دو گروه ایران و آمریکا برخوردارند. برقرار بودن تغییرناپذیری شکلی بدان معنا است که دانش‌آموزان دو کشور ایران و آمریکا برای دادن پاسخ به سؤال‌های علوم از چارچوب مفهومی یکسانی استفاده می‌کنند (چونگ و رنسولد، ۲۰۰۲؛ هورن و مک آردل، ۱۹۹۲) و لذا می‌توان از روش‌های DIF مبتنی بر نظریه سؤال - پاسخ برای بررسی کارکرد افتراقی تک تک سؤال‌ها استفاده نمود.

انجام تحلیل‌های DIF مستلزم انتخاب یکی از مدل‌های دو ارزشی IRT به عنوان مدل پایه است. مهم‌ترین مفروضه مدل‌های دو ارزشی IRT، مفروضه تک‌بعدی بودن سؤال‌ها است. یکی از مناسب‌ترین روش‌های آزمون این مفروضه، انجام تحلیل عاملی در سطح سؤال است. بنابراین، پیش از انتخاب یکی از مدل‌های ۱، ۲، و ۳ پارامتری به عنوان مدل پایه، سؤال‌های هر یک از بلوک‌ها با استفاده از تحلیل عاملی غیرخطی که از طریق برنامه NOHARM (فریزر، ۱۹۸۸) قابل اجرا است برای دو کشور ایران و آمریکا به صورت

جداگانه مورد تحلیل قرار گرفت. نتایج حاصل از این تحلیل، که به دلیل محدودیت جا و جلوگیری از حجیم شدن مقاله در اینجا ارائه نشده است، نشان داد که شواهد کافی برای رد فرضیه تک بعدی بودن بلوک‌های درس علوم تیمز ۲۰۰۷ وجود ندارد. این یافته حاکی از این است که سئوال‌های تمام ۱۴ بلوک در دو گروه ایران و آمریکا صرفاً یک سازه یا عامل مکنون را می‌سنجند و مفروضه تک‌بعدی بودن در مورد تمام آن‌ها صادق است. به منظور انتخاب مدل پایه، مدل‌های لوجستیک ۱، ۲ و ۳ پارامتری به هر یک از بلوک‌ها و به صورت جداگانه به داده‌های ایران و آمریکا برازش داده شد و با استفاده از آماره نسبت درست‌نمایی (-2LL) که با G^2 نشان داده می‌شود، میزان برازش این مدل‌ها با یکدیگر مورد مقایسه قرار گرفت. نتایج حاصل از این تحلیل‌ها که با بهره‌گیری از نرم‌افزار MULTILOG صورت گرفت نشان داد که مدل لوجستیک دو پارامتری در مقایسه با دو مدل دیگر، از برازش بهتری با داده‌ها برخوردار است. بنابراین، این مدل به عنوان مدل پایه در نظر گرفته شد.

پس از تعیین مدل پایه، به منظور شناسایی سئوال‌های دارای DIF، هر یک از بلوک‌ها با استفاده از نرم‌افزار IRTL RDIF به صورت جداگانه مورد تحلیل قرار گرفتند. شناسایی سئوال‌های دارای DIF در دو مرحله صورت گرفت. در واقع، برنامه IRTL RDIF در مورد هر بلوک، دو بار اجرا گردید. در مرحله اول که هدف آن شناسایی سئوال‌های لنگر اولیه است، برنامه IRTL RDIF هر یک از سئوال‌ها را به عنوان سئوال مورد مطالعه و بقیه سئوال‌ها را به عنوان سئوال لنگر در نظر می‌گیرد و بر این اساس تعیین می‌کند که آیا سئوال مورد مطالعه، دارای DIF است یا نه. با اجرای این مرحله، سئوال‌های لنگر اولیه شناسایی گردید. آنگاه سایر سئوال‌ها که دارای DIF تشخیص داده شدند از فایل داده‌ها حذف گردیدند. به عبارت دقیق‌تر، پس از اجرای مرحله اول، برای هر یک از بلوک‌ها فایل جدیدی تشکیل گردید که صرفاً سئوال‌های لنگر اولیه را شامل می‌شد. آنگاه این فایل‌های داده مجدداً با برنامه IRTL RDIF مورد تحلیل قرار گرفتند. هدف از این کار،

خالص‌سازی متغیر هم‌تاسازی به منظور شناسایی سؤال‌های لنگر نهایی بود.

پس از شناسایی سؤال‌های لنگر نهایی، برنامه IRTL RDIF یکبار دیگر در مورد فایل اولیه داده‌های هر یک از بلوک‌ها اجرا گردید. در این اجرا، سؤال‌هایی که به عنوان سؤال لنگر نهایی شناسایی شده بودند، به عنوان سؤال لنگر به برنامه معرفی شدند. پس از اجرای این مرحله، سؤال‌های دارای DIF شناسایی گردیدند. به منظور کنترل نرخ خطای نوع اول از اصلاح به روش بنجامینی - هاجبرگ (B-H؛ تیسن، استنبرگ، و کوانگ، ۲۰۰۲) که از طریق برنامه EXCELL صورت گرفت استفاده گردید. پس از اجرای روش B-H سؤال‌های نهایی واجد DIF شناسایی گردید. پس از شناسایی سؤال‌های لنگر و سؤال‌های دارای DIF، سؤال‌های هر یک از بلوک‌ها به صورت جداگانه و از طریق برنامه MULTILOG مورد تحلیل قرار گرفتند. در اجرای این تحلیل که به شیوه چند گروهی صورت گرفت، پارامترهای سؤال‌های لنگر بین دو گروه ایران و آمریکا یکسان قرار داده شدند، در حالیکه پارامترهای سؤال‌های واجد DIF برای دو گروه به صورت جداگانه برآورد گردیدند. البته، پارامتر a سؤال‌هایی که دارای DIF یکنواخت بودند نیز بین دو گروه یکسان قرار داده شدند. به عبارت دقیق، در مرحله برآورد نهایی پارامترهای سؤال و توانایی، پارامتر a و b سؤال‌های لنگر و همچنین پارامتر a سؤال‌های دارای DIF یکنواخت بین دو گروه یکسان قرار داده شد و پارامتر a و b سؤال‌های دارای DIF غیریکنواخت و همچنین پارامتر a سؤال‌های دارای DIF یکنواخت بین دو گروه آزاد گذاشته شد و برای هر گروه به صورت جداگانه برآورد گردیدند. نتایج حاصل از این تحلیل در جدول ۴ ارائه شده است. در این جدول علاوه بر برآوردهای نهایی پارامترهای سؤال‌ها، لنگر یا DIF بودن سؤال، نوع DIF و همچنین شاخص اندازه اثر و جهت DIF که به ترتیب با NCDIF و CDIF نشان داده می‌شوند، ارائه شده است.

شاخص NCDIF همانند یک شاخص اندازه اثر تفسیر می‌شود. این شاخص برای شناسایی سؤال‌های دارای DIF در چارچوب رویکرد DFIT بکار می‌رود. برای آزمون

معناداری این شاخص و همچنین برای شاخص‌های CDIF و DTF آماره‌های χ^2 ارائه شده است. اما بدلیل حساسیت زیاد آماره χ^2 به حجم نمونه، بر اساس داده‌های شبیه‌سازی، برای شاخص NCDIF و DTF نقاط برش ارائه شده است.

فلر (۱۹۹۳) و فلاورز، اوشیما و راجو (۱۹۹۵) براساس داده‌های شبیه‌سازی نقاط برش را برای آماره NCDIF ارائه کرده‌اند. مقدار نقطه برش، بسته به تعداد گزینه‌های پاسخ متفاوت است. برای سؤال‌های دو ارزشی نقطه برش ۰/۰۰۶ پیشنهاد شده است. نقطه برش برای شاخص DTF که بیانگر بزرگی کارکرد افتراقی آزمون است از جمع نقاط برش سؤال‌های دارای DIF بدست می‌آید. به عنوان مثال، نقطه برش شاخص DTF برای یک آزمون ۴۰ سؤالی دو ارزشی که فرضاً ۲۵ سؤال آن دارای DIF است برابر با ۰/۱۵ (۲۴/۰۰۶=۲۵×) خواهد بود.

اگر علامت شاخص CDIF مثبت باشد بیانگر این است که سؤال به نفع گروه مرجع (آمریکا) دارای DIF است و اگر علامت آن منفی باشد حاکی از این است که سؤال به نفع گروه هدف (ایران) دارای کارکرد افتراقی است.

جدول ۴: برآوردهای نهایی پارامترهای سئوال‌های لنگر و دارای DIF همراه با شاخص اندازه اثر

شخص CDIF	شخص اندازه اثر	نوع DIF	DIF یا لنگر	آمریکا		ایران		کد سئوال
				b	a	b	a	
۰/۳۴	۰/۰۶	یکنواخت	DIF	۰/۶۲	۰/۹۹	-۰/۵۶	۰/۹۹	S032115
-۰/۱۶	۰/۰۲	یکنواخت	DIF	۰/۳۷	۰/۹۱	۱/۱۰	۰/۹۱	S032565
۰/۲۶	۰/۰۴	یکنواخت	DIF	۰/۵۱	۰/۵۲	-۰/۹۸	۰/۵۲	S032403
۰/۶۱	۰/۱۹	یکنواخت	DIF	۲/۴۶	۰/۳۷	-۲/۷۱	۰/۳۷	S032273
۰/۳۳	۰/۰۶	یکنواخت	DIF	۰/۷۷	۱/۴۷	-۰/۲۵	۱/۴۷	S032019A
۰/۰۰۳	۰/۰۰۴	غیر یکنواخت	DIF	۱/۶۸	۱/۱۴	۰/۹۳	۱/۹۵	S032019B
۰/۴۹	۰/۱۳	یکنواخت	DIF	۰	۱/۱۴	-۱/۹۰	۰/۶۸	S032516
۰/۱۴	۰/۰۱	یکنواخت	DIF	۰/۹۶	۰/۶۸	۱/۰۴	۰/۳۳	S032620
-	-	-	لنگر	-۰/۱۱	۱/۰۶	-۰/۱۱	۱/۰۶	S032693A
-۰/۰۵	۰/۰۰۲	یکنواخت	DIF	-۱/۰۴	۱/۳۸	-۰/۹۸	۱/۳۸	S032693B
-۰/۱۳	۰/۰۱	یکنواخت	DIF	۰/۴۰	۱/۱۳	۰/۹۳	۱/۱۳	S032695
۰/۱۰	۰/۰۰۷	غیر یکنواخت	DIF	-۰/۰۲	۳/۵۲	-۰/۱۲	۲/۳۶	S032697A
۰/۰۶	۰/۰۰۳	غیر یکنواخت	DIF	۰/۰۸	۴/۲۰	۰/۰۳	۳	S032697B
-	-	-	لنگر	۰/۲۵	۴/۴۲	۰/۲۵	۴/۴۲	S032697D
-۰/۰۱	۰/۰۳	یکنواخت	DIF	۰/۸۵	۰/۵۶	۰/۴۵	۰/۵۶	S042009
-	-	-	لنگر	-۱/۷۵	۰/۹۹	-۱/۷۵	۰/۹۹	S042313
-۰/۰۲	۰/۰۵	یکنواخت	DIF	۱/۷۳	۰/۷۸	۰/۵۱	۰/۷۸	S042059
-	-	-	لنگر	۱/۲۳	۱/۱۶	۱/۲۳	۱/۱۶	S042011
۰/۰۱	۰/۰۳	غیر یکنواخت	DIF	۱/۶۹	۰/۷۲	۳/۷۷	۰/۱۷	S042028
-۰/۰۲	۰/۰۵	یکنواخت	DIF	۰/۵۹	۰/۶۲	۲/۶۹	۰/۶۲	S042001
-۰/۰۰۸	۰/۰۱	غیر یکنواخت	DIF	-۰/۲۸	۰/۳۸	۰/۲۴	۰/۶۳	S042276
۰/۰۰	۰/۰۰۴	غیر یکنواخت	DIF	-۰/۸۴	۰/۷۱	-۱/۱۱	۰/۲۷	S042279
۰/۰۰۱	۰/۰۰۱	یکنواخت	DIF	۲/۴۳	۱/۶۴	۱/۷۴	۱/۶۴	S042083
-	-	-	لنگر	۱/۰۲	۱/۴۸	۱/۰۲	۱/۴۸	S042106
۰/۰۳	۰/۱۴	غیر یکنواخت	DIF	۰/۶۰	۰/۶۵	-۱/۲۷	۱/۳۰	S042071
۰/۰۲	۰/۰۶	یکنواخت	DIF	۱/۲۴	۱/۰۸	-۰/۰۵	۱/۰۸	S042101
۰/۰۲	۰/۰۴	یکنواخت	DIF	-۰/۲۹	۱/۲۵	-۰/۹۸	۱/۲۵	S042307
-	-	-	لنگر	-۰/۴۰	۰/۷۲	-۰/۴۰	۰/۷۲	S042405
-۰/۰۰۴	۰/۰۰۹	یکنواخت	DIF	۰/۷۳	۱/۷۶	۱/۳۱	۱/۷۶	S042244A
-	-	-	لنگر	۱/۵۲	۱/۱۹	۱/۵۲	۱/۱۹	S042244B
-۰/۰۱	۰/۰۱	غیر یکنواخت	DIF	۰/۳۶	۰/۸۸	۰/۴۴	۱/۵۴	S042153
۰/۲۵	۰/۲۳	یکنواخت	DIF	۱/۰۳	۰/۹۴	-۱/۳۸	۰/۹۴	S022183
-۰/۱۵	۰/۰۸	یکنواخت	DIF	-۱/۱۹	۱/۱۸	-۰/۰۸	۱/۱۸	S022276
-	-	-	لنگر	-۰/۶۵	۱/۱۱	-۰/۶۵	۱/۱۱	S022115
-۰/۰۴	۰/۰۰۵	یکنواخت	DIF	۰/۲۵	۱/۱۳	-۰/۰۴	۱/۱۳	S022022

شخص CDIF	شخص اندازه اثر	نوع DIF	DIF یا لنگر	آمریکا		ایران		کد ستوال
				b	a	b	a	
-	-	-	لنگر	-۱/۰۴	۱/۰۱	-۱/۰۴	۱/۰۱	S022019
-	-	-	لنگر	-۰/۳۸	۰/۹۴	-۰/۳۸	۰/۹۴	S022002
۰/۰۳	۰/۰۰۴	یکنواخت	DIF	-۰/۶۳	۰/۸۳	-۰/۹۶	۰/۸۳	S022294
-	-	-	لنگر	۵/۸۵	۳/۸۴	۵/۸۵	۳/۸۴	S022106
۰/۰۳	۰/۰۰۴	یکنواخت	DIF	۵/۶۹	۰/۷۲	۳/۱۴	۰/۷۲	S022244
۰/۰۴	۰/۰۰۶	یکنواخت	DIF	-۰/۲۳	۱/۰۹	-۰/۵۴	۱/۰۹	S022150
-۰/۰۹	۰/۰۴	غیر یکنواخت	DIF	-۰/۳۰	۱/۴۳	۱/۶۲	۰/۴۶	S022042
-	-	-	لنگر	۲/۰۶	۰/۹۶	۲/۰۶	۰/۹۶	S022289
۰/۱۲	۰/۰۶	غیر یکنواخت	DIF	۲/۵۳	۰/۸۱	۰/۲۲	۱/۲۶	S022069
-	-	-	لنگر	-۰/۲۶	۰/۹۲	-۰/۲۶	۰/۹۲	S022268
-۰/۰۱	۰/۰۰۱	یکنواخت	DIF	-۲/۳۰	۱/۳۲	-۲/۱۱	۱/۳۲	S042013
-	-	-	لنگر	-۰/۲۸	۰/۶۰	-۰/۲۸	۰/۶۰	S042006
۰/۰۶	۰/۰۲	یکنواخت	DIF	۰/۳۱	۱/۰۹	-۰/۲۵	۱/۰۹	S042310
-۰/۰۲	۰/۰۰۱	یکنواخت	DIF	۰/۷۲	۰/۷۵	۰/۹۷	۰/۷۵	S042052
۰/۱۲	۰/۰۷	یکنواخت	DIF	۰/۴۳	۰/۵۳	-۱/۶۲	۰/۵۳	S042054
-۰/۰۴	۰/۰۱	یکنواخت	DIF	۱/۱۸	۰/۷۵	۲/۰۹	۰/۷۵	S042043
-	-	-	لنگر	۱/۷۷	۱/۱۵	۱/۷۷	۱/۱۵	S042196
-	-	-	لنگر	۱/۱۵	۰/۴۵	۱/۱۵	۰/۴۵	S042061
۰/۱۰	۰/۰۸	یکنواخت	DIF	۱/۶۰	۱/۳۶	۰/۲۳	۱/۳۶	S042292
۰/۰۵	۰/۰۲	غیر یکنواخت	DIF	-۰/۳۵	۱/۸۰	-۰/۷۷	۰/۷۱	S042109
-۰/۰۸	۰/۰۳	یکنواخت	DIF	-۰/۱۸	۰/۸۴	۰/۷۶	۰/۸۴	S042232A
-	-	-	لنگر	۲/۵۵	۲/۲۳	۲/۵۵	۲/۲۳	S042232B
-۰/۰۲	۰/۰۰۳	یکنواخت	DIF	۴/۹۹	۰/۰۸	۳/۸۹	۰/۰۸	S042232C
-۰/۰۴	۰/۰۳	غیر یکنواخت	DIF	-۰/۲۹	۱/۱۸	۱/۶۵	۰/۳۰	S042294
۰/۱۳	۰/۰۸	یکنواخت	DIF	۰/۱۹	۱/۱۸	-۰/۸۵	۱/۱۸	S042149
۰/۰۷	۰/۰۲	غیر یکنواخت	DIF	۰/۵۵	۱/۷۵	۰/۳۲	۱/۰۳	S042155
-۰/۰۶	۰/۰۲	یکنواخت	DIF	-۰/۴۵	۰/۷۳	۰/۲۷	۰/۷۴	S042150
۰/۱۰	۰/۰۲	یکنواخت	DIF	-۰/۶۰	۱/۱۲	۱/۱۵	۱/۱۲	S022290
-	-	-	لنگر	۰/۸۳	۰/۹۱	۰/۸۳	۰/۹۱	S022292
-	-	-	لنگر	-۰/۱۰	۰/۸۹	-۰/۱۰	۰/۸۹	S022054
-	-	-	لنگر	۰/۵۰	۱/۱۰	۰/۵۰	۱/۱۰	S022181
۰/۲۷	۰/۱۵	یکنواخت	DIF	۰/۱۶	۱/۱۶	-۱/۳۸	۱/۱۶	S022208
۰/۰۸	۰/۰۱	غیر یکنواخت	DIF	-۰/۴۴	۱/۷۱	-۰/۷۳	۱/۸۷	S022078
-۰/۱۰	۰/۰۲	یکنواخت	DIF	-۰/۵۲	۰/۷۰	۰/۳۱	۰/۷۰	S022126
-	-	-	لنگر	۰/۹۷	۰/۸۳	۰/۹۷	۰/۸۳	S022281

شخص CDIF	شخص اندازه اثر	نوع DIF	DIF یا لنگر	آمریکا		ایران		کد سوال
				b	a	b	a	
۰/۲۴	۰/۱۱	یکنواخت	DIF	-۰/۱۱	۱/۲۳	-۱/۴۰	۱/۲۸	S032385
-	-	-	لنگر	-۰/۸۸	۱/۰۲	-۰/۸۸	۱/۰۲	S032035
-۰/۰۲	۰/۰۰۱	یکنواخت	DIF	-۰/۷۶	۰/۹۸	-۰/۶۴	۰/۹۸	S032519
۰/۱۸	۰/۰۶	یکنواخت	DIF	۰/۸۵	۰/۹۰	-۰/۴۵	۰/۹۰	S032683
-۰/۱۸	۰/۰۶	یکنواخت	DIF	-۱/۱۷	۰/۷۳	۰/۳۲	۰/۷۳	S032258
-۰/۰۳	۰/۰۰۱	یکنواخت	DIF	۱/۱۶	۱/۰۵	۱/۴۳	۱/۰۵	S032120A
-۰/۰۲	۰/۰۰۴	غیر یکنواخت	DIF	۱	۱/۲۱	۰/۶۴	۲/۱۹	S032120B
۰/۰۰۴	۰/۰۰۷	غیر یکنواخت	DIF	-۰/۸۸	۱/۱۵	-۰/۶۲	۰/۷۲	S042304
-۰/۰۰۷	۰/۰۲	غیر یکنواخت	DIF	-۰/۳۹	۰/۹۰	-۰/۸۰	۱/۶۸	S042038
-	-	-	لنگر	۰/۰۳	۱/۲۶	۰/۰۳	۱/۲۶	S042298
-۰/۰۰۶	۰/۰۳	یکنواخت	DIF	۱/۲۷	۱/۱۱	۰/۲۶	۱/۱۱	S042261
۰/۰۰۸	۰/۰۹	یکنواخت	DIF	-۰/۹۶	۱/۰۷	۰/۳۲	۱/۰۷	S042051A
۰/۰۰۴	۰/۰۰۹	یکنواخت	DIF	۰/۶۰	۱/۵۰	۱/۱۱	۱/۵۰	S042051B
-۰/۰۰۴	۰/۰۱	یکنواخت	DIF	۰/۲۷	۱/۵۴	-۰/۱۵	۱/۵۴	S042076
-۰/۰۰۲	۰/۰۰۳	یکنواخت	DIF	۲/۲۱	۲/۰۷	۱/۳۴	۲/۰۷	S042404
۰/۰۰۲	۰/۰۰۵	یکنواخت	DIF	۰/۵۶	۰/۷۰	۱/۰۷	۰/۷۰	S042306
-۰/۰۰۷	۰/۰۳	غیر یکنواخت	DIF	۰/۸۰	۱/۲۰	-۰/۰۵	۱/۷۰	S042403
-	-	-	لنگر	-۱/۰۱	۱/۰۵	-۱/۰۱	۱/۰۵	S042272
۰/۰۰۴	۰/۰۲	غیر یکنواخت	DIF	۰/۴۷	۰/۷۸	۰/۸۷	۱/۲۹	S042100
-	-	-	لنگر	۳/۷۹	۳/۶۴	۳/۷۹	۳/۶۴	S042238A
۰/۰۰۹	۰/۰۴	غیر یکنواخت	DIF	۰/۵۱	۱/۰۴	۴/۲۷	۰/۲۹	S042238B
۰/۰۰۹	۰/۰۲	غیر یکنواخت	DIF	-۱/۰۱	۱/۳۷	-۱/۳۶	۰/۳۰	S042141
-۰/۰۰۸	۰/۱۱	یکنواخت	DIF	۰/۸۲	۰/۸۰	-۱/۰۴	۰/۸۰	S042215
-۰/۱۲	۰/۰۳	یکنواخت	DIF	-۱/۴۳	۱/۰۳	-۲/۶۸	۱/۰۳	S032606
۰/۲۳	۰/۰۵	یکنواخت	DIF	۰/۲۲	۰/۷۳	۱/۷۹	۰/۷۳	S032015
۰/۴۹	۰/۲۲	یکنواخت	DIF	۰/۶۴	۳/۱۰	۰/۴۰	۳/۱۰	S032310A
۰/۳۸	۰/۱۲	غیر یکنواخت	DIF	۰/۱۸	۵/۰۹	۰/۵۵	۴/۴۵	S032310B
۰/۱۲	۰/۰۳	غیر یکنواخت	DIF	-۰/۰۰۷	۳/۰۲	۰/۷۴	۴/۷۶	S032310D
-۰/۰۰۵	۰	یکنواخت	DIF	۰/۳۶	۰/۷۹	۰/۳۴	۰/۷۹	S032680
-۰/۱۵	۰/۰۳	یکنواخت	DIF	۰/۹۷	۰/۲۵	-۱/۶۰	۰/۲۵	S032672
-	-	-	لنگر	-۳/۲۸	۰/۷۴	-۳/۲۸	۰/۷۴	S032392
-۰/۰۷	۰/۰۰۵	یکنواخت	DIF	۱/۱۵	۰/۴۸	۰/۴۹	۰/۴۸	S032425
۰/۱۴	۰/۰۲	غیر یکنواخت	DIF	۰/۱۳	۰/۵۸	۳/۹۳	۰/۱۰	S032257
۰/۲۹	۰/۱۰	یکنواخت	DIF	-۱/۳۶	۰/۴۹	۱/۳۱	۰/۴۹	S032663

شخص CDIF	شخص اندازه اثر	نوع DIF	DIF یا لنگر	آمریکا		ایران		کد ستوال
				b	a	b	a	
-۰/۰۹	۰/۰۰۸	یکنواخت	DIF	۰/۲۷	۰/۵۴	-۰/۴۳	۰/۵۳	S032660
۰/۰۷	۰/۰۰۴	یکنواخت	DIF	۰/۶۸	۰/۸۶	۱/۰۷	۰/۸۶	S032555
۰/۰۳	۰/۰۰۱	یکنواخت	DIF	۱/۴۶	۰/۸۶	۱/۷۰	۰/۸۶	S032122
۰/۰۶	۰/۰۰۹	یکنواخت	DIF	-۰/۹۰	۰/۹۵	-۱/۳۷	۰/۹۵	S042053
۰/۰۵	۰/۰۰۷	یکنواخت	DIF	۰/۰۷	۰/۹۳	-۰/۳۱	۰/۹۳	S042408
۰/۰۶	۰/۰۰۸	یکنواخت	DIF	۰/۱۲	۰/۷۴	-۰/۴۲	۰/۷۴	S042015
-	-	-	لنگر	-۰/۲۴	۰/۴۲	-۰/۲۴	۰/۴۲	S042309
-۰/۱۴	۰/۰۷	غیر یکنواخت	DIF	-۲/۲۱	۰/۸۴	-۰/۵۹	۱/۴۷	S042049A
-۰/۰۵	۰/۰۰۷	یکنواخت	DIF	۰/۳۲	۱/۶۰	-۰/۰۷	۱/۶۰	S042049B
۰/۱۰	۰/۰۳	یکنواخت	DIF	-۱/۱۵	۱/۳۳	-۲/۰۷	۱/۳۳	S042182
۰/۰۹	۰/۰۲	یکنواخت	DIF	۲/۲۱	۰/۹۸	۱/۰۷	۰/۹۸	S042402
۰/۰۳	۰/۰۰۴	یکنواخت	DIF	۱/۳۰	۱/۶۴	۰/۹۷	۱/۶۴	S042228A
-	-	-	لنگر	-۰/۵۶	۱/۸۵	-۰/۵۶	۱/۸۵	S042228B
-۰/۱۲	۰/۰۴	یکنواخت	DIF	-۰/۱۳	۲/۴۳	۰/۴۵	۲/۲۱	S042228C
-۰/۰۶	۰/۰۱	یکنواخت	DIF	-۱/۴۸	۰/۶۸	-۰/۹۶	۰/۵۱	S042126
۰/۰۶	۰/۰۱	غیر یکنواخت	DIF	۱/۶۷	۰/۳۵	۰/۱۵	۰/۶۵	S042210
۰/۲۴	۰/۱۵	یکنواخت	DIF	۰/۷۴	۱/۴۴	-۰/۶۷	۱/۴۴	S042176
-۰/۰۷	۰/۰۲	غیر یکنواخت	DIF	-۰/۶۸	۱/۳۲	-۰/۲۶	۰/۸۰	S042211
۰/۱۵	۰/۰۶	یکنواخت	DIF	-۰/۲۸	۱/۴۹	-۱/۱۳	۱/۴۹	S042135
-	-	-	لنگر	-۰/۳۱	۰/۶۸	۰/۳۱	۰/۶۸	S042257
-۰/۰۰۳	۰/۰۰۴	یکنواخت	DIF	۰/۵۴	۰/۵۸	۱/۰۸	۰/۵۸	S032542
-۰/۰۰۷	۰/۰۵	یکنواخت	DIF	-۰/۳۸	۰/۹۷	۰/۶۸	۰/۹۷	S032645
-	-	-	لنگر	۰/۸۱	۰/۶۴	۰/۸۱	۰/۶۴	S032530A
۰/۰۰۶	۰/۰۳	یکنواخت	DIF	۱/۹۱	۰/۵۷	۰/۴۹	۰/۵۷	S032530B
۰/۰۰۱	۰/۰۰۴	یکنواخت	DIF	۲/۷۷	۰/۹۲	۱/۹۱	۰/۹۲	S032530D
-	-	-	لنگر	۰/۰۸	۱/۲۷	۰/۰۸	۱/۲۷	S032007
۰/۰۰۴	۰/۰۰۹	یکنواخت	DIF	-۰/۰۶	۰/۸۹	-۰/۵۲	۰/۸۹	S032502
-۰/۰۰۲	۰/۰۰۶	غیر یکنواخت	DIF	۱/۸۸	۰/۶۸	۰/۷۰	۱/۳۹	S032679
-	-	-	لنگر	۰/۷۴	۰/۲۸	۰/۷۴	۰/۲۸	S032184
۰/۰۰۸	۰/۰۳	یکنواخت	DIF	۰/۹۸	۰/۵۸	-۰/۴۳	۰/۵۸	S032394
-	-	-	لنگر	-۰/۲۳	۱/۳۷	-۰/۲۳	۱/۳۷	S032151
۰/۰۰۴	۰/۰۵	غیر یکنواخت	DIF	-۰/۲۰	۲/۶۴	۰/۷۷	۱/۳۰	S032651A
-۰/۰۰۱	۰/۰۰۲	یکنواخت	DIF	۱/۰۳	۱/۱۵	۱/۳۱	۱/۱۵	S032651B
-	-	-	لنگر	۰/۳۶	۱/۴۸	۰/۳۶	۱/۴۸	S032665A

شخص CDIF	شخص اندازه اثر	نوع DIF	DIF یا لنگر	آمریکا		ایران		کد سوال
				b	a	b	a	
-	-	-	لنگر	۰/۷۸	۱/۹۵	۰/۷۸	۱/۹۵	S032665B
۰	۰/۰۰۱	یکنواخت	DIF	۰/۸۱	۱/۸۳	۰/۶۸	۱/۸۳	S032665C
۰/۱۱	۰/۰۵	یکنواخت	DIF	-۲/۸۷	۰/۵۸	-۰/۸۸	۰/۵۸	S042073
۰/۰۴	۰/۰۲	غیر یکنواخت	DIF	۰/۴۷	۰/۸۳	۴/۰۹	۰/۲۳	S042017
-۰/۰۳	۰/۰۰۵	یکنواخت	DIF	۰/۸۵	۰/۹۴	۰/۴۷	۰/۹۴	S042007
-	-	-	لنگر	۴/۵۴	۰/۱۲	۴/۵۴	۰/۱۲	S042024
-	-	-	لنگر	-۱/۲۲	۱/۱۱	-۱/۲۲	۱/۱۱	S042095
۰/۱۳	۰/۰۸	یکنواخت	DIF	-۰/۴۰	۰/۹۷	۰/۹۲	۰/۹۷	S042022
-	-	-	لنگر	-۲/۲۶	۱/۲۸	-۲/۲۶	۱/۲۸	S042063
۰/۰۶	۰/۰۳	غیر یکنواخت	DIF	-۰/۴۰	۰/۹۴	۰/۸۵	۰/۴۲	S042197
۰/۰۸	۰/۰۴	یکنواخت	DIF	۱/۲۱	۰/۹۶	۳/۶۹	۰/۹۵	S042297
-	-	-	لنگر	۲/۶۰	۱/۰۷	۲/۶۰	۱/۰۷	S042305
۰/۱۸	۰/۱۵	یکنواخت	DIF	-۱/۲۰	۰/۹۲	۰/۶۹	۰/۹۲	S042112
۰/۰۳	۰/۰۰۴	یکنواخت	DIF	-۱/۲۰	۰/۸۵	-۰/۸۳	۰/۸۵	S042173A
-۰/۰۸	۰/۰۳	یکنواخت	DIF	۱/۰۱	۰/۶۰	-۰/۱۹	۰/۶۰	S042173B
-۰/۰۲	۰/۰۱	غیر یکنواخت	DIF	-۱/۱۳	۰/۴۸	-۰/۸۶	۱/۲۸	S042173C
۰/۰۳	۰/۰۰۸	غیر یکنواخت	DIF	-۰/۶۳	۰/۹۳	-۰/۳۴	۱/۶۹	S042173D
-۰/۰۴	۰/۰۰۷	غیر یکنواخت	DIF	-۱/۰۶	۱/۱۰	-۱/۱۵	۱/۵۹	S042173E
-۰/۰۲	۰/۰۳	غیر یکنواخت	DIF	۰/۵۹	۱/۶۱	۰/۱۱	۳/۵۴	S042173
-۰/۱۰	۰/۰۵	یکنواخت	DIF	۲/۱۴	۰/۴۴	-۰/۰۶	۰/۴۴	S042407
-۰/۰۲	۰/۰۰۱	یکنواخت	DIF	۰/۷۲	۱/۱۴	۰/۵۶	۱/۱۴	S042278
-۰/۰۷	۰/۰۳	یکنواخت	DIF	۳/۰۴	۰/۶۴	۱/۳۳	۰/۶۴	S042274
-	-	-	لنگر	۴/۲۶	۳/۸۰	۴/۲۶	۳/۸۰	S042311
-۰/۰۴	۰/۰۰۸	یکنواخت	DIF	-۰/۳۷	۰/۹۶	-۰/۷۸	۰/۹۶	S042317
-	-	-	لنگر	-۱/۵۲	۱/۱۱	-۱/۵۲	۱/۱۱	S032465
۰/۰۵	۰/۰۲	غیر یکنواخت	DIF	-۰/۴۰	۱/۰۸	-۱/۱۴	۰/۳۵	S032315
-	-	-	لنگر	۰/۴۲	۰/۹۸	۰/۴۲	۰/۹۸	S032306
-۰/۰۲	۰/۰۰۴	غیر یکنواخت	DIF	۱/۴۳	۰/۵۹	-۰/۹۵	۰/۹۲	S032640
۰/۰۶	۰/۰۰۷	یکنواخت	DIF	۰/۶۲	۰/۴۲	-۰/۲۳	۰/۴۲	S032579
۰/۰۹	۰/۰۲	غیر یکنواخت	DIF	۰/۱۴	۰/۸۲	-۰/۵۳	۱/۲۷	S032570
۰/۱۵	۰/۰۵	غیر یکنواخت	DIF	۱/۹۰	۰/۳۷	-۰/۴۵	۰/۷۶	S032024
۰/۰۹	۰/۰۲	یکنواخت	DIF	۲/۵۴	۱	۱/۲۰	۱	S032272
-	-	-	لنگر	۰/۸۰	۰/۸۶	۰/۸۰	۰/۸۶	S032141
-	-	-	لنگر	-۰/۸۵	۱/۳۲	-۰/۸۵	۱/۳۲	S032060

شخص CDIF	شخص اندازه اثر	نوع DIF	DIF یا لنگر	آمریکا		ایران		کد ستوال
				b	a	b	a	
۰/۱۱	۰/۰۲	یکنواخت	DIF	-۰/۴۰	۱/۲۶	-۰/۹۷	۱/۲۶	S032463
-	-	-	لنگر	-۰/۵۲	۲/۵۴	-۰/۵۲	۲/۵۴	S032650A
-	-	-	لنگر	-۰/۱۶	۴/۵۶	-۰/۱۶	۴/۵۶	S032650B
-	-	-	لنگر	۰/۰۳	۳/۶۰	۰/۰۳	۳/۶۰	S032650D
۰/۰۰۵	۰/۰۰۶	غیر یکنواخت	DIF	۰/۰۷	۰/۶۲	۱/۰۹	۰/۱۸	S032514
-۰/۰۵	۰/۰۳	یکنواخت	DIF	-۱/۶۴	۰/۸۸	-۰/۷۱	۰/۸۸	S042042
-۰/۰۴	۰/۰۳	یکنواخت	DIF	۰/۳۸	۱/۲۸	۱/۱۴	۱/۲۸	S042030
۰/۰۳	۰/۰۲	غیر یکنواخت	DIF	۰/۴۷	۰/۴۶	-۰/۴۶	۰/۸۳	S042003
۰/۰۳	۰/۰۰۹	یکنواخت	DIF	-۱/۷۰	۰/۶۷	-۲/۵۰	۰/۶۷	S042110
-	-	-	لنگر	۰/۶۲	۱/۲۹	۰/۶۲	۱/۲۹	S042222A
-	-	-	لنگر	۰/۱۲	۱/۵۰	۰/۱۲	۱/۵۰	S042222B
-۰/۰۶	۰/۰۵	یکنواخت	DIF	-۱/۵۵	۰/۸۶	-۰/۴۰	۰/۸۶	S042222C
۰/۰۵	۰/۰۴	غیر یکنواخت	DIF	-۱/۲۶	۰/۹۴	-۱/۷۳	۱/۷۴	S042065
-	-	-	لنگر	-۰/۴۴	۱/۳۵	-۰/۴۴	۱/۳۵	S042280
-	-	-	لنگر	۰/۹۵	۰/۵۹	-۰/۹۵	۰/۵۹	S042088
۰/۰۴	۰/۰۳	غیر یکنواخت	DIF	-۰/۵۰	۱/۳۳	-۱/۵۶	۰/۶۳	S042218
۰/۰۲	۰/۰۱	غیر یکنواخت	DIF	۱/۳۹	۱/۰۶	۰/۶۲	۱/۴۷	S042104
-۰/۰۸	۰/۰۹	یکنواخت	DIF	۰/۱۵	۱/۴۵	۱/۶۰	۱/۴۵	S042064
۰/۰۰۷	۰/۰۰۴	غیر یکنواخت	DIF	۰/۱۵	۱/۳۱	-۰/۰۵	۱/۸۲	S042273
۰/۰۰۲	۰/۰۰۱	غیر یکنواخت	DIF	-۰/۵۶	۱/۲۷	-۰/۵۹	۱/۱۱	S042301
۰/۰۵	۰/۰۳	یکنواخت	DIF	-۰/۳۴	۰/۳۴	-۲/۳۵	۰/۳۴	S042312
۰/۰۶	۰/۰۴	یکنواخت	DIF	۰/۲۵	۰/۹۶	-۰/۶۲	۰/۹۶	S042217
۰/۰۳	۰/۰۲	غیر یکنواخت	DIF	۰/۷۶۵	۱/۱۷	۰/۱۳	۱/۵۶	S042406
-	-	-	لنگر	۲/۲۰	۰/۲۹	۲/۲۰	۰/۲۹	S032611
-	-	-	لنگر	-۰/۸۵	۱/۱۱	-۰/۸۵	۱/۱۱	S032614
-۰/۱۰	۰/۰۹	یکنواخت	DIF	-۰/۸۹	۱/۲۳	۰/۲۲	۱/۲۳	S032451
-۰/۰۷	۰/۰۵	یکنواخت	DIF	-۰/۲۲	۰/۷۳	۱/۱۶	۰/۷۳	S032156
۰/۰۵	۰/۰۲	یکنواخت	DIF	۰/۳۷	۱/۱۲	-۰/۲۱	۱/۱۲	S032056
۰/۱۱	۰/۱۰	غیر یکنواخت	DIF	۰/۰۰۷	۱/۱۹	-۱/۸۵	۰/۶۸	S032087
۰/۰۳	۰/۰۰۶	یکنواخت	DIF	۱/۱۳	۰/۵۹	۰/۵۱	۰/۵۹	S032279
۰/۰۷	۰/۰۴	یکنواخت	DIF	۱/۲۰	۱/۰۲	۰/۱۳	۱/۰۲	S032238
۰/۰۲	۰/۰۱	یکنواخت	DIF	۱/۵۰	۱/۷۸	۰/۹۸	۱/۷۸	S032369
-۰/۰۳	۰/۰۰۸	یکنواخت	DIF	۱/۲۱	۰/۷۰	-۰/۶۵	۰/۷۰	S032160
۰/۰۵	۰/۰۲	یکنواخت	DIF	-۰/۲۸	۰/۹۱	-۰/۹۳	۰/۹۲	S032654

کد سوال	ایران		آمریکا		DIF یا لنگر	نوع DIF	شاخص اندازه اثر	شاخص CDIF
	a	b	a	b				
S032126	۰/۷۹	-۰/۴۰	۰/۷۹	-۰/۴۰	لنگر	-	-	-
S032510	۰/۸۳	-۱/۱۸	۱/۳۶	-۰/۹۰	DIF	غیر یکنواخت	۰/۰۰۳	۰
S032158	۱/۱۶	-۰/۹۱	۱/۱۶	-۰/۹۱	لنگر	-	-	-
S042258	۰/۷۴	-۰/۴۶	۰/۷۴	۱/۰۹	DIF	یکنواخت	۰/۰۶	۰/۱۱
S042005	۰/۵۰	-۰/۷۶	۰/۵۰	-۰/۶۸	DIF	یکنواخت	۰	۰/۰۰۴
S042016	۰/۸۹	۱/۶۸	۰/۸۹	۰/۹۴	DIF	یکنواخت	۰/۰۰۹	-۰/۰۴
S042300A	۱/۹۲	۰/۱۶	۱/۹۲	۰/۱۶	لنگر	-	-	-
S042300B	۰/۴۷	۴/۰۵	۰/۴۷	۴/۰۵	لنگر	-	-	-
S042300C	۰/۴۷	۲/۵۱	۱/۲۱	۰/۰۴	DIF	یکنواخت	۰/۰۴	-۰/۰۵
S042319	۱/۹۸	۰/۹۹	۱/۹۸	۰/۲۶	DIF	یکنواخت	۰/۰۳	-۰/۰۵
S042068	۰/۴۵	۰/۳۴	۰/۴۴	۳/۳۹	DIF	یکنواخت	۰/۰۷	۰/۱۱
S042216	۰/۸۰	-۱/۰۸	۰/۸۰	-۰/۴۶	DIF	یکنواخت	۰/۰۱	۰/۰۶
S042249	۱/۱۳	-۰/۹۹	۱/۱۳	-۰/۵۴	DIF	یکنواخت	۰/۰۱	۰/۰۶
S042094	۱/۱۵	۰/۹۹	۱/۱۵	۰/۳۴	DIF	یکنواخت	۰/۰۲	-۰/۰۵
S042293A	۱/۴۸	-۱/۱۸	۱/۴۸	-۱/۱۸	لنگر	-	-	-
S042293B	۱/۹۲	۰/۷۱	۰/۹۳	۲/۶۱	DIF	غیر یکنواخت	۰/۰۲	۰/۰۲
S042195	۰/۸۲	۲/۱۷	۰/۸۲	۳/۱۲	DIF	یکنواخت	۰/۰۰۴	۰/۰۲
S042400	۱/۲۵	۱/۴۹	۱/۲۵	۱/۷۲	DIF	یکنواخت	۰/۰۰۱	۰/۰۰۷
S042401	۳/۳۳	۴/۴۴	۳/۳۳	۴/۴۴	لنگر	-	-	-
S042164	۱/۰۵	-۰/۶۶	۱/۷۷	-۰/۳۶	DIF	غیر یکنواخت	۰/۰۲	۰/۰۶

بررسی یافته‌های جدول ۴ نشان می‌دهد که از ۲۲۸ سوال مورد مطالعه، ۵۹ سوال لنگر و ۱۶۹ (۷۴٪) سوال دارای DIF است که ۵۱ سوال دارای DIF غیر یکنواخت و ۱۱۸ سوال دارای DIF یکنواخت است.

از ۱۶۹ سوال دارای DIF، ۶۴ سوال (۳۸٪) به نفع گروه دانش آموزان ایران و ۱۰۵ سوال (۶۲٪) به نفع گروه دانش آموزان آمریکا از کارکرد افتراقی برخوردارند. در جدول ۵ تعداد سوال‌هایی که به نفع هر کدام از گروه‌ها دارای DIF یکنواخت یا غیر یکنواخت است ارائه شده است.

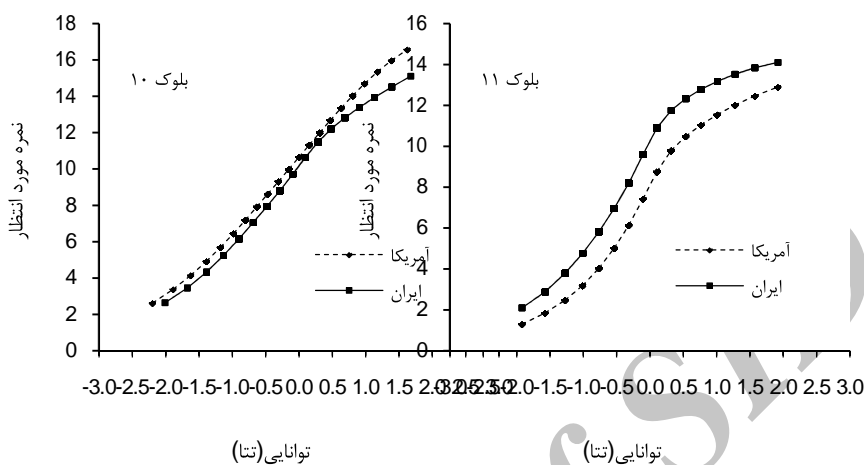
جدول ۵: فراوانی و درصد سئوال‌های دارای DIF

کشور	نوع DIF	
	یکنواخت	غیریکنواخت
ایران	۵۰ (۳۰٪)	۱۴ (۸٪)
آمریکا	۶۸ (۴۰٪)	۳۷ (۲۲٪)

در جدول ۶ شاخص DTF که کارکرد افتراقی آزمون (کل سئوال‌های یک مجموعه) را نشان می‌دهد به تفکیک بلوک ارائه شده است. آزمون χ^2 بدلیل حساسیت به حجم نمونه، روش مناسبی برای آزمون معناداری شاخص DTF نیست، بنابراین از نقاط برش که در اولین ستون سمت چپ جدول ارائه شده است برای این منظور استفاده می‌شود. بررسی یافته‌های جدول ۶ نشان می‌دهد که به غیر از بلوک‌های ۲، ۶ و ۹، سایر بلوک‌ها از کارکرد افتراقی برخوردارند. به عبارت دیگر، ۱۱ بلوک از ۱۴ بلوک دارای کارکرد افتراقی است. در واقع، ۷۹٪ بلوک‌ها بر له یا علیه گروه دانش‌آموزان ایران دارای کارکرد افتراقی هستند.

جدول ۶: شاخص DTF همراه با نقاط برش به تفکیک بلوک

بلوک	DTF	ریشه دوم DFT	مقدار χ^2	df	P	نقطه برش DTF
۱	۱/۹۹	۱/۴۱	۴۶۰۰۲/۰۲	۵۷۱	۰/۰۰۰	۰/۰۷۲
۲	۰/۰۰۷	۰/۰۸۴	۳۱۲۶	۵۷۲	۰/۰۰۰	۰/۰۷۲
۳	۰/۲۸	۰/۵۳	۵۴۷۴۵	۵۶۴	۰/۰۰۰	۰/۰۴۸
۴	۰/۲۵	۰/۵۰	۳۱۱۰/۱۲	۵۶۹	۰/۰۰۰	۰/۰۷۸
۵	۰/۵۱	۰/۷۲	۳۱۶۹۵/۸۴	۵۶۶	۰/۰۰۰	۰/۰۶
۶	۰/۰۰۴	۰/۰۶	۶۱۹/۶۱	۵۶۰	۰/۰۴	۰/۰۷۸
۷	۱/۳۲	۱/۴۵	۱۶۶۹/۱۵	۵۵۷	۰/۰۰۰	۰/۰۷۸
۸	۰/۳۹	۰/۶۲	۶۶۴۹۵/۸۵	۵۶۸	۰/۰۰۰	۰/۰۸۴
۹	۰/۰۱	۰/۱۰	۷۵۳/۷۴	۵۷۶	۰/۰۰۰	۰/۰۶
۱۰	۰/۲۵	۰/۵۰	۴۴۶۶/۱۶	۵۷۴	۰/۰۰۰	۰/۱۰۲



۰/۰۴۸	۰/۰۰۰	۵۷۹	۶۰۲۸۶/۱۱	۰/۷۳	۰/۵۳	۱۱
۰/۰۸۴	۰/۰۰۰	۵۶۲	۲۳۱۸۲/۸۸	۰/۳۰	۰/۰۹	۱۲
۰/۰۶	۰/۰۰۰	۵۶۱	۳۱۹۱۲/۲۰	۰/۳۴	۰/۱۱	۱۳
۰/۰۷۸	۰/۰۰۰	۵۶۹	۳۴۹۹/۷۱	۰/۵۱	۰/۲۶	۱۴

شکل ۳: نمودار تابع پاسخ آزمون (بلوک)، TRF، برای دو گروه ایران و آمریکا

شکل ۳ نیز کارکرد افتراقی آزمون یا بلوک‌ها (DTF) را به صورت نموداری برای ۵ بلوک که ۴ بلوک آن بر له یا علیه دانش‌آموزان ایران از کارکرد افتراقی یکنواخت و ۱ بلوک نیز دارای کارکرد افتراقی غیر یکنواخت است نشان می‌دهد. به طور کلی، ۳۷٪ از ۱۴ بلوک به نفع گروه هدف (دانش‌آموزان ایران) و ۲۱٪ بلوک‌ها به نفع گروه مرجع (دانش‌آموزان آمریکا) از کارکرد افتراقی یکنواخت برخوردارند. ۲۱٪ بلوک‌ها دارای کارکرد افتراقی غیریکنواخت و ۲۱٪ مابقی فاقد کارکرد افتراقی هستند.

بحث و نتیجه‌گیری

در مباحث مربوط به روش‌شناسی سنجش‌های بین‌المللی، معادل‌بودن و مقایسه‌پذیری از اهمیت اساسی برخوردار است (ارسی‌کان، ۱۹۹۸). با توجه به اهمیت و ضرورت

تغییرناپذیری سازه‌ای آزمون‌هایی که در مطالعات بین فرهنگی و بین‌المللی بکار می‌روند (به عنوان مثال برن و استوارت، ۲۰۰۶؛ لیتل، ۱۹۹۷) و با توجه به یافته‌های مربوط به تغییرناپذیری سازه‌ای و کارکرد افتراقی سؤال‌های ریاضیات و علوم تیمز (به عنوان مثال به ارسی کان، ۱۹۹۸؛ ارسی کان و کاه، ۲۰۰۵؛ وو، لی، و زامبو، ۲۰۰۷) پژوهش حاضر تلاش کرد تا به سؤال‌های زیر پاسخ دهد:

۱- آیا ساختار عاملی/سازه‌ای آزمون علوم تیمز ۲۰۰۷ در بین دانش‌آموزان ایران و آمریکا یکسان و قابل مقایسه است؟

۲- آیا سؤال‌های آزمون علوم تیمز ۲۰۰۷ بر له یا علیه دانش‌آموزان ایرانی دارای DIF است و به چه میزان؟

۳- تأثیر سؤال‌های دارای DIF بر عملکرد دانش‌آموزان ایرانی در آزمون علوم پایه هشتم تیمز ۲۰۰۷، چگونه و به چه میزان است؟

به طور کلی نتایج حاصل از تحلیل‌های مربوط به سؤال نشان می‌دهد که تمام ۱۴ بلوک، از تغییرناپذیری شکلی در بین دو گروه ایران و آمریکا برخوردارند. برقرار بودن تغییرناپذیری شکلی بدان معنا است که دانش‌آموزان دو کشور ایران و آمریکا برای دادن پاسخ به سؤال‌های علوم تیمز ۲۰۰۷ از چارچوب مفهومی یکسانی استفاده می‌کنند (چونگ و رنولد، ۲۰۰۲؛ هورن و مک آردل، ۱۹۹۲) و این آزمون در هر دو جامعه، سازه واحد و یکسانی را اندازه می‌گیرد. این یافته، هر چند که به درس علوم تیمز ۲۰۰۷ مربوط می‌شود، لیکن با نتایج حاصل از پژوهش وو، لی و زامبو (۲۰۰۷) که درس ریاضی تیمز ۱۹۹۹ را مورد مطالعه قرار دادند، از همخوانی برخوردار است. آن‌ها نیز در پژوهش خود به این نتیجه رسیدند که سؤال‌های ریاضی تیمز ۱۹۹۹ از تغییرناپذیری شکلی در بین کشورهای مورد مطالعه برخوردارند.

یافته‌های حاصل از تحلیل‌های مربوط به سؤال دوم نشان داد که از ۲۲۸ سؤال مورد مطالعه، ۵۹ سؤال لنگر و ۱۶۹ سؤال دارای DIF است که ۵۱ سؤال دارای DIF غیریکنواخت و ۱۱۸ سؤال دارای DIF یکنواخت است. از ۱۶۹ سؤال دارای DIF، ۶۴

سئوال (۳۸٪) به نفع گروه دانش‌آموزان ایران و ۱۰۵ سئوال (۶۲٪) به نفع گروه دانش‌آموزان آمریکا، یا به عبارت دیگر علیه دانش‌آموزان ایران از کارکرد افتراقی برخوردارند. این یافته با نتایج حاصل از پژوهش ارسی کان و کاه (۲۰۰۵) و ارسی کان و مک کریس (۲۰۰۲) از همخوانی برخوردار است. ارسی کان و کاه (۲۰۰۵) در پژوهش خود دریافتند که ۳۷٪ از سئوال‌های علوم تیمز ۱۹۹۵ در بین دانش‌آموزان انگلیسی زبان و فرانسوی زبان کانادا از DIF برخوردارند، در حالیکه درصد سئوال‌های دارای DIF در بین دانش‌آموزان آمریکایی و فرانسوی ۷۹٪ بود. همچنین آن‌ها دریافتند که ۳۹٪ سئوال‌ها در بین دانش‌آموزان کشور انگلیس و فرانسه از DIF برخوردارند. در واقع، آن‌ها دریافتند که از بین سئوال‌های مورد مطالعه در درس علوم، تعداد ۵۲ سئوال در بین دانش‌آموزان انگلیسی و فرانسوی زبان کانادا، تعداد ۵۴ سئوال در بین دانش‌آموزان کشور انگلیس و فرانسه و تعداد ۱۱۰ سئوال در بین دانش‌آموزان کشور فرانسه و آمریکا از DIF برخوردارند. در تمام این ۳ مقایسه‌ای که این پژوهشگران انجام دادند، بیش از نیمی از سئوال‌های دارای DIF، به نفع گروه هدف، یعنی دانش‌آموزان فرانسوی زبان عمل می‌کردند. زیاد بودن تعداد سئوال‌های دارای DIF در بین دانش‌آموزان آمریکا و فرانسه بیانگر این است که هم عامل زبان و هم عوامل فرهنگی، اجتماعی و آموزشی بر نحوه عملکرد سئوال تأثیر مهمی دارند. تعداد سئوال‌های دارای DIF در بین دانش‌آموزان انگلیسی و فرانسوی زبان کانادا که از شرایط فرهنگی، اجتماعی و آموزشی مشابهی برخوردارند در مقایسه با تعداد سئوال‌های دارای DIF در بین دانش‌آموزان کشور آمریکا و فرانسه، بسیار کمتر است. در پژوهش حاضر نیز درصد بالایی از سئوال‌ها (۷۴٪) از DIF برخوردار بودند که می‌تواند همانند مورد فوق ناشی از تفاوت‌های عمیق فرهنگی، زبان‌شناختی، اجتماعی، اقتصادی، آموزشی و حتی سیاسی باشد.

در پژوهش ارسی کان و کاه (۲۰۰۵) در درس ریاضی نیز وضعیت و الگوی سئوال‌های دارای DIF مشابه درس علوم بود. در این درس نیز تعداد سئوال‌های دارای DIF در بین دانش‌آموزان کشور آمریکا و فرانسه بیشتر از تعداد این سئوال‌ها در بین دانش‌آموزان

انگلیسی زبان و فرانسوی زبان کانادا و همچنین بیشتر از تعداد این سؤال‌ها در بین دانش‌آموزان کشور انگلیس و فرانسه بود.

ارسی کان و مک کریس (۲۰۰۲) نیز دریافتند که در درس ریاضی تیمز ۱۹۹۵، از ۱۵۶ سؤال مورد مطالعه، ۲۲ سؤال در بین دانش‌آموزان فرانسوی و انگلیسی زبان کشور کانادا، ۶۱ سؤال در بین دانش‌آموزان کشور فرانسه و انگلستان، و از بین ۱۵۴ سؤال، ۹۱ سؤال در بین دانش‌آموزان کشور فرانسه و آمریکا دارای DIF هستند. تقریباً ۴۵٪ سؤال‌های دارای DIF به نفع دانش‌آموزان فرانسوی زبان عمل می‌کردند. در درس علوم در مقایسه با درس ریاضی نسبت بیشتری از سؤال‌ها دارای DIF بودند. از ۱۴۰ سؤال مورد مطالعه، تعداد ۵۲ سؤال در بین دانش‌آموزان فرانسوی و انگلیسی زبان کانادا، تعداد ۵۴ سؤال در بین دانش‌آموزان کشور فرانسه و انگلستان، و ۹۱ سؤال در بین دانش‌آموزان کشور فرانسه و آمریکا دارای DIF بودند. در واقع، بین ۵۴ تا ۶۵ درصد سؤال‌های دارای DIF به نفع دانش‌آموزان فرانسوی زبان عمل می‌کردند. این پژوهشگران دریافتند که در ۶۴ سؤال از ۲۹۶ سؤال تیمز، مشکلات مرتبط با ترجمه و انطباق وجود دارد. ارسی کان و مک کریس به این نتیجه رسیدند که در حدود ۳۵ تا ۴۰ درصد سؤال‌های دارای DIF ناشی از مشکلات ترجمه و انطباق سؤال‌ها و تفاوت‌ها در برنامه‌های درسی کشورها است.

وجود DIF ناشی از ترجمه و انطباق سؤال‌ها، تفاوت‌های فرهنگی، اجتماعی، زبان‌شناختی و آموزشی در یک سؤال، به آن معناست که سؤال، ویژگی مورد نظر را که برای اندازه‌گیری آن تهیه شده است اندازه نمی‌گیرد و از این رو، بر روی آزمون و استنباط‌هایی که از نمره‌های آن به عمل می‌آید تأثیر منفی می‌گذارد. یکی از راه‌های بهبود کیفیت آزمون و کاهش DIF این است که تعدادی از سؤال‌ها از همان ابتدا به زبان مادری دانش‌آموزان کشورهای مشارکت‌کننده طراحی و اجرا شوند. علاوه بر این، برای اجتناب از مشکلات مرتبط با ترجمه، انجام تحلیل‌های DIF در مرحله مقدماتی پرورش آزمون یا حتی پس از اجرای آزمون می‌تواند سودمند باشد. در مرحله مقدماتی، سؤال‌هایی که دارای DIF تشخیص داده می‌شوند را می‌توان حذف یا اصلاح کرد. از نتایج حاصل از

تحلیل‌های DIF پس از اجرای آزمون نیز می‌توان برای شناسایی سؤال‌های دارای DIF و حذف آنها به هنگام انجام مقایسه‌های بین‌المللی استفاده نمود.

برای پاسخگویی به سؤال سوم از شاخص DTF که تأثیر DIF را بر عملکرد دانش‌آموزان در آزمون نشان می‌دهد، استفاده گردید. یافته‌های حاصل از این تحلیل نشان داد که به غیر از بلوک‌های ۲، ۶ و ۹ سایر بلوک‌ها از کارکرد افتراقی برخوردارند. به عبارت دیگر، ۱۱ بلوک از ۱۴ بلوک دارای کارکرد افتراقی است. در واقع، ۷۹٪ بلوک‌ها دارای کارکرد افتراقی هستند که ۳۷٪ به نفع گروه هدف (دانش‌آموزان ایران) و ۲۱٪ به نفع گروه مرجع (دانش‌آموزان آمریکا) عمل می‌کنند. ۲۱٪ بلوک‌ها نیز دارای کارکرد افتراقی غیریکنواخت و ۲۱٪ مابقی فاقد DIF کارکرد افتراقی هستند.

در زیر محتوای دو سؤال S042173 و S032310 که از DIF یکنواخت علیه دانش‌آموزان ایران برخوردارند ارائه شده است. سؤال S032310 که در دفترچه شماره ۷ به عنوان سؤال شماره ۳۲ آمده است به این شکل است. "گیاهان سبز هنگام فتوسنتز، اکسیژن و غذا تولید می‌کنند. کلروفیل یکی از چیزهایی است که برای فتوسنتز لازم است. نام دو عامل دیگر که برای انجام فتوسنتز لازم هستند را بنویسید". سؤال S042173 نیز که در دفترچه شماره ۹ به عنوان سؤال شماره ۴۹ آمده است به این شکل است. "وقتی که یک مایع به گاز تبدیل می‌شود، چه ویژگی‌ها یا خاصیت‌هایی تغییر می‌کند و چه ویژگی‌ها یا خاصیت‌هایی ثابت می‌ماند". در ستون مناسب در جدول زیر علامت (X) بگذارید.

تغییر می‌کند	ثابت می‌ماند
	چگالی
	جرم
	حجم
	اندازه‌ی مولکول‌ها
	سرعت مولکول‌ها

به منظور شناسایی علل احتمالی کارکرد افتراقی دو سؤال فوق، از تعدادی از معلمان درس علوم پایه هشتم و همچنین از تعدادی از دبیران زیست‌شناسی، نظر خواهی به عمل

آمد. به طور خلاصه، به نظر این افراد شاید بتوان گفت که عدم تسلط خود معلمان بر مفاهیم مطرح شده در این سئوال‌ها و نداشتن توانایی در برقراری ارتباط بین این مفاهیم، عدم آموزش عملی و آزمایشگاهی این مفاهیم و اکتفا کردن به آموزشهای صرفاً نظری، ضعف محتوای دروس در آموزش و ایجاد مهارت تفکر واگرا و چندبُعدی در دانش‌آموزان و همچنین گسترش فرهنگ آپارتمان‌نشینی که باعث ضعیف شدن ارتباط و تعامل دانش‌آموزان با طبیعت شده است را می‌توان از جمله علل احتمالی عملکرد ضعیف دانش‌آموزان ایرانی در این نوع سئوال‌ها محسوب کرد.

به طور کلی نتایج این پژوهش نشان می‌دهد، اگرچه ۶۲٪ سئوال‌های درس علوم تیمز ۲۰۰۷ علیه دانش‌آموزان ایران از کارکرد افتراقی برخوردارند، لیکن این کارکرد افتراقی تأثیر چندانی بر عملکرد دانش‌آموزان ایران و رتبه ایران در تیمز ندارد، زیرا از ۱۱ بلوک دارای کارکرد افتراقی، تعداد ۶ بلوک به نفع دانش‌آموزان ایران و ۳ بلوک به نفع دانش‌آموزان آمریکا از کارکرد افتراقی یکنواخت و ۲ بلوک دیگر نیز از کارکرد افتراقی غیر یکنواخت برخوردارند. این یافته با نتیجه حاصل از پژوهش بیتون (۱۹۹۸)، و ارسی کان و مک کریس (۲۰۰۲) هماهنگی دارد. بیتون در پژوهش خود که بر اساس داده‌های تیمز ۱۹۹۵ صورت گرفت به این نتیجه رسید که عادلانه بودن یا نبودن سئوال‌ها، هر چند ممکن است جایگاه یک کشور را به میزان چند رتبه تغییر دهد اما باعث نمی‌شود که رتبه یک کشور از ضعیف به متوسط یا از متوسط به قوی تغییر یابد. او به طور کلی به این نتیجه رسید که استفاده یا عدم استفاده از سئوال‌هایی که مناسب یک کشور هستند تأثیر بسیار اندکی بر جایگاه نسبی یا رتبه کشورها دارد و تأثیر اساسی بر تصویر کلی و عملکرد کلی کشورها نمی‌گذارد.

ارسی کان و مک کریس (۲۰۰۲) نیز به این نتیجه رسیدند که تأثیر سئوال‌های دارای DIF بر عملکرد دانش‌آموزان کشورهای مورد مقایسه به اندازه کافی بزرگ نیست که بتواند جایگاه یک کشور را به شکل اساسی تغییر بدهد.

به طور خلاصه، چنین به نظر می‌رسد که بر اساس یافته‌های پژوهش حاضر می‌توان

نتیجه گرفت که عملکرد ضعیف دانش‌آموزان ایران در درس علوم تیمز ۲۰۰۷ را نمی‌توان به کارکرد افتراقی سؤال‌ها نسبت داد و بایستی علل آن را جایی دیگر جستجو کرد. با این حال، وجود سؤال‌های دارای DIF می‌تواند مقایسه‌پذیری نمره‌ها را به شیوه‌های پیچیده‌تری تحت تأثیر قرار دهد. برای مثال، سؤال‌های دارای DIF ممکن است تأثیر یکنواختی بر عملکرد دانش‌آموزان نداشته باشند. در واقع، ممکن است مقایسه‌پذیری نمره‌های برخی از دانش‌آموزان، مانند دانش‌آموزان ضعیف، به میزان بیشتری تحت تأثیر قرار گیرد. از اینرو، به منظور افزایش روایی و صحت و درستی استنباطهایی که از داده‌های مطالعات تیمز در پژوهش‌ها به عمل می‌آید، پیشنهاد می‌شود که در مراحل مقدماتی تهیه آزمون‌های تیمز با انجام تحلیل‌های DIF سؤال‌های دارای DIF شناسایی شده و اصلاح یا حذف شوند. همچنین پیشنهاد می‌شود که دست‌اندرکاران و مسئولان ملی پروژه تیمز بایستی تلاش کنند تا برگزارکنندگان بین‌المللی تیمز را متقاعد کنند، تا به جای ترجمه سؤال‌ها از زبان انگلیسی به فارسی، اجازه دهند تعدادی از سؤال‌های نسخه فارسی آزمون‌های تیمز از همان ابتدا به زبان فارسی تهیه و اجرا شوند.

یافته ضمنی پژوهش حاضر این است که مسئولان وزارت آموزش و پرورش و متولیان امر آموزش در کشور بایستی تلاش کنند تا با تهیه محتواهای آموزشی مناسب، مفاهیم مهم هر حوزه را به صورت زنجیره‌وار و در ارتباط با یکدیگر آموزش دهند و تلاش کنند تا توانایی و مهارت تفکر واگرا و چند بُعدی را در دانش‌آموزان پرورش دهند. همچنین بایستی تلاش شود تا از روش سنتی آموزش که به شکل سخنرانی و عمدتاً نظری صورت می‌گیرد فاصله گرفت و دانش‌آموزان را درگیر فعالیتهای عملی و آزمایشگاهی نمود. در ارتباط با معلمان و دبیران نیز بایستی تلاش شود تا با برگزاری دوره‌های آموزشی، توانایی و مهارت معلمان را در حوزه‌ها و موضوعات مرتبط با دروسی که تدریس می‌کنند، افزایش داد تا بتوانند با برقراری ارتباط درست و دقیق بین مفاهیم و موضوعات مورد تدریس، آموخته‌های قبلی و یادگیریهای جدید دانش‌آموزان، یا به عبارت دیگر آموخته‌های جزیره‌وار و از هم گسیخته دانش‌آموزان، را به شکل شبکه‌ای نظام‌مند به یکدیگر متصل نمایند.

منابع

- کریمی، عبدالعظیم. (۱۳۸۸). مجموعه سئوال‌های علوم و ریاضیات تیمز TIMSS (پایه سوم راهنمایی). تهران: پژوهشگاه مطالعات آموزش و پرورش.
- Barrett, P. (2007). Structural Equation Modeling: Adjusting model fit. *Personality and Individual Differences*, 42, 815-824.
- Beaton, A. E. (1998). Comparing cross-national student performance on TIMSS using different test items. *International Journal of Educational Research*, 29, 529-542.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. NY: The Guilford Press.
- Byrne, B. M., & Stewart, S. M. (2003). The MACS approach to testing for multigroup invariance of a second-order factor structure: A walk through the process. *Structural Equation Modeling*, 13, 287-321.
- Campbell, H. L., Barry, C. L., Joe, J. N. & Finney, S. J. (2008). Configural, metric and scalar invariance of the modified achievement goal questionnaire across africanamerican and white university students. *Educational and Psychological Measurement*, 68, 988-1007.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233-255.
- DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, 2, 292-307.
- Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement*, 23, 355-368.
- Dorans, N. J., & Kulick, E. (2006). Differential item functioning on the Mini-Mental State Examination : An application of the Mantel-Haenszel and standardization procedures. *Medical Care*, 44(Suppl 3), S107-S114.

- Ercikan, K.(1998). Translation effects in international assessments. *International Journal of Educational Research*, 29, 543-553.
- Ercikan, K., & Koh, K. (2005). Examining the construct comparability of the English and French versions of TIMSS. *International Journal of Testing*, 5, 23-35.
- Ercikan, K., & McCreith, T. (2002). Effects of adaptations on comparability of test items and test scores. In D. F. Robitaille & A. E. Beaton, (Eds.), *Secondary Analysis of the TIMSS Data*(391-405). Kluwer Academic Publishers.
- Fleer, P.F.(1993) A Monte Carlo assessment of a new measure of item and test bias. Illinois Institute of Technology. *Dissertation Abstracts International*. 54(04B):2266.
- Flowers, C. P., Oshima, T. C., & Raju, N.S.(1999). A description and demonstration of the polytomous DFIT framework. *Applied Psychological Measurement*. 23,309-32.
- Frase, C.(1988). NOHARM: An IBM PC program for fitting both unidimensional and multidimensional normal ogive models of latent trait theory. Armidale, Australia:The University of New England.
- Floyd; F.J. , & Widaman, K. F. (1995). Factor analysis in development and refinement of clinical assessment instrument . *Psychological Assessment*, 3, 286-299.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide in measurement invariance in aging research. *Experimental Aging Research*, 18, 117-144.
- Kline, P. (2005). Principles and practice of structural equation modeling (2nd ed.). New York: Guilford Press.
- Kline, P. (2011). Principles and practice of structural equation modeling (3rd ed.). New York: Guilford Press.
- Lee, Y-S., Park, Y-S., & Tayalan, D. (2011). A cognitive diagnostic modeling of attribute mastery in Massachusetts, Minnesota, and U.S national sample using TIMSS 2007. *International Journal of testing*, 11, 144-177.
- Little, T. D.(1997). Mean and covariance structures(MACS) analysis of cross-cultural data: Practical and theoretical issues. *Multivariate Behavioral Research*, 32, 53-76.
- Muthén, L. K., & Muthén, B. O. (1998-2010). *Mplus user's guide* (6th ed.). Los Angeles: Muthén & Muthén.
- Olson, J.F., Martin, M.O., & Mullis, I.V.S. (Eds.) (2008). Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College. Available in PDF format at: <http://timss.bc.edu/timss2007/techreport.html>
- Orlando-Edelen, M., Thissen, D., Teresi, J. A., Kleinman, M., & Ocepek-

- Welikson, K. (2006). Identification of differential item functioning using item response theory and the likelihood-based model comparison approach: Application to the Mini-mental status examination. *Medical Care, 44*, S134–S142.
- Raju, N. S. (1999). DFIT5P: A fortran program for calculating dichotomous DIF/DTF [computer program]. Chicago, IL: Illinois Institute of Technology.
- Raju, N. S., Van der Linden, W., & Fleer, P. (1995). An IRT-based internal measure of test bias with applications for differential item functioning. *Applied Psychological Measurement, 19*, 353-368.
- Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two Approaches for exploring measurement invariance. *Psychological Bulletin, 114*, 552-566.
- Teresi, J. A. (2006). Overview of quantitative measurement methods: Equivalence, invariance and differential item functioning in health applications. *Medical Care*, (Suppl 3), 44, S39-S49.
- Teresi, J. A., Kleinman, M., & Ocepek-Welikson, K. (2000). Modern psychometric methods for detection of differential item functioning: Application to cognitive assessment measures. *Statistics in Medicine, 19*, 1651-1683.
- Teresi, J. A., Ocepek-Welikson, K., Kleinman, M., Cook, K. F., et al. (2007). Evaluation measurement equivalence using the item response theory log-likelihood ratio (IRTLR) method to assess differential item functioning (DIF): Applications (with illustration) to measure physical functioning ability and general distress. *Quality Life Research, 16*(Suppl 1), 43-68.
- Teresi, J. A., Ocepek-Welikson, K., Kleinman, M., Eimicke, J. P., et al. (2009). Analysis of differential item functioning in the depression item bank from the Patient Reported Outcome Measurement Information System (PROMIS): An item response approach. *Psychology Science Quarterly, 51*, 148-180.
- Thissen, D. (2001). *IRTLRDIF* (version 2.02b): Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning [Computer program]. Chapel Hill, NC: L. L. Thurstone Psychometric Laboratory.
- Thissen, D., Chen, W., & Bock, D. (2003). *MULTIOLOG*: multiple category item analysis and test scoring using item response theory (Version 7.03). Lincolnwood, IL: Scientific Software International.
- Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive

- rate in multiple comparisons. *Journal of Educational and Behavioral Statistics*, 24, 77-83.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the Measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4-70.
- Wu, A. D., Li, Z. & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical assessment Research & Evaluation*, 12, 1-26. Available Online: <http://pareonline.net/pdf/v12n2.pdf>

Archive of SID