Educational Measurement and Evaluation Studies

National Organization for Educational Testing

Print ISSN: 2476-2865 Online ISSN: 2783-0942

Equating and Linking Scores in National Exams

Soleyman Zolfagharnaasab[1] , Ali MoghadamZade[2], Negar SharifiYegane [3]

1. Research Expert of National Organization of Educational Testing (NOET), Iran. (Corresponding Author), Email: salarnik2001@yahoo.com

2.Associate Professor, Faculty of Psychology and Education, University of Tehran, Tehran, Iran. Email: amoghadamzadeh@ut.ac.ir

3.Research Expert of National Organization of Educational Testing (NOET), Iran. Email: nsh_yeganeh@yahoo.com

Article Info

Article Type:
Research Article

Received:2023.07.05
Received in revised form:2023.11.11
Accepted: 2023.12.10
Published online: **2023.12.27**

ABSTRACT

Objective: According to the approval of the Supreme Council of Cultural Revolution the National Organization of Educational Testing is supposed to hold a national exam twice a year and the score of each test would have credit for two years. The purpose of this research is to introduce and implement appropriate methods of equating in the framework of the classical test theory in order to compare the test scores of different courses to choose applicants in a fair way.

Methods: Based on common-item nonequivalent groups design three test forms X, Y and Z of differential calculus were experimentally implemented on the group of 600, 1111 and 2200 examinees. The number of test items were 21, 21 and 20 with six common item in total. Data were analyzed in R software with the package equate (Albano & Albano, 2018). The equipercentile function was used to equate the test forms; The pre-smoothing process with linear logarithmic transformations were done so that the parameters were estimated with less standard error and scores with more accuracy.

Results: The average difficulty of these test forms were 9.03, 7.61 and 7.79 respectively. Also, all three forms have different skewness 0.203, 0.518 and 0.392, respectively.

Conclusion: Due to the different difficulty and right skewness of the distribution of the tests scores, it is suggested to equate them with the equipercentile function in a nonequivalent anchor test design. The tests for which common item cannot be developed, it is recommended to link their scores by the same equipercentile function.

Keywords: equipercentile, pre-smoothing, linear equating, anchor items, national exams

Cite this article: Zolfagharnasab, Soleyman; Moghadamzadeh, Ali; Sharifi Yeganeh, Negar (2023). Equating and Linking Scores in National Exams. *Educational Measurement and Evaluation Studies*, 13 (44):111-132 Pages.
DOI:10.22034/emes.2024.561619.2412

© The Author(s).

Publisher: National Organization of Educational Testing (NOET)

مطالعات اندازه‌گیری و ارزشیابی آموزشی
سازمان سنجش آموزش کشور
شاپا چاپی: ۲۸۶۵-۲۴۷۶ شاپا الکترونیکی: ۰۹۴۲-۲۷۸۳

هم‌ترازسازی و پیوند نمرات در آزمون‌های سراسری

سلیمان ذوالفقارنسب[۱]، علی مقدم‌زاده[۲] و نگار شریفی‌یگانه[۳]

۱. کارشناس پژوهشی سازمان سنجش آموزش کشور، تهران، ایران؛ (نویسنده مسئول)، رایانامه: salarnik2001@yahoo.com
۲. دانشیار دانشکده روانشناسی و علوم تربیتی دانشگاه تهران، ایران. رایانامه: amoghadamzadeh@ut.ac.ir
۳. کارشناس پژوهشی سازمان سنجش آموزش کشور، تهران، ایران. رایانامه: nsh_yeganeh@yahoo.com

اطلاعات مقاله

نوع مقاله:
مقاله پژوهشی

دریافت: ۱۴۰۲/۰۴/۱۴
اصلاح: ۱۴۰۲/۰۸/۲۰
پذیرش: ۱۴۰۲/۰۹/۱۹
انتشار : ۱۴۰۲/۱۰/۰۶

چکیده

هدف: بنا به مصوبه شورای عالی انقلاب فرهنگی قرار است سازمان سنجش در سال دو بار آزمون سراسری برگزار کند و نمره هر آزمون تا دو سال اعتبار خواهد داشت. هدف این پژوهش معرفی و اجرای روش‌های مناسب هم‌ترازی در چهارچوب تئوری کلاسیک آزمون سازی است تا بتوان نمرات آزمون‌های دوره‌های مختلف را مقایسه کرد و بر اساس آنها دانشجویان را به شیوه‌ای عادلانه انتخاب کرد.

روش پژوهش: بر اساس طرح گروه‌های نابرابر با سوالات مشترک سه فرم آزمون X ، Y و Z حساب دیفرنسیل روی ۳ گروه ۶۰۰، ۱۱۱۱ و ۲۲۰۰ نفری به طور آزمایشی اجرا شده. تعداد سوال هر فرم در کل ۲۱، ۲۱ و ۲۰ به ترتیب هر کدام با ۶ سوال مشترک بوده است. داده‌ها در نرم افزار R با پکیج equate (آلبانو، ۲۰۱۸) تحلیل شده‌اند. تابع هم‌ترازی هم‌صدک برای هم‌ترازکردن این فرم‌ها به کار رفته؛ فرایند پیش هم‌وارسازی با تبدیلات لگاریتم خطی صورت گرفته تا پارامترها با خطای معیار کمتر و نمرات با دقت بیشتری برآورد شوند.

یافته‌ها: میانگین دشواری این سه فرم آزمون X ، Y و Z به ترتیب، ۹/۰۳، ۷/۶۱ و ۷/۷۹ است. هم‌چنین هر سه فرم دارای چولگی متفاوت به ترتیب ۰/۲۰، ۰/۵۲ و ۰/۳۹ می‌باشد.

نتیجه‌گیری: در طرح گروه‌های نابرابر به دلیل دشواری متفاوت فرم‌های آزمون‌ها نیازمند سوالات مشترک است و چولگی راست توزیع نمرات مستلزم استفاده از تابع هم‌ترازی هم صدک به عنوان مناسب‌ترین تابع برای هم‌ترازکردن نمرات است؛ برای آزمون‌هایی که نمی‌توان سوالات مشترک تهیه کرد، پیشنهاد می‌شود به دلیل چولگی راست توزیع نمرات با همان تابع هم‌ترازی هم‌صدک آنها را پیوند داد.

واژه‌های کلیدی: هم‌ترازسازی هم‌صدک، پیش هم‌وارسازی، هم‌ترازی خطی، آزمون لنگر، آزمون سراسری

استناد: ذوالفقارنسب، سلیمان، مقدم زاده، علی و شریفی یگانه، نگار (۱۴۰۲). هم‌تراز سازی و پیوند نمرات در آزمون‌های سراسری. *مطالعات اندازه‌گیری و ارزشیابی آموزشی*، ۱۳ (شماره۴۴)، ۱۱۱-۱۳۲ صفحه.
DOI:10.22034/emes.2024.561619.2412
ناشر: سازمان سنجش آموزش کشور

حق‌مؤلف © نویسندگان.

مقدمه

با توجه به مصوبه شماره ۳۲۱۷ در تاریخ ۱۴۰۰/۰۴/۱۵ در شورایعالی انقلاب فرهنگی "سیاست ها و ضوابط ساماندهی سنجش و پذیرش متقاضیان ورود به آموزش عالی (پس از دوران متوسطه)" که در نه ماده به تصویب رسید، ماده دوم آن سازمان سنجش را ملزم می‌کند آزمون‌های تخصصی برای هر رشته در سال دو بار اجرا شوند و نتیجه هر آزمون حداکثر برای دو سال برای هر فرد در همان گروه آزمایشی معتبر باشد (شورایعالی انقلاب فرهنگی، ۱۴۰۰). بدین ترتیب هر داوطلب می‌تواند در دو سال متوالی در چند دوره آزمون شرکت کند و اگر در طی دو سال در همه آزمون‌ها شرکت کند حداکثر سه (یا چهار) نمره در این آزمون‌های اختصاصی خواهد داشت. این کار مستلزم ساخت فرم‌های چندگانه‌ای از یک آزمون برای هر تک درس اختصاصی (مثلا زیست، ریاضی، تاریخ و نظایر آن) است که یک توانایی معین را اندازه بگیرند و هم‌زمان سوالات این فرم‌های آزمون در ضرایب پایایی، ضریب دشواری و ضریب تمیز موازی و یا معادل یکدیگر باشند. به گونه‌ای که نمرات آنها قابل مقایسه باشد، سوای این‌که فرد در کدام دوره آزمون شرکت کرده و یا کدام فرم آزمون را امتحان داده است. با این حال ساخت فرم‌های موازی و یا معادل از یک آزمون و سپس مقایسه نمرات افراد روی آنها به سادگی امکان‌پذیر نیست. برخی از سوالات یک فرم آزمون ممکن است ساده‌تر و یا دشوارتر از سوالات فرم‌های دیگر باشد. یا گروه آزمون‌شوندگان یک دوره قوی‌تر از دوره‌های دیگر باشند. شیوه‌های هم‌ترازسازی[1] این فرم‌های آزمون برای برطرف‌کردن این مشکلات ایجاد شده‌اند. به بیان ساده هم‌ترازسازی شیوه‌های طراحی آزمون و روش‌های آماری هستند که اجازه می‌دهند نمرات روی یک فرم آزمون را با نمرات روی فرم‌های دیگر همان آزمون مقایسه کرد (شیا و نورسینی[2]، ۱۹۹۵).

مبانی نظری و پیشینه پژوهش

در عمل ساخت آزمون‌های چندگانه‌ای که افراد مختلف هر یک فرم‌های متفاوتی از آنها را امتحان داده‌اند و سپس هم‌ترازکردن نمرات آنها به گونه‌ای که بتوان افراد را عادلانه رتبه‌بندی کرد دشوار است. یک آزمون در نظر بگیرید. نمرات یک گروه معین که تنها از روی یک فرم معین آزمون بدست می‌آید متشکل از چهار نوع واریانس است: یکی مربوط به واریانس تفاوت توانایی آزمون‌شوندگان است (واریانس نمره افرادν_p) و دیگری واریانس مربوط به مجموعه سوالات همان فرم آزمون است (واریانس سوالات آزمونν_i) بعلاوه واریانس خطای تعامل p و i به همراه خطای باقیمانده e که عملیاتی نشده $\nu_{pi,e}$ (برنان، ۲۰۰۱، صص. ۲۶۹-۲۶۸).

فرمول ۱

$$X_{piv} = \mu_v + \nu_p + \nu_i + \nu_{pi,e}$$

واریانس ν_p تفاوت توانایی افراد را ایجاد می‌کند. این همان واریانسی است که موردنظر و مطلوب سازندگان آزمون و دست‌اندرکاران برنامه سنجش و پذیرش است. اما واریانس دوم ν_i یا واریانس مجموعه سوالات برآمده از تغییرات ناشی از سوالات آزمون است و نامطلوب است. به عنوان مثال فرض بر این است اگر یک متر برای اندازه‌گیری طول قد افراد به‌کار می‌بریم هر واحد آن دقیقا ۱۰ میلی متر باشد تا بتوان افراد را بر اساس این اندازه‌های استاندارد با یکدیگر مقایسه کرد. اما اگر هر واحد آن میلی مترهای متفاوتی داشته باشند این تفاوت‌ها در برآورد اندازه قد افراد و مقایسه آنها نیز تاثیرگذار است. حالا اگر دو متر متفاوت X و Y در زمان‌های متفاوت به‌کار گرفته شود چگونه می‌توان اندازه‌های قد این دو گروه از افراد را با یکدیگر مقایسه کرد؟

علاوه بر این‌ها واریانس خطای باقیمانده‌ای $\nu_{pi,e}$ نیز وجود دارد که ناشی از تعامل این دو واریانس ν_p و ν_i به همراه واریانس e عملیاتی نشده است.

حال اگر دو یا سه فرم آزمون وجود داشته باشد که سوالات آنها از لحاظ دشواری، پایایی و ضریب تمیز با یکدیگر تا اندازه‌ای تفاوت دارند و در دوره‌های متوالی اجرا شده‌اند و برای هر فرد در هر دوره نمره متفاوتی بدست آمده، این تفاوت دشواری سوال‌های هر فرم در رتبه‌بندی افراد را چگونه می‌توان توجیه کرد؟

در این پژوهش تنها با توجه به نیاز سازمان سنجش برای هم‌ترازکردن سه یا چهار فرم آزمون سراسری در سال و شرایط خاصی که نمرات هر خرده آزمون دارد به معرفی انواع توابع هم‌ترازی (مثل رتبه‌های صدکی، هم‌ترازی خطی، هم‌ترازی میانگین، هم‌ترازی هم‌صدک[3])، ارتباط تئوریک آنها و روش‌های برآورد پارامترها پرداخته شده است و کدهای اجرایی مورد نیاز هر روش در پژوهشی با همین نام به تفصیل آمده است.

---

[1]. equating

[2]. Shea and Norcini

[3]. باید توجه داشت هم‌ترازی با رتبه های صدکی و هم‌ترازی هم‌صدک دو تابع و دو روش متفاوت هستند اما مفاهیم زیربنایی آنها یکسان است.

اما در این نوشته تنها نتایج هم‌ترازی دو آزمون X و Y با سوالات لنگر آمده است. در پایان نمرات سه آزمون X ، Y و Z هم‌زمان و بدون سوالات لنگر هم‌تراز شده‌اند. به هر حال برای هم‌ترازکردن این فرم‌های آزمون‌ها تابع هم‌ترازی هم‌صدک[1] با هم‌وارسازی به کار رفته است.

مدل هم‌ترازسازی هم‌صدک بر اساس ارتباطی که با مدل‌های خطی دارند (به فرمول های ۱۰ و ۱۱ نگاه کنید)، روش بسیار مناسبی برای آزمون‌های سراسری از هر نوع (بدون سوال لنگر و یا با سوال لنگر) می‌باشند. این مدل ترکیبی قوی از رتبه‌های صدکی[2] (مدل‌های ناپارامتریک) و روش خطی (مدل‌های پارامتریک) به همراه هم‌وارسازی است؛ در این پژوهش روش هم‌ترازسازی هم‌صدکی ارایه شده که بتوان نمرات را نخست بر اساس توزیع فراوانی آنها هم‌وارسازی[3] و سپس نمرات آزمون‌ها را هم‌تراز کرد. این تابع هم‌ترازی هم برای آزمون‌هایی با سوالات لنگر و هم بدون سوالات لنگر قابل کاربرد است. در هم‌ترازسازی هم‌صدک دو نمره یکی روی فرم X و دیگری روی فرم Y می‌توان هم‌تراز در نظر گرفت اگر رتبه‌های صدکی آنها در هر گروه برابر باشد (آنگوف[4]، ۱۹۷۱).

در طرح جمع‌آوری داده‌ها روی گروه‌های نابرابر با سوالات لنگر، نمره فرم‌های آزمون‌ها با کمک سوالات لنگر با یکدیگر هم‌تراز می‌شوند. فرض بر این است که دو گروه احتمالا تنها در توانایی با یکدیگر متفاوت (یا نابرابر) هستند و اگر تفاوتی ناشی از سوالات دو فرم v_i آزمون X یا Y است با کمک سوالات لنگر U تعدیل می‌شود. از طرفی این سوالات لنگر بر روی هر دو گروه اجرا می‌شود تا عمل‌کرد ترکیبی گروه‌ها بر روی دو فرم قابل انتقال و مشخص شود. هدف سوالات لنگر این است که مشخص کنند تا چه اندازه تفاوت نمرات روی دو فرم آزمون ناشی از تفاوت در توانایی واقعی v_p افراد دو گروه است و نه تفاوت سوالات v_i دو فرم آزمون. از آن‌جایی که سوالات لنگر تفاوت‌های گروهی را کنترل می‌کنند حتی اگر گروه‌های شرکت‌کننده در آزمون سراسری دقیقا (از لحاظ توانایی) برابر نباشند با سوالات لنگر این نابرابری توانایی آنها کنترل می‌شود (دورانز، موسس و ایگنور[5]، ۲۰۱۰). آزمون‌شوندگان تنها یک فرم آزمون می‌گیرند که در آن سوالات متفاوتی وجود دارد اما سوالات مشترک با فرم دیگر نیز وجود دارد که این تفاوت‌ها را کنترل می‌کنند. این روش برای برخی از دروس (مثل ریاضی و نظایر آن) در برنامه‌های سنجشی به آسانی امکان‌پذیر است. اما ایرادتی نیز دارد از جمله این‌که سوالات لنگر باید به تعداد کافی در آزمون باشند (مثلا یک سوم) و خوب طراحی شده باشند به گونه‌ای که از لحاظ دشواری یک پیوستار از آسان، متوسط تا دشوار را پوشش دهند. ایراد دیگر آن این است چون پیشتر اجرا شده‌اند ممکن است برای آزمون‌شوندگان آشنا باشند و میزان لو رفتن آنها به نسبت دیگر سوالات بالا است. در برنامه‌های سنجشی که سوالات پس از آزمون در جامعه پخش می‌شوند به دو روش با سوالات لنگر برخورد می‌شود: یکی سوالات لنگری که در نمره کل سهم آنها محاسبه می‌شود به نام سوالات لنگر درونی. دیگری سوالات لنگری که پس از فرایند هم‌ترازی سهم آنها از نمره کل جدا می‌شود و به نام سوالات لنگر بیرونی معروف هستند. چون سازمان سنجش نمی‌تواند سوالات لنگر بیرونی داشته باشد–بنا به دلایل قانونی–به ناچار باید از سوالات لنگر درونی استفاده کند که سهم آنها در نمره کل محاسبه می‌شود.

در ادامه بنا به الزامات سازمان "هم‌زمان" سه دوره آزمون (مثلا چندین فرم یک آزمون X، Y و Z) هم‌تراز خواهد شد بدون این‌که این تکرار فرایند ضربدری باعث از بین رفتن هزینه و زمان (و یا وقت و انرژی) عوامل اجرایی در سازمان و سردرگمی آزمون‌شوندگان شود و این فرایند به نصف کاهش یابد.

به هر حال در ادبیات و پیشینه هم‌ترازسازی "چهار طرح جمع‌آوری داده‌ها" وجود دارد:

۱) طرح تک گروهی که در آن یک گروه آزمون‌شونده دو (یا چند) فرم آزمون می‌گیرند.

۲) طرح گروه‌های مستقل که در آن گروه‌های متفاوتی از آزمون‌شوندگان فرم‌های متفاوتی از یک آزمون را می‌گیرند.

۳) طرح گروه‌های نابرابر که در آن هر گروه از آزمون‌شوندگان هر دو (یا همه) فرم‌های آزمون را می‌گیرند و

[1]. equipercentile method
[2]. Percentile Ranks
[3]. Smoothing method
[4]. Angoff
[5]. Dorans, Moses & Eignor

۴) طرح گروه‌های نابرابر با سوالات مشترک[1] (NEAT) در هر آزمون. در این طرح دو (یا چند) گروه از آزمون‌شوندگان فرم‌های متفاوتی از یک آزمون می‌گیرند که در هر فرم یک خرده آزمون لنگر[2] وجود دارد که سوالات آن خرده آزمون برای همه گروه‌ها مشترک است (رایان، بروکمان[3]، ۲۰۰۹؛ شیا و نورسینی، ۱۹۹۵).

سه طرح نخست جمع‌آوری داده ارتباطی با الزامات سازمان سنجش ندارد. تنها در این پژوهش به "طرح گروه‌های نابرابر" پرداخته شده که از عمومیت و کلیت بیشتری برخوردار است. سپس در درون این طرح به جزئیاتی پرداخته شده که مربوط به نوع[4] تابع هم‌ترازسازی و سپس شیوه‌های[5] برآورد پارامترهای هم‌ترازی بر اساس جدول ۱ است.

برای هر یک از این طرح‌های جمع‌آوری داده، گونه‌های مختلفی از توابع هم‌ترازسازی وجود دارد که برای انتقال مقیاس نمرات از روش‌های مختلف آماری استفاده می‌کنند تا نمره مشاهده شده فرم‌های مختلف آزمون را با یکدیگر هم‌تراز کنند و یا پیوند[6] دهند. از جمله مهم‌ترین این توابع می‌توان به تابع خطی، تابع خطی تعمیم یافته، تابع هم‌صدک و نظایر آنها نام برد (وون داویر و وون داویر[7]، ۲۰۰۴).

جدول ۱انواع توابع هم‌ترازسازی و شیوه‌های برآورد پارامترها

(برگرفته از آلبانو، ۲۰۱۶، ص.۹)

نوع تابع هم‌ترازی	روش					
	اسمی	توکر	لوین	براون	فراوانی	زنجیره ای
میانگین	*	*	*	*		❋
خطی		❋	*	*		*
خطی تعیمیم یافته	❋	❋	❋	❋		
هم‌ترازی هم‌صدک					❋	❋
Circle-arc	❋	❋	❋	❋		❋
چندلنگری[8]	❋	❋		❋	❋	

به طور کل داشتن مفروضه‌های متفاوت در رابطه با توزیع نمرات $P(x)$ و $G(y)$ در فرم‌های مختلف آزمون X و Y منجر به تصمیم‌گیری به انتخاب انواع متفاوت توابع هم‌ترازی و روش‌های برآورد پارامترها و آماره‌های نمرات می‌شود (وون داویر و چن، ۲۰۱۳). در هر برنامه هم‌ترازسازی سه گام برداشته می‌شود: نخست انتخاب طرح هم‌ترازسازی، سپس انتخاب نوع تابع هم‌ترازی و در نهایت شیوه برآورد پارامترها. هر سه این گام ها می بایست به تناسب نیاز و الزامات یکدیگر باشند. به عبارتی هنگامی که آزمون‌های اجرا شده مبتنی بر طرح گروه‌های نابرابر است (مثل تکرار دوبار در سال آزمون سراسری)، باید نوع تابع هم‌ترازی متناسب با این طرح انتخاب شود و سپس شیوه برآورد پارامترها متناسب با نوع تابع هم‌ترازی باشد.

پیشینه تجربی و نظری موضوع

در همین زمینه می‌توان به پژوهش‌هایی اشاره کرد که در ایران انجام شده. به عنوان مثال مقدم‌زاده (۱۳۹۴) در مطالعه‌ای روش بهینه هم‌وارسازی داده‌ها را در فرایند هم‌ترازسازی بررسی کرد. نتایج او نشان داد که برای داده‌های آزمون تولیمو فرم X در سه حجم ۲۰۰، ۵۰۰ و ۱۰۰۰ نفر مدل لگاریتم خطی بهترین روش هم‌وارسازی انتخابی است. هم‌چنین به موازات افزایش حجم نمونه و طول آزمون، برازش مربوط به هم‌وارسازی کرنل نیز بهبود می‌یابد. رضوانی‌فر، فلسفی‌نژاد و دلاور (۱۳۹۵) در پژوهشی هم‌ترازسازی نمرات دروس ریاضی و فیزیک رشتهٔ علوم

1. Non-Equivalent Groups with Anchor Test (NEAT) design
2. anchor test
3. Ryan & Brockmann
4. Function Type
5. Method
6. link
7. von Davier
8. Multiple anchors

تجربی آزمون کنکور سراسری سال‌های ۱۳۸۸ و ۱۳۸۹ را براساس نظریه کلاسیک و نظریه جدید اندازه‌گیری مورد بررسی قرار دادند. نتایج بیانگر آن بود که در طرح گروه‌های همسان و هم در طرح گروه‌های ناهمسان، روش هم‌صدک بدون هم‌وارسازی کمترین خطای استاندارد را داشته است و طرح گروه‌های ناهمسان با روش هم‌صدک بدون هم‌وارسازی به نسبت طرح گروه‌های همسان خطای استاندارد کمتری داشت.

بهمن آبادی، فلسفی نژاد، دلاور، فرخی و مینایی (۱۳۹۹) با استفاده از نمونه ۵۰۰۰ نفری از میان داوطلبان گروه علوم ریاضی و فنی کنکور سراسری سال ۱۳۹۶ و ۱۳۹۷ شش روش هم‌ترازسازی تک‌بعدی و چندبعدی را براساس طرح گروه‌های تصادفی به منظور شناسایی رویکرد بهینه هم‌ترازسازی آزمون‌های چندبعدی مورد مقایسه قرار دادند. در این مطالعه روش هم‌صدک به لحاظ مقاوم بودن در برابر مفروضه تک‌بعدی بودن به عنوان ملاکی برای مقایسه روش‌ها درنظر گرفته شد. نتایج آنها نشان داد که در هم‌ترازسازی آزمون‌های چندبعدی بهینه‌ترین روش، هم‌ترازسازی چندبعدی کامل است و پس از آن روش نمره مشاهده شده چندبعدیِ تک‌بعدی شده روش مناسبی است، اما استفاده از روش‌های تک‌بعدی نمره مشاهده شده و نمره واقعی در چنین شرایطی، کارایی مناسبی ندارد.

به هر حال برای این‌که نمرات این آزمون‌ها را بتوان "به شیوه‌ای عادلانه" تفسیر کرد و بر اساس آنها افراد را انتخاب کرد نه تنها آزمون‌ها باید هم‌تراز شوند بلکه باید نمرات آنها روی یک مقیاس مشترک قرار بگیرند تا تفسیرپذیر شوند (تیان[1]، ۲۰۱۱، ص.۲). به بیان ساده هم‌ترازکردن شیوه‌های طراحی آزمون و روش‌های آماری هستند که اجازه می‌دهند نمرات روی یک فرم آزمون با نمرات روی فرم‌های دیگر همان آزمون قابل مقایسه شوند. سوال اساسی در این پژوهش با توجه به آزمون‌های سراسری این است که چگونه می‌توان نمرات آزمون‌ها را –با درصدی از خطای قابل قبول– هم‌تراز کرد به گونه‌ای که نمره در یک فرم آزمون برابر با نمره در فرم دیگر همان آزمون در زمانی دیگر باشد؟ در ادامه تنها با توجه به شرایط سازمان سنجش و آزمون‌هایی که تهیه می‌کند این شیوه‌ها بررسی شده‌اند.

انواع توابع هم‌ترازسازی و شیوه‌های برآورد پارامترهای آن بر اساس تئوری کلاسیک آزمون‌سازی

با توجه به گروه‌های آزمون‌شوندگان طرح گروه‌های نابرابر با سوالات لنگر یکی از مناسب‌ترین طرح‌های هم‌ترازی برای جمع‌آوری داده‌های آزمون‌ها می‌باشد. سپس نوبت آن است که یک نوع تابع هم‌ترازی برای انتقال نمرات فرم‌های چندگانه این آزمون‌ها بر روی یک مقیاس مشترک انتخاب شود. انواع توابع هم‌ترازسازی در چهارچوب دو تئوری روانسنجی ایجاد شده‌اند: تئوری مرسوم (کلاسیک) آزمون‌سازی و تئوری سوال پاسخ[2]. در چهارچوب تئوری کلاسیک چندین نوع هم‌ترازسازی وجود دارد که رایج‌ترین و پرکاربردترین آنها نوع هم‌ترازی خطی و هم‌ترازی هم‌صدک[3] است. این توابع هم‌ترازی را می‌توان هم برای آزمون‌هایی با سوالات لنگر و هم بدون سوالات لنگر به‌کاربرد. اساسا به‌کاربردن سوالات لنگر هنگامی الزام‌آور می‌شود که فرض تفاوت توانایی دو گروه آزمون‌شونده و همچنین فرض دشواری متفاوت فرم‌های آزمون پذیرفته باشیم. این روش آزمون‌سازی در چهارچوب دو تئوری کلاسیک و سوال پاسخ عمومیت دارد. با این تفاوت که در تئوری سوال پاسخ به جای سرجمع نمره سوالات، این پارامترهای تک تک سوالات هستند که با یکدیگر هم‌تراز می‌شوند و طبیعتا روش نیرومندتر و دقیقتری است برای انتقال و هم‌ترازکردن نمرات.

معمولا برای همه روش‌های کلاسیک (به استثنای توابع هم‌ترازی هم‌صدک و توابع آرک سینوس که جزء روش‌های خطی نیستند) فرمول عمومی انتقال خطی به کار می‌رود. اما، توابع هم‌ترازسازی خطی در راه و روشی که برای محاسبه ضرایب ثابت انتقال a و b به کار می برند متفاوت هستند. هدف هم‌ترازسازی در این تبدیل خطی، محاسبه ضرایب a و b است به گونه‌ای که فرمول ۲ را برآورده سازد.

فرمول ۲

$$Y = aX + b$$

باید خاطرنشان شود که هم‌ترازسازی فراسوی معادله رگرسیون Y از روی X است؛ *تفاوت هم‌ترازی با پیش‌بینی (رگرسیون) این است که هم‌ترازی دو نمره به صورت متقارن است اما پیش‌بینی (در رگرسیون) متقارن نیست. به عبارتی در رگرسیون پیش‌بینی y از روی x با پیش‌بینی x از روی y متفاوت است. مقادیری که به عنوان y پیش‌بینی می‌شود دقیقا برابر با نمره x نیست چه بسا اگر معادله رگرسیون را برعکس کنیم و با y مقادیر x را پیش‌بینی کنیم عددی متفاوت مثل z بدست آید. بنابراین هنگامی که یک نمره در فرم آزمون جدیدی را برای هم‌ترازکردن با فرم*

[1] Tian
[2] item response theory
[3] equipercentile method

قدیم به‌کار می‌بریم تلاش نمی‌کنیم که این نمره را برای پیش‌بینی نمره در فرم قدیم به کار بریم. فرایند متفاوتی به‌کار می‌بریم. بنابراین هم‌ترازسازی پیش‌بینی نیست (لی وینگ استون، ۲۰۱۴، ص. ۱۱).

تابع هم‌ترازی خطی

در هم‌ترازسازی خطی فرض بر این است که دو آزمون در میانگین [دشواری] و انحراف استاندارد تفاوت دارند، اما شکل توزیع دو فرم آزمون یکسان و مشابه است. با توجه به این مفروضه اگر تفاوت در دشواری دو آزمون هست نمرات روی دو فرم آزمون را می‌توان با z استاندارد با یکدیگر تطبیق داد (موراکی، هومبو و لی[1]، ۲۰۰۰).

فرمول ۳

$$z = \frac{X-\mu_X}{\sigma_X} = \frac{Y-\mu_Y}{\sigma_Y}$$

X نمره در آزمون X و μ_X و σ_X به ترتیب میانگین و انحراف معیار آن

Y نمره در آزمون Y و μ_Y و σ_Y به ترتیب میانگین و انحراف معیار آن

با این تبدیل یک نمره در فرم Y برابر با نمره در فرم X خواهد بود اگر نمره z استاندارد آنها مشابه یکدیگر باشد (آلبانو[2]، ۲۰۱۶). هنگامی که شرایط این تبدیل مناسب باشد فرمول بالا را می‌توان برای هم‌ترازی و تبدیل مقیاس به صورت تابع عمومی $Y = aX + b$ تبدیل کرد (آنگوف، ۱۹۷۱، ص.۸۷). این تابع فرض می‌گیرند که تفاوت دشواری بین دو فرم X و Y بر اساس یک مقدار ثابت a تغییر می‌کند. به هر حال در هم‌ترازسازی خطی، شیب خط را بر اساس انحراف استاندارد دو فرم X و Y برآورد می‌کنند:

فرمول ۴

$$a = \frac{\sigma_Y}{\sigma_X} \quad و \quad b = \mu_Y - \frac{\sigma_Y}{\sigma_X}\mu_X$$

بنابراین شکل دیگر تابع عمومی خطی برای هم‌ترازکردن و یا پیوند نمرات بر اساس فرمول زیر است:

فرمول ۵

$$lin_Y(x) = \frac{\sigma_Y}{\sigma_X}x + \mu_Y - \frac{\sigma_Y}{\sigma_X}\mu_X$$

با این حال هم‌ترازسازی خطی هنگامی مناسب است که توزیع دو نمره تنها در میانگین و انحراف استاندارد با یکدیگر تفاوت داشته باشند. اگر تفاوت‌ها فراتر از این دو گشتاور باشند، هم‌ترازسازی خطی مناسب نیست (موراکی، هومبو و لی[3]، ۲۰۰۰).

لازم به تاکید است هنگامی که شکل توزیع نمرات خام دو فرم X و Y شبیه به هم باشد هم‌ترازی خطی تقریب بسیار نزدیکی (یا مشابه) است با هم‌ترازی هم‌صدک که در ادامه آمده. اگر بتوانیم فرض بگیریم تفاوت در شکل توزیع نمرات دو فرم به اندازه کافی ناچیز هستند که بتوان آنها را نادیده گرفت هم‌ترازی خطی بسیار مناسب است. برخلاف هم‌ترازی هم‌صدک، این هم‌ترازی تایید شده است و رها از هر نوع خطای هم‌وارسازی است، که می‌تواند خطاهای جدی ایجاد کند در دامنه نمراتی که داده‌ها در آن نقاط اندک و یا نامنظم هستند (آنگوف، ۱۹۷۱، صص. ۸۷-۸۸).

تابع هم‌ترازی میانگین

هنگامی که انحراف استاندارد دو توزیع با یکدیگر برابر باشد، $\sigma_X = \sigma_Y$، هم‌ترازسازی خطی (در فرمول ۵) برابر می‌شود با هم‌ترازی بر اساس میانگین[4]:

فرمول ۶

$$x - \mu_X = y - \mu_y$$

در هم‌ترازسازی میانگین فرض می گیرد که تفاوت تنها بین میانگین دو فرم آزمون هست و این تفاوت در امتداد دامنه مقیاس نمرات یکسان است. این مفروضه در واقعیت کمتر پیش می‌آید و در عمل از هم‌ترازی با میانگین کمتر استفاده می‌شود (تیان[5]، ۲۰۱۱).

[1] Muraki, Hombo & Lee
[2] Albano
[3] Muraki, Hombo & Lee
[4] Mean equating
[5] Tian

ویژگی‌های هم‌ترازی خطی

ویژگی‌های هم‌ترازسازی خطی این است که در این تبدیل میانگین و انحراف استاندارد متفاوت هر دو توزیع X و Y با یکدیگر برابر می‌شوند اما شکل توزیع آنها تغییر پیدا نخواهد کرد (چن، هوانگ و مک گرگور[1]، ۲۰۰۹). چون فرض اولیه بر آن است که شکل توزیع هر دو یکسان است:

فرمول ۷

$$\sigma[l_Y(X)] = \sigma(Y) \quad \text{و} \quad E[l_Y(X)] = \mu(Y)$$

اما، مشکل هم‌ترازسازی خطی این است که اگر توزیع اولیه نمرات چوله باشند حتی پس از تبدیل خطی، توزیع آنها چوله می ماند.

از دیگر مشکلات هم‌ترازسازی خطی این است که نمرات تبدیل‌شده ممکن است خارج از دامنه نمرات مشاهده شده قرار بگیرند. البته می‌توان هنگامی که نمرات تبدیل‌شده فراتر از دامنه نمرات ممکن می روند برای آن کمینه و بیشینه نمرات ممکن را تعیین کرد که به آن کوتاه‌سازی[2] (یا برش‌دادن) می‌گویند. به عنوان مثال اگر آزمونی دامنه نمراتی برابر با ۱۰۰ تا ۶۰۰ داشته باشد اگر پس از معادل‌سازی به نمراتی فراتر از نمرات ممکن ۶۰۰ رسید می‌توان دامنه نمرات را با کوتاه‌سازی به مقیاس نمرات کمینه و بیشینه ممکن برگرداند (چن، هوانگ و مک گرگور[3]، ۲۰۰۹). مشکل سوم در هم‌ترازی خطی ممکن است دو نمره در فرم‌های X و Y با z یکسان رتبه‌های صدکی متفاوتی داشته باشند.

به هر حال هنگامی که در آزمون‌گیری‌های چندگانه روی جمعیت فرض شده توزیع نمرات گروه‌های متفاوت داوطلبان در حد قابل قبولی نسبتا پایدار و یکسان است، اگر در میانگین و واریانس توزیع نمرات آزمون‌ها تفاوت‌های زیادی دیده شود هم‌ترازسازی خطی بی‌اعتبار می‌شود. معمولا در این موارد طرح گروه‌های نابرابر با سوالات لنگر مناسب‌تر است (موراکی، هومبو و لی[4]، ۲۰۰۰).

تابع هم‌ترازی هم‌صدک

هم‌ترازی هم‌صدک هنگامی مناسب است که توزیع نمرات چولگی داشته باشد و هم‌زمان رابطه خطی بین نمرات دو فرم را نتوان توجیه کرد. هم‌ترازی هم‌صدک $Equip_{XYT}(x)$ دربرگیرنده تعیین یک نمره در هر توزیع است که رتبه صدکی مشابه‌ای داشته باشد. سپس آن نمرات با رتبه صدکی یکسان را هم‌تراز می نامند (آنگوف، ۱۹۷۱). هنگامی که رابطه بین توزیع نمرات غیرخطی است، برای محاسبه نمرات X و Y تابع توزیع تجمعی آنها را برابر قرار می‌دهند $Y: P(x) = G(y)$. با حل این معادله برای y، تابع پیوند هم‌ترازی هم‌صدک $Equip_{XYT}(x) = e_Y(x)$ به صورت زیر ایجاد می‌شود:

فرمول ۸

$$\begin{aligned} e_Y(x) = y = Q^{-1}[P(x)] \quad &, \quad -0.5 \le x \le Kx + 0.5 \\ = \frac{\frac{P(x)}{100} - G(y^*_{Up} - 1)}{G(y^*_{Up}) - G(y^*_{Up} - 1)} + (y^*_U - 0.5) \quad &, \quad 0 \le P(x) \le 100 \\ = Kx + 0.5 \quad &, \quad P(x) \le 100 \end{aligned}$$

$P(x)$ توزیع تجمعی گسسته x

$G(y)$ توزیع تجمعی گسسته y

$Q(y)$ رتبه صدکی Y

Q^{-1} وارون تابع رتبه صدکی برای فرم Y

K تعداد سوالات در فرم Y

y^*_{Up} برابر است با کوچکترین نمره صحیح با درصد تجمعی برابر با $100[G(y)]$ که بالاتر از $P(x)$ است[5].

از تابع معکوس Q^{-1} بالا برای پیدا کردن رتبه صدکی یک نمره در فرم X با نمره‌ای برابر با رتبه صدکی آن روی فرم Y استفاده می‌شود.

[1] Chen, Huang & MacGregor
[2] Truncation
[3] Chen, Huang & MacGregor
[4] Muraki, Hombo & Lee

[5] در اینجا نماد y^*_U و y^*_L به ترتیب نشانگر بالاتر Upper و پایین تر Lower است

یک شکل جایگزین[1] دیگر برای تابع هم‌ترازسازی هم‌صدک وجود دارد که نتایج یکسانی به بار می‌آورد مگر این‌که برخی از احتمالات برابر با صفر باشد. فرم جایگزین به این صورت است:

فرمول ۹

$$Equip_{XYT}(x) = \frac{\frac{P(x)}{100} - G(y^*_{Lo})}{G(y^*_{Lo}+1) - G(Y^*_{Lo})} + (y^*_{Lo} + 0.5) \quad , \quad 0 \le P(x) \le 100$$
$$= -.5 \quad , \quad P(x) = 0$$

برای $0 \le P(x) \le 100$ ، مقدار y^*_L برابر است با بزرگترین نمره صحیح با درصد تجمعی $100[G(y)]$ که پایین‌تر از $P(x)$ باشد. در همه نقاط نمره $0,1, \dots, K_Y$ اگر $g(y)$ غیر صفر باشد، آنگاه $e_Y(x) = e_{YLo}(x) = e_{YUp}(x)$ و بدین ترتیب می‌توان از هر یک از عبارت‌ها استفاده کرد (هه[2]، ۲۰۰۷، صص.۶۱-۶۲).

ارتباط بین توابع هم‌ترازی خطی و هم‌صدک

برای هر جامعه‌ی هدف فرضی T، اگر دو گروه (یا دو نمونه افراد از آن جمعیت) به مجموعه سوالات دو آزمون X و Y پاسخ دهند و اگر به ترتیب تابع توزیع تجمعی $P_T(x)$ و $G_T(y)$ آنها پیوسته باشد و تابع توزیع خطی P_0 و G_0 دارای توزیعی با میانگین صفر $\mu_{XT} = 0$ و واریانس یک ${\sigma_{XT}}^2 = 1$ باشد به این ترتیب داریم:

فرمول ۱۰

$$P_T(x) = P_0\left(\frac{x - \mu_{XT}}{\sigma_{XT}}\right)$$
$$G_T(y) = G_0\left(\frac{y - \mu_{YT}}{\sigma_{YT}}\right)$$

به عبارتی مقدار پرانتز به ارزش $P_0(x)$ می گراید و دو طرف معادله برابر می‌شود. بنابراین هم‌ترازسازی هم‌صدک $Equip_{XYT}(x)$ برابر با هم‌ترازی خطی $Lin_{XYT}(x)$ می‌شود بعلاوه یک مقدار $R(x)$. به این ترتیب داریم:

فرمول ۱۱

$$Q^{-1}[P(x)] = Equip_{XYT}(x) = Lin_{XYT}(x) + R(x)$$

جایی که $R(x)$ برابر است با :

$$\sigma_{YT} r\left(\frac{x - \mu_{XT}}{\sigma_{XT}}\right)$$

و مقدار پرانتز $r(z)$ تابعی است از

$$r(z) = G_0^{-1}\big(P_0(z)\big) - z$$

(به نقل از وون داویر و چن، ۲۰۱۳، ص.۹-۴).

هم‌ترازی هم‌صدک هنگامی مناسب است که X و Y به طور غیرخطی در دشواری متفاوت باشند. به این معنی که تفاوت‌های دشواری دو فرم X و Y در امتداد مقیاس نمرات آنها به طور بلقوه در هر یک از نقاط نمره نوسان می‌کند. بنابراین در مقایسه با روش هم‌ترازی خطی، روش هم‌ترازی هم‌صدک نسبت به خطای نمونه‌گیری[3] بسیار حساس‌تر است چون دربرگیرنده برآورد پارامترهای بسیار بیشتری است تا تنها یک نقطه نمره منحصر به فرد روی فرم X (آلبانو، ۲۰۱۶، ص.۵).

در هم‌ترازسازی هم‌صدک $e_Y(x)$ نمرات هم‌تراز شده همیشه در دامنه نمرات ممکن قرار می‌گیرند. این یکی از معایب هم‌ترازی هم‌صدک است که نمی‌توان روابط هم‌ترازی را فراتر از بزرگترین نمرات مشاهده شده تعیین کرد. به عبارتی هم‌ترازی تنها می‌تواند در دامنه نمرات مشاهده

1 An alternate form
2 Heh
3 Sampling error

شده برای نمونه‌های به کار رفته در فرایند معادل‌سازی ایجاد شود. برای نمرات خارج از دامنه نمرات مشاهده شده برونیابی[1] لازم است (سوامی ناتان، ----)[2].

در مواردی که بر اثر خطاهای تصادفی در برآورد هم‌ترازی توزیع نمرات بی‌نظم[3] است، برای کاهش خطاهای تصادفی و ایجاد نمودار هموارتر می‌توان روش هم‌وارسازی[4] را در روش هم‌ترازسازی هم‌صدک به‌کاربرد. خطای استاندارد هم‌ترازی هم‌صدک بزرگتر از هم‌ترازی خطی است. از مشکلات دیگر هم‌ترازی هم‌صدک این است که روش‌های مختلف هم‌وارسازی ممکن است نتایج هم‌ترازی متفاوتی ایجاد کنند.

به هرحال در هم‌ترازسازی هم‌صدک، اگر نمرات در دو فرم در جمعیت هدف[5] رتبه‌های صدکی یکسانی داشته باشند معادل در نظر گرفته می‌شوند. پس از هم‌ترازسازی، نمرات معادل شده گروه هدف در فرم جدید X با نمرات گروه مرجع در فرم قدیم Y توزیع یکسانی خواهد داشت (آنگوف، ۱۹۷۱؛ چن، هوانگ و مک گرگور[6]، ۲۰۰۹).

هم‌ترازی با آزمون لنگر[7](سوالات مشترک)

هم‌ترازی با سوالات لنگر پیچیده‌تر از گونه‌های دیگر هم‌ترازی است. موضوع به این سادگی نیست که تنها دو توزیع نمره در گروه‌های آزمون‌شونده همسان را هم‌تراز کنیم. ما به سوالات لنگر نیاز داریم چون نمی‌توانیم فرض کنیم دو گروهی که دو آزمون متفاوت گرفته‌اند در توانایی و مهارت با هم برابرند. تا اندازه‌ای باید از اطلاعات بدست آمده از نمرات سوالات لنگر استفاده کنیم تا بتوانیم تفاوت‌های بین گروه‌هایی که فرم جدید X و فرم مرجع Y را گرفته‌اند تعدیل کنیم. باید فرض بگیریم که در گروه‌های نمونه‌ای که هم‌تراز می‌کنیم نوعی از اطلاعات می‌توانیم محاسبه کنیم که قابل تعمیم است به جمعیت هدف فرضی T (لی وینگ استون، ۲۰۱۴، ص. ۵۰).

در طرح گروه‌های نابرابر، پارامترهای دو نمونه A و B از آزمون‌شوندگان روی دو فرم آزمون X و Y به همراه سوالات مشترک U در هر آزمون محاسبه می‌شود. سپس برآوردهای جداگانه‌ای که از سوالات مشترک در دو گروه شده پایه و مبنایی می‌شود برای انتقال مقیاس[8]. در این گونه هم‌ترازی دو گروه متفاوت از آزمون‌شوندگان در دو آزمون متفاوت شرکت می‌کنند اما در هر دو آزمون یک مجموعه سوال لنگر (آزمون لنگر یا سوالات مشترک) وجود دارد. این خرده آزمون لنگر تفاوت‌های گروهی را تحت کنترل در می‌آورد.

جدول ۰ طرح گروه‌های نابرابر با سوالات لنگر

جمعیت	نمونه	فرم X	لنگر U	فرم Y
P	A	*	*	
Q	B		*	*

از آنجا که برای هم‌ترازسازی خطی از معادله‌های زیر استفاده می‌شود:

فرمول ۱۲

$$Y = aX + b$$
$$b = \mu_Y - \frac{\sigma_Y}{\sigma_X}\mu_X = M_{yt} - aM_{xt} \quad \text{و} \quad a = \frac{\sigma_Y}{\sigma_X} = \frac{s_{yt}}{s_{xt}}$$

چهار معادله باید حل شود تا ضرایب بدست آیند (شیا و نورسینی[9]، ۱۹۹۵، ص. ۲۶۴؛ آنگوف، ۱۹۸۴، ص.۱۰۵):

فرمول ۱۳

$$M_{xt} = M_{xa} + b_{xua}(M_{ut} - M_{ua})$$
$$M_{yt} = M_{yb} + b_{yub}(M_{ut} - M_{ub})$$
$$s_{xt}^2 = s_{xa}^2 + b_{xua}^2(s_{ut}^2 - s_{ua}^2)$$

[1] extrapolation

[2] لازم به توضیح است که پکیج equate که برای همترازکردن نمرات این پژوهش به‌کار رفته این محدودیت‌ها را ندارد.

[3] irregular

[4] Smoothing method

[5] Target population

[6] Chen, Huang & MacGregor

[7] Non-equivalent Anchor Test (NEAT) Design

[8] Scale transformation

[9] Shea and Norcini

$$s_{yt}^2 = s_{yb}^2 + b_{yub}^2(s_{ut}^2 - s_{ub}^2)$$

فهرست همه مولفه های این معادله $Y = aX + b$ طولانی است. اما محاسبه عبارات و در نهایت انتقال نمرات را می‌توان با پکیج equate (آلبانو، ۲۰۱۸) در نرم افزار R انجام دارد.

در شیوه‌ای که توکر[1] برای برآورد پارامترهای تابع هم‌ترازسازی خطی با سوالات لنگر ارایه داد (به نقل از آلبانو، ۲۰۱۶، ص. ۹) ارتباط بین نمره کل و نمره سوالات لنگر به صورت شیب رگرسیون تعریف می‌شود، جایی که γP برابر است با شیب پیش‌بینی X از روی سوالات مشترک U برای جمعیت P و γQ برابر است با شیب حاصل از پیش‌بینی Y از روی سوالات مشترک U برای جمعیت Q:

فرمول ۱۴

$$\gamma Q = \frac{\sigma_{X_Q,U_Q}}{\sigma_{U_Q}^2} \qquad \text{و} \qquad \gamma \mathrm{P} = \frac{\sigma_{X_p,U_p}}{\sigma_{U_p}^2}$$

روش توکر که اصول آن مبتنی بر رگرسیون است فرض می گیرد که در روی هر دو جمعیت P و Q دو اصل وجود دارد:

۱) ضرایب بدست آمده از پیش‌بینی X از روی U برای هر مقدار U در امتداد خط رگرسیون با یکدیگر یکسان هستند. ۲) واریانس X به شرط U نیز برای تمام مقادیر U یکسان است. این مفروضه برای پیش‌بینی Y از روی U و هم‌چنین واریانس Y به شرط U نیز برای تمام مقادیر U یکسان است.

لوین (۱۹۵۵) روشی بر اساس نمره مشاهده شده و نمره واقعی در چهارچوب تئوری کلاسیک برای برآورد پارامترهای تابع هم‌ترازی خطی پیشنهاد داد که بعدها بوسیله کولن و برنان (۲۰۰۴) توسعه داده‌شد.

فرمول ۱۵

$$x = \tau_x + \varepsilon_x$$
$$y = \tau_y + \varepsilon_y$$
$$u = \tau_u + \varepsilon_u$$

بر اساس تئوری کلاسیک نمره مشاهده شده (مثلا x) به صورت نمره واقعی τ بعلاوه مقداری خطای تصادفی ε بیان می‌شود. برای هم‌ترازسازی فرض بر این است که از دو جمعیت P و Q نمونه‌های آزمون‌شوندگان A و B انتخاب شده‌اند. هم‌چنین در روش لوین فرض می گیرد که در روی هر دو جمعیت های P و Q سه مفروضه وجود دارد:

نخست همبستگی بین نمرات حاصل از سوالات فرم X با سوالات مشترک U برابر با ۱ است، به همین ترتیب همبستگی بین سوالات فرم Y با سوالات مشترک U نیز برابر با ۱ است؛

دوم ضرایب بدست آمده از پیش‌بینی (رگرسیون) نمره‌های واقعی در فرم X از روی سوالات مشترک U برابر هستند. به عبارتی ضرایب شیب $\gamma = a$ برای پیش‌بینی X از روی U برای همه مقادیر U یکسان و مشابه است. به همین ترتیب برای پیش‌بینی Y از روی U همین مفروضه وجود دارد؛

سوم واریانس خطای اندازه‌گیری[2] ε (بر روی جمعیت) برای X و Y و U یکسان است (برای توضیحات بیشتر نگاه کنید به آلبانو، ۲۰۱۶، صص.۹-۱۰). این مفروضه‌ها امکان برآورد γ را فراهم می‌سازند به گونه‌ای که بتوان، در چهارچوب نمره واقعی در تئوری کلاسیک آزمون‌سازی، نمرات فرم‌های آزمون را با سوالات لنگر به صورت زیر عملیاتی کرد:

فرمول ۱۶

$$\gamma \mathrm{Q} = \frac{\sigma_{X_Q}^2}{\sigma_{Y_Q,U_Q}} \quad \text{و} \quad \gamma \mathrm{P} = \frac{\sigma_{X_P}^2}{\sigma_{X_P,U_P}}$$

بنابراین برای تابع هم‌ترازی داریم:

فرمول ۱۷

$$lin_Y(x) = \frac{\gamma \mathrm{Q}}{\gamma \mathrm{P}} X(x - \mu \mathrm{X}_P) + \mu \mathrm{Y}_Q + \gamma \mathrm{Q}\left(\mu \mathrm{U}_P - \mu \mathrm{U}_Q\right)$$

1Tucker
2 measurement error variance

مزایای هم‌ترازسازی با سوالات لنگر آن است که قابلیت اجرای بالایی در برنامه‌های سنجشی دارد. نیازی نیست گروه‌ها با یکدیگر برابر باشند. آزمون‌شوندگان تنها یک فرم آزمون بعلاوه خرده آزمون لنگر U می‌گیرند. بدین ترتیب یک فرم جدید آزمون X را می‌توان به فرم پیشین Y هم‌تراز کرد بدون این‌که نیاز باشد همه سوالات فرم قدیمی Y را دوباره اجرا کنیم.

معایب این روش این است که حتما سوالات لنگر (یا مشترک U) در دو فرم باید به تعداد کافی و از لحاظ کیفی مناسب باشند. سوالات لنگر در دو گروه متفاوت A و B از آزمون‌شوندگان، باید کارکرد یکسانی داشته باشند (به عنوان مثال دشواری برابری در دو گروه داشته باشند). معمولا اگر با آزمون پیشین هم‌ترازسازی صورت بگیرد، سوالات لنگری که قبلا اجرا شده‌اند برای آزمون‌شوندگان بعدی شناخته شده است. اما همان‌طور که پیشتر گفته شد می توان تصمیم گرفت آیا نمره سوالات لنگر در تهیه نمره کل باشند یا نه.

روش پژوهش

طرح جمع‌آوری داده‌های این پژوهش بر اساس طرح گروه‌های نابرابر است که در آن دو یا چند گروه (نابرابر در توانایی) از آزمون‌شوندگان در دوره‌های متفاوت از سال تحصیلی فرم‌های متفاوتی از یک آزمون می‌گیرند. الزام طرح گروه‌های نابرابر داشتن سوالات لنگر[1] است. در این پژوهش علاوه بر هم‌ترازی با سوالات لنگر، شیوه‌های پیوند دادن نمرات بدون سوالات لنگر نیز توضیح داده شده. در طرح گروه‌های نابرابر اگر سوالات لنگر وجود داشته باشند هم‌ترازی کامل است، اما اگر سوالات لنگر نباشند صرفا نمرات فرم‌های آزمون‌ها با یکدیگر "پیوند" داده می‌شوند. برای این کار چه سوالات لنگر باشند چه نباشند گونه‌های مختلفی از توابع هم‌ترازسازی نمرات و انتقال مقیاس آنها وجود دارد (به جدول ۱ نگاه کنید). از جمله توابع هم‌ترازی از این نوع که می توان نام برد عبارتند از: رتبه صدکی، میانگین، خطی و هم‌صدک (وون داویر و وون داویر[2]، ۲۰۰۴) که هر یک مزایا و معایب خاص خود را دارند. این گونه‌های متفاوت هم‌ترازی در برگیرنده گستره‌ای از روش‌های آماری (مثل روش توکر، لوین و نظایر آن) هستند که بسته به شرایط آزمون نمره مشاهده شده فرم‌های مختلف یک آزمون را پارامتریزه و سپس مقیاس آنها را بر روی یکدیگر برمی‌گردانند[3].

الف) جامعه، نمونه و روش نمونه‌گیری و حجم نمونه

جامعه این پژوهش داوطلبان بلقوه در آزمون سراسری هستند که در سال های گذشته دیپلم گرفته‌اند و یا در حال گرفتن آن هستند. نمونه این پژوهش یک گروه آزمایشی ریاضی و فنی است که نمرات آنها روی یک آزمون اختصاصی (حسابان) تحلیل شده. حجم نمونه برابر ۶۰۰ و ۱۱۱۱ و ۲۲۰۰ داوطلب شرکت‌کننده در آزمون X و Y و Z است.

ب) ابزار گردآوری داده و نحوه ارزیابی روایی و اعتبار ابزار

ابزار گردآوری داده‌ها سه فرم از یک آزمون آزمایشی تخصصی گروه‌های ریاضی و فنی (حسابان) بوده است: آزمون X با ۲۱ سوال (۶ سوال لنگر و ۱۵ سوال متفاوت)، آزمون Y با ۲۱ سوال (۶ سوال لنگر و ۱۵ سوال متفاوت) و آزمون Z با ۲۰ سوال (۶ سوال لنگر و ۱۴ سوال متفاوت).

ج) روش‌های آماری تجزیه و تحلیل داده‌ها

چون طرح جمع‌آوری داده‌ها بر اساس گروه‌های نابرابر است الزاما باید از سوالات لنگر استفاده شود. اما انواع توابع هم‌ترازسازی از جمله رتبه‌های صدکی، هم‌ترازسازی خطی، هم‌ترازی هم‌صدک، چه با سوالات لنگر و چه بدون سوالات لنگر روی این داده‌ها اجرا شده و نتایج گزارش شده‌اند. این‌کار با نرم افزار R و به وسیله پکیج equate (آلبانو[4]، ۲۰۱۸) است.

یافته‌ها

در ادامه نتایج توابع هم‌ترازی به کار رفته برای هم‌ترازکردن نمره آزمون‌ها آمده است. آماره ها و ویژگی‌های توزیع نمرات هر فرم در جداول خلاصه شده است. فرض بر این است که فرم آزمون X همین روزها اجرا شده و فرم آزمون Y قدیمی در شش ماه پیش و فرم آزمون Z یک سال و نیم پیش اجرا شده است. در پیشینه پژوهشی آزمون‌های قدیمی که آزمون‌های جدید بر روی آنها برگردانده می‌شود را آزمون مرجع می‌خوانند. به عنوان مثال برای آزمون X ، آزمون مرجع Y می‌باشد. حالا اگر آزمون Z هم باشد، این آزمون برای هر دو آزمون X و Y مرجع می‌باشد.

1 Non-Equivalent Groups with Anchor Test (NEAT) design
2 von Davier
3 Scale transformation
4 Albano

نتایج تابع هم‌ترازی هم‌صدک در طرح گروه‌های نابرابر با سوالات لنگر

همان‌طور که گفته شد، هنگامی‌که فرض نابرابری دو توزیع نمرات X و Y باشد و آزمون‌شوندگان احتمالا از دو جامعه نمونه برداری شده‌اند که نمی‌توان فرض برابری میانگین و انحراف معیار توزیع آنها را پذیرفت (به عنوان مثال نظام قدیم و جدید) چاره‌ای جز ساخت سوالات لنگر نیست. چون هر دو گروه به سوالات لنگر جواب می‌دهند، این سوالات لنگر کمک می‌کنند تا پارامترهای توانایی هر دو گروه بر اساس سوالات لنگر به یک مقدار معین برگشت داده شوند. در جدول زیر آماره‌های ۲۱ سوال و ۶ سوال لنگر آنها به طور جداگانه آمده است.

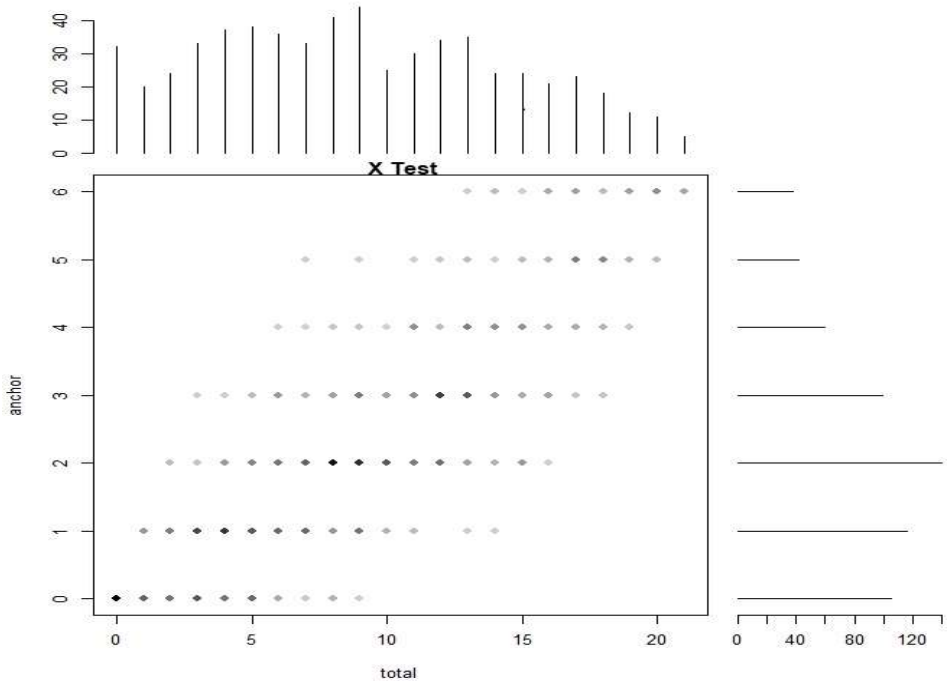
جدول ۳ گشتاورها و آماره‌های توصیفی دو آزمون X و Y با سوالات لنگر و تعداد n نمونه آزمون‌شوندگان

آزمون	میانگین	انحراف معیار	چولگی	کشیدگی	کمینه	بیشینه	nتعداد
x.total	۹/۰۳	۵/٤۸	۰/۲۰	۲/۱	۰	۲۱	٦۰۰
x.anchor	۲/۲۹	۱/۷٤	۰/۵۲	۲/٤۱	۰	٦	٦۰۰
y.total	۷/٦۱	۵/۲٦	۰/۵۲	۲/٤۲	۰	۲۱	۱۱۱۱
y.anchor	۱/۹۲	۱/٦۸	۰/٦٤	۲/۵۲	۰	٦	۱۱۱۱

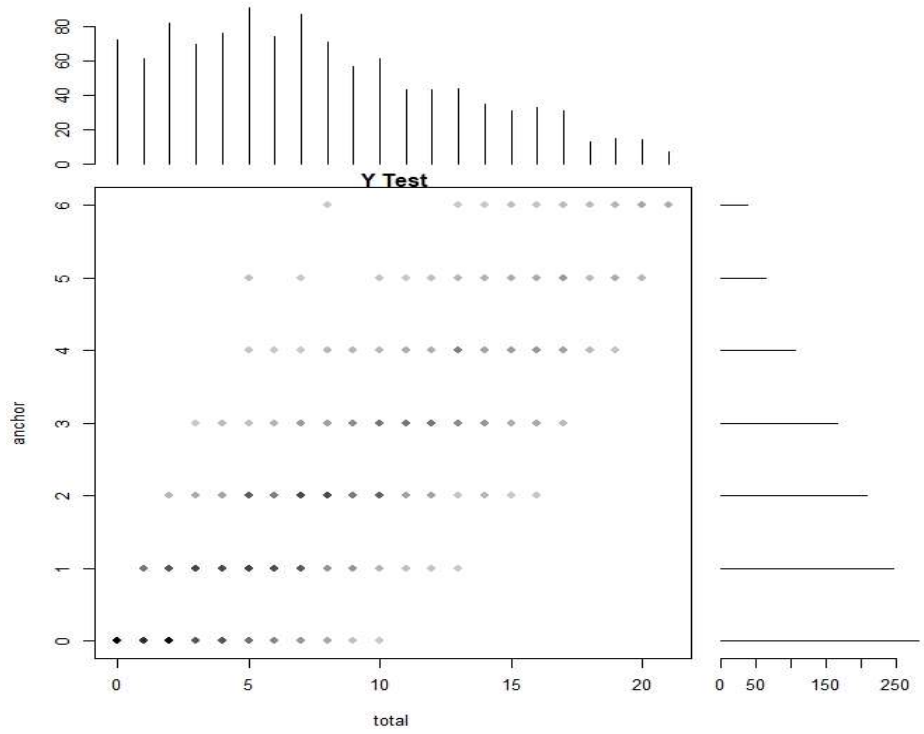
جدول ٤ جزئیات گشتاورها و آماره‌های توصیفی دو آزمون X و Y با سوالات لنگر U و تعداد n نمونه آزمون‌شوندگان با مقادیر هم‌وارسازی شده

برآورد فراوانی در هم‌ترازی هم‌صدک: برگردان neat.xx بر روی neat.yy							
طرح: گروه‌های نابرابر							
روش هم‌وارسازی: پیش هم‌وارسازی با لگاریتم خطی							
وزن ترکیبی برای x: ۰/۳۵							
خلاصه آماری:							
آزمون	میانگین	انحراف معیار	چولگی	کشیدگی	کمینه	بیشینه	nتعداد
x.count	۹/۰۳	۵/٤۷	۰/۲۰	۲/۱	۰	۲۱	٦۰۰
x.smooth	۹/۰۳	۵/٤۷	۰/۲۰	۲/۱	۰	۲۱	٦۰۰
x.synth	۸/۲۵	۵/٤۳	۰/۳۲	۲/۱٤	۰	۲۱	۹۳۱/۸۱
y.count	۷/٦۱	۵/۲٦	۰/۵۲	۲/٤۲	۰	۲۱	۱۱۱۱
y.smooth	۷/٦۱	۵/۲٦	۰/۵۲	۲/٤۲	۰	۲۱	۱۱۱۱
y.synth	۷/۸۳	۵/۳۱	۰/٤۹	۲/۳۸	۰	۲۱	۹۳۱/۸
yx.obs	۸/۵۹	۵/٤۱	۰/۳۷	-۲/۲۷	۰/۰۰۲	۲۱/۱	٦۰۰
xu.count	۲/۲۹	۱/۷٤	۰/۵۲	۲/٤۱	۰	٦	٦۰۰
xu.smooth	۲/۲۹	۱/۷٤	۰/۵۲	۲/٤۱	۰	٦	٦۰۰
xu.synth	۲/۰۰	۱/۷۰	۰/٦۱	۲/٤۹	۰	٦	۹۳۱/۸
yu.count	۱/۹۲	۱/٦۸	۰/٦٤	۲/۵۲	۰	٦	۱۱۱۱
yu.smooth	۱/۹۲	۱/٦۸	۰/٦٤	۲/۵۲	۰	٦	۱۱۱۱
yu.synth	۲/۰۰	۱/۷۰	۰/٦۱	۲/٤۹	۰	٦	۹۳۱/۸

در زیر نمودارهای شرطی توزیع نمرات سوالات هر فرم آزمون X و Y آمده است. نقاط پر رنگ در این نمودارها نشان می‌دهند به شرط این‌که افراد در آزمون لنگر به یک تعداد مشخص سوالات لنگر جواب داده باشند (مثلا نمره ۳ گرفته باشند) توزیع تراکمی نمرات آنها روی سوالات دیگر آزمون چگونه است. در این مثال برای نمره ۳ در آزمون لنگر احتمالا نمرات بین۱۰ تا ۱۵ بیشترین فراوانی را دارند.

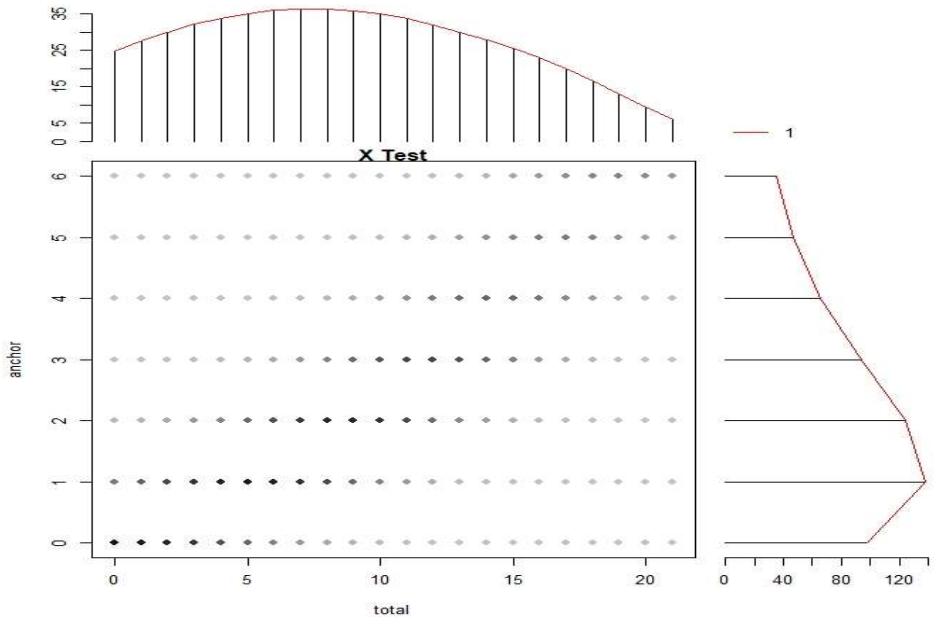


نمودار۱ توزیع شرطی نمرات در آزمون X به شرط نمره در آزمون لنگر، بدون بوت استراپینگ و فرایند هم‌وارسازی

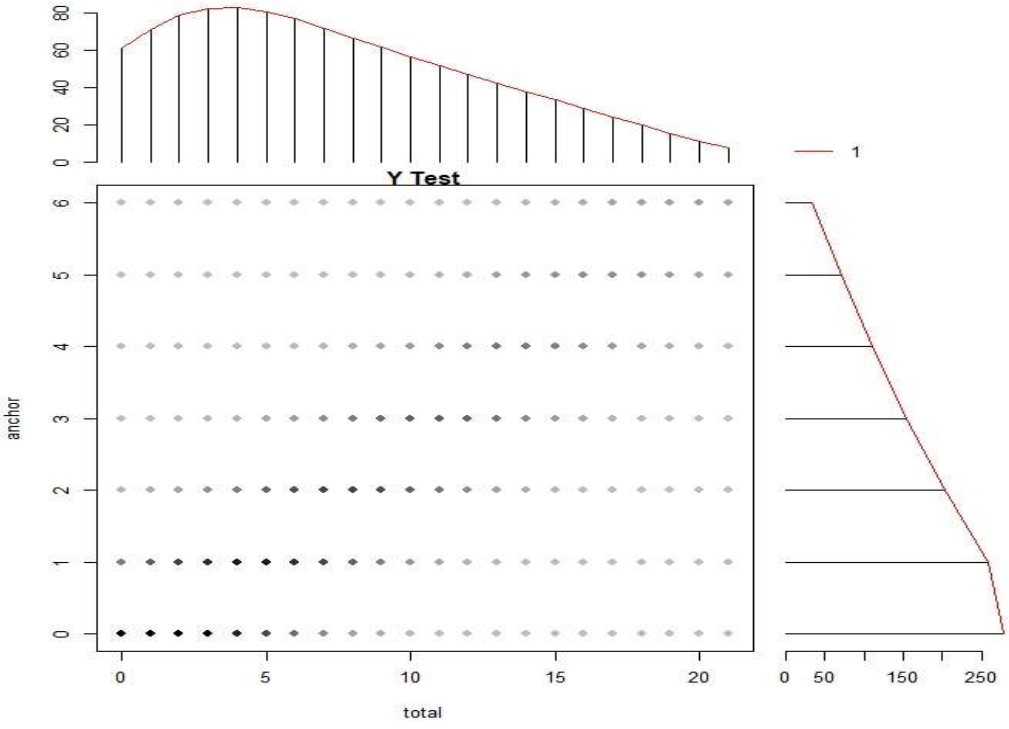


نمودار ۲ توزیع شرطی نمرات در آزمون Y به شرط نمره در آزمون لنگر، بدون بوت استراپینگ و فرایند هم‌وارسازی

دو نمودار زیر همان داده‌ها می‌باشند هنگامی که فرایند بوت استراپینگ برای هموار سازی توزیع لگاریتمی آنها صورت گرفته.



نمودار ۳ توزیع شرطی نمرات در آزمون X به شرط نمره در آزمون لنگر، با بوت استراپینگ و فرایند هم‌وارسازی



نمودار٤ توزیع شرطی نمرات در آزمون Y به شرط نمره در آزمون لنگر،

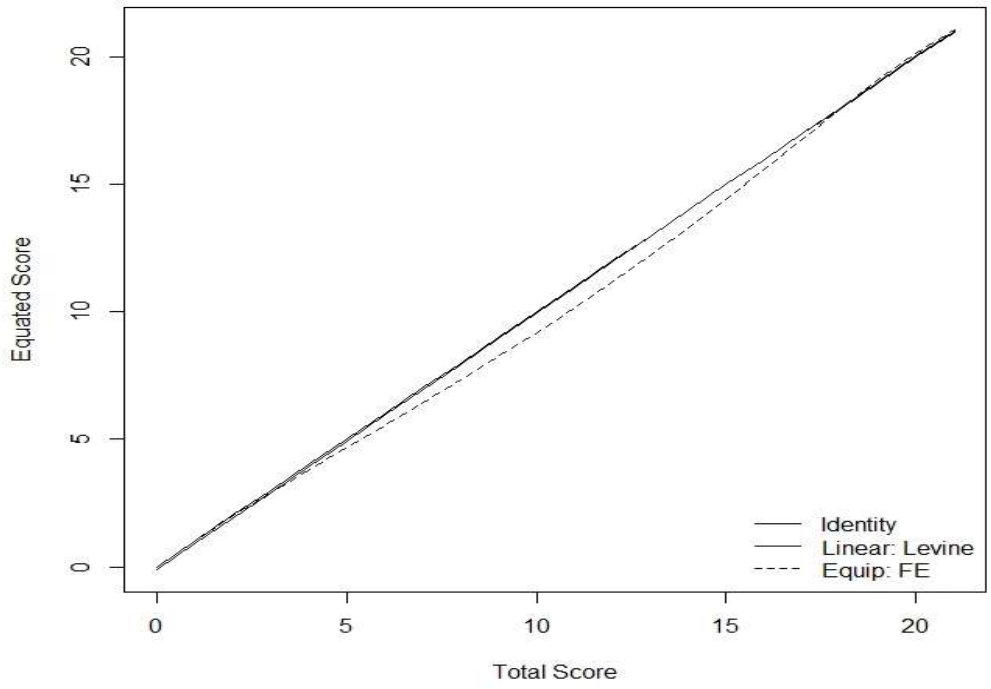
با بوت استراپینگ و فرایند هم‌وارسازی

جدول زیر توزیع نمرات برگشت داده شده آزمون X به آزمون Y را نشان می دهد. همان‌طور که می بینیم بر اساس هم‌ترازی هم‌صدک نمره ۱۳ در آزمون X برابر با نمره ۱۲/۲۲ در آزمون Y است و بر اساس هم‌ترازی خطی نمره ۱۳ در آزمون X برابر با نمره ۱۲/۹۸ در آزمون Y است.

جدول ۵ نمرات تبدیل‌شده آزمون جدید X بر روی آزمون قدیم Y

بر اساس "هم‌ترازی هم‌صدک در مقایسه با هم‌ترازی خطی" در طرح گروه‌های نابرابر با آزمون لنگر

هم‌ترازی هم‌صدک		هم‌ترازی خطی	
آزمون جدید x	هم‌ترازی yx	آزمون جدید x	هم‌ترازی yx
۰	-۰/۰۰۲	۰	-۰/۱۲٤
۱	۰/۹۸۷	۱	۰/۸۸
۲	۱/۹۵	۲	۱/۸۹
۳	۲/۸۹	۳	۲/۹۰
٤	۳/۷۹	٤	۳/۹۱
٥	٤/٦۸	٥	٤/۹۲
٦	٥/٥٥	٦	٥/۹۳
۷	٦/٤۳	۷	٦/۹۳
۸	۷/۳۲	۸	۷/۹٤
۹	۸/۲٤	۹	۸/۹٥
۱۰	۹/۱۸	۱۰	۹/۹٦
۱۱	۱۰/۱٦	۱۱	۱۰/۹۷
۱۲	۱۱/۱۷	۱۲	۱۱/۹۷
۱۳	۱۲/۲۲	۱۳	۱۲/۹۸
۱٤	۱۳/۳۱	۱٤	۱۳/۹۹
۱٥	۱٤/٤۳	۱٥	۱٤/۱۰
۱٦	۱٥/٦۰	۱٦	۱٦/۰۱
۱۷	۱٦/۷۹	۱۷	۱۷/۰۱
۱۸	۱۷/۹۸	۱۸	۱۸/۰۲
۱۹	۱۹/۱۱	۱۹	۱۹/۰۳
۲۰	۲۰/۱۷	۲۰	۲۰/۰٤
۲۱	۲۱/۱۰	۲۱	۲۱/۰٥



نمودار۵ توزیع نمره هم‌تراز شده دو آزمون X و Y به دو روش هم‌صدک و خطی در مقایسه با یکدیگر

نتایج تابع هم‌ترازی هم‌صدک هم‌زمان سه فرم آزمون X ، Y و Z بدون سوالات لنگر

در این بخش هم‌زمان سه فرم آزمون X و Y و Z یک‌جا و بدون در نظر گرفتن سوالات لنگر هم‌تراز شده‌اند. برای کاهش خطای معیار برآورد، توزیع داده‌های هر سه آزمون بر اساس لگاریتم خطی نخست هم‌وارسازی شده‌اند و سپس بر اساس هم‌ترازی هم‌صدک مقیاس آنها بر روی یکدیگر برگردانده شده است.

جدول ٦ آماره‌های توصیفی گشتاورهای سه آزمون X و Y و Z و تعداد n نمونه آزمون‌شوندگان

آزمون	میانگین	انحراف معیار	چولگی	کشیدگی	کمینه	بیشینه	nتعداد
X	۹/۰۳	۵/٤۷	۰/۲۰	۲/۱	۰	۲۱	٦۰۰
Y	۷/٦۱	۵/۲٦	۰/۵۲	۲/٤۲	۰	۲۱	۱۱۱۱
Z	۷/۸۰	۵/۱۰	۰/۳۹	۲/۲۵	۰	۲۰	۲۲۰۰

ردیف آخر جدول زیر مقادیر هم‌تراز شده آزمون X بر روی Y را نشان می دهد.

جدول ۷ آماره‌های توصیفی گشتاورهای دو آزمون X و Y و مقادیر هموار شده و هم‌تراز شده آزمون X به آزمون Y

پیونددادن هم‌صدک X به Y							
طرح: گروه‌های نابرابر							
روش هم‌وارسازی: پیش هم‌وارسازی با لگاریتم خطی							
خلاصه آماری							
آزمون	میانگین	انحراف معیار	چولگی	کشیدگی	کمینه	بیشینه	nتعداد
x.count	۹/۰۳	۵/٤۷	۰/۲۰	۲/۱۰	۰	۲۱	٦۰۰

x.smooth	۹/۰۳	۵/٤۷	۰/۲۰	۲/۱۰	۰	۲۱	٦۰۰
y.count	۷/٦۱	۵/۲٦	۰/۵۲	۲/٤۲	۰	۲۱	۱۱۱۱
y.smooth	۷/٦۱	۵/۲٦	۰/۵۲	۲/٤۲	۰	۲۱	۱۱۱۱
yx.obs	۷/٦۱	۵/۲٦	۰/۵۱	۲/۳۸	-۰/۱٤	۲۰/۷۷	٦۰۰

ردیف آخر جدول زیر مقادیر هم‌تراز شده آزمون Y بر روی Z را نشان می دهد.

جدول ۸ آماره‌های توصیفی گشتاورهای دو آزمون Y و Z و مقادیر هموار شده و هم‌تراز شده آزمون Y به آزمون Z

پیونددادن هم‌صدک Y به Z							
طرح: گروه‌های نابرابر							
روش هم‌وارسازی: پیش هم‌وارسازی با لگاریتم خطی							
خلاصه آماری							
آزمون	میانگین	انحراف معیار	چولگی	کشیدگی	کمینه	بیشینه	nتعداد
y.count	۷/٦۱	۵/۲٦	۰/۵۲	۲/٤۲	۰	۲۱	۱۱۱۱
y.smooth	۷/٦۱	۵/۲٦	۰/۵۲	۲/٤۲	۰	۲۱	۱۱۱۱
z.count	۷/۸۰	۵/۱۰	۰/۳۹	۲/۲۵	۰	۲۰	۲۲۰۰
z.smooth	۷/۸۰	۵/۱۰	۰/۳۹	۲/۲۵	۰	۲۰	۲۲۰۰
zy.obs	۷/۷۹	۵/۱۰	۰/۳۹	۲/۲٦	۰/۰۸	۲۰/۰۷	۱۱۱۱

جدول زیر نمرات آزمون جدید X و نمرات تبدیل‌شده X به Y و همچنین نمرات تبدیل‌شده Y به Z را نشان می دهد. انتظار می رود فردی که در آزمون X نمره ۱۳ گرفته باشد این نمره در آزمون Y معادل با ۱۱/۰۴ و در آزمون Z همین نمره معادل با ۱۳/۱۲ باشد.

جدول ۹ نمرات تبدیل‌شده آزمون جدید X بر روی آزمون قدیم Y

و آزمون Y بر روی آزمون قدیم تر Z بر اساس هم‌ترازی هم‌صدک با هم‌وارسازی (لگاریتم خطی)

آزمون جدید X	برگردان X بر روی Y x1y2	برگردان Y بر روی Z y2z3
۰	-۰/۱٤	۰/۰۸
۱	۰/٦۱	۱/۱۸
۲	۱/۳۵	۲/۲٤
۳	۲/۱۱	۳/۲۹
٤	۲/۸۸	٤/۳۲
۵	۳/٦۷	۵/۳٤
٦	٤/٤۸	٦/۳٦
۷	۵/۳۱	۷/۳٦
۸	٦/۱۷	۸/۳٦
۹	۷/۰۷	۹/۳۵
۱۰	۷/۱۰	۱۰/۳۲
۱۱	۸/۹۷	۱۱/۲۷
۱۲	۹/۹۸	۱۲/۲۱
۱۳	۱۱/۰٤	۱۳/۱۲
۱٤	۱۲/۱۵	۱٤/۰۲

۱۵	۱۳/۲۹	۱٤/۹۰
۱٦	۱٤/٤٥	۱٥/۷۷
۱۷	۱٥/٦۷	۱٦/٦۲
۱۸	۱٦/۹۱	۱۷/٤٦
۱۹	۱۸/۱٦	۱۸/۳۲
۲۰	۱۹/٤۳	۱۹/۱۹
۲۱	۲۰/۷۷	۲۰/۰۷

بحث

در این پژوهش به طور خلاصه روش‌های هم‌ترازسازی مناسب آزمون های سازمان سنجش در چهارچوب تئوری کلاسیک بیان شدند. با توجه به گسترگی روش‌های هم‌ترازسازی سعی شده تنها به روش‌هایی پرداخته شود که مناسب سازمان سنجش برای هم‌ترازکردن نمرات آزمون سراسری باشد. هر موقعیت و یا نوع آزمون (با لنگر و یا بدون لنگر) پیش‌بینی شده و سپس توابع هم‌ترازسازی برای آن موقعیت‌ها اجرا و نتایج آنها به صورت عملی توضیح و تشریح شده است. بسته به ویژگی‌ها و شرایط آزمون و هدف از آزمون هر یک از روش‌های هم‌ترازی نقاط قوت و ضعف خود را دارند. به عنوان مثال تابع هم‌ترازی با رتبه‌های صدکی هنگامی مناسب هستند که نتوان فرض نرمال بودن توزیع نمرات گروه‌های آزمون‌شونده را پذیرفت و یا توزیع داده‌ها چولگی داشته باشند. اگر توزیع نمرات دارای چولگی مثبت باشد و نمرات بالا فراوانی کمی داشته باشند این روش مناسب و سر راست است. اما این روش تنها به شما می‌گوید چه نسبتی از جامعه یک نمره معین را گرفته‌اند و بدین ترتیب با این نسبت های یکسان هر نمره را با نمره دیگر هم‌تراز می‌کند. در آزمون سراسری هم‌تراز کردن نمرات نه تنها مستلزم نسبت (یا فراوانی) جامعه آزمون‌شوندگان در گرفتن یک نمره معین است، بلکه دشواری فرم‌های آزمون هم باید در نظر گرفته شود: به عبارتی نیازمند روشی است که بتواند علاوه بر تفاوت توزیع توانایی آزمون‌شوندگان، تفاوت توزیع دشواری آزمون‌ها را نیز کنترل کند. بنابراین روش‌هایی لازم است که لایه های درونی تر آزمون را هم در نظر بگیرد. به عنوان مثال هم‌ترازی خطی، گشتاورهای یکم و دوم (یعنی میانگین و واریانس) توزیع نمرات آزمون را نیز در نظر می گیرد.

اما، یکی از ایراداتی که هم‌ترازی خطی دارد این است که در این نوع هم‌ترازی تنها نمراتی که در دامنه میانگین قرار دارند به خوبی هم‌تراز می‌شوند اما در مقادیر فرین توزیع، یعنی همان نقاطی که رتبه‌های بسیار خوب قرار دارند به دلیل فراوانی بسیار کم آنها به خوبی هم‌تراز نمی‌شوند. بویژه در هم‌ترازی خطی فرض اصلی بر این است که توزیع دو فرم آزمون همانند یکدیگر هستند و تنها در میانگین و انحراف معیار با هم تفاوت دارند. عمدتا دیده می‌شود بسیاری از توزیع نمرات آزمون‌های سازمان سنجش ممکن است تحت تاثیر دشواری سوالات فرم‌ها و یا توانمندی افراد آزمون‌شونده در دوره‌های مختلف یکسان نباشند. در این موارد آزمون‌هایی که سوالات لنگر دارند برای هم‌ترازکردن این نمرات مناسب‌تر است. اما با توجه به محدودیت‌های سازمان که امکان اجرای آزمون‌هایی با سوالات لنگر ندارد و از طرف دیگر تفاوت توزیع داده‌های فرم‌های آزمون سراسری و از همه مهمتر چولگی راست توزیع نمرات، هم‌ترازی هم‌صدک بهینه ترین روش برای برگرداندن مقیاس نمرات بر روی یکدیگر است. در این روش نمرات تبدیل به مقادیر لگاریتمی می‌شوند، سپس هم‌وارسازی صورت گرفته و سپس بر اساس تابع هم‌ترازی هم‌صدک که بیشتر یک آماره ناپارامتریک است نمرات افراد رتبه‌بندی و سپس هم‌تراز می‌شوند بنابراین در برآورد نمرات و هم‌ترازکردن آنها ناهمانندی توزیع نمرات فرم‌های آزمون کمتر اثر می‌گذارد.

نتیجه‌گیری

الف-برای آزمون‌هایی که می‌توان سوال لنگر تهیه کرد:

"هم‌ترازی هم‌صدک با سوالات لنگر" لایه های ژرف‌تری از آزمون‌ها را با یکدیگر هم‌تراز می‌کند. اساسا هم‌ترازی با سوالات لنگر به همین منظور توسعه داده شده است. برای هر درسی که امکان دارد (مثل ریاضی و یا فیزیک) بهتر است از هم‌ترازی با سوالات لنگر استفاده شود. با توجه به شرایط آزمون سراسری که در آن امکان دارد گروه‌های نابرابری هر فرم آزمون را پاسخ دهند، سوالات لنگر برای برخی دروس می‌تواند دشواری متفاوت آزمون‌ها روی گروه‌های آزمون‌شونده (یا شرکت کنندگانی) با توزیع توانایی‌های متفاوت را تعدیل کند و در نهایت نمرات با خطای کمتری

هم‌تراز می‌شوند. از طرفی به دلیل چولگی راست توزیع نمرات فرم‌ها "هم‌ترازی هم‌صدک" چون هم از مزایای آمار رتبه ای (ناپارامتریک) استفاده می‌کند و هم از ویژگی‌های آمارهای پارامتریک، برای تبدیل داده‌ها و سپس هم‌وارسازی آنها بسیار مناسب است.

ب-برای آزمونهایی که نمی‌توان سوال لنگر تهیه کرد:

اما در صورت نبود امکان تهیه آزمون‌های لنگر، از تابع هم‌ترازی هم‌صدک با تبدیل خطی و هم‌وارسازی (اما بدون سوال لنگر) استفاده شود. همان‌طور که در پیشینه نظری این پژوهش نیز گفته شد اگر تابع توزیع تجمعی $P_T(x)$ یک آزمون (مثلا آزمون X) پیوسته باشد و هم‌زمان بر اساس مفروضه‌های توزیع نرمال، تابع توزیع نمرات P_0 آن دارای توزیعی با میانگین صفر $\mu_{XT} = 0$ و واریانس یک $\sigma_{XT}{}^2 = 1$ باشند، تابع توزیع هم‌صدک برابر با تابع توزیع خطی خواهد بود: $P_T(x) = P_0$ (به فرمول های ۱۰ و ۱۱ نگاه کنید). قاعده کلی این است اگر در داده‌ها مفروضه‌های توزیع نرمال وجود داشته باشد، حتی اگر روش‌های ناپارامتریک برای داده‌ها استفاده شود نتایج یکسانی بدست می‌آید. با این برتری که مدل‌های ناپارامتریک در مقابل انحراف‌های جزئی از مفروضه‌های توزیع‌های نرمال مقاومت بیشتری دارند و برآوردهای واقعی‌تری از رتبه‌بندی افراد به دست می‌دهند. بنابراین اگر توزیع داده‌ها چوله باشد (مثل نمرات آزمون‌های سراسری) بهتر است از هم‌ترازی هم‌صدک با هم‌وارسازی استفاده شود.

هم‌ترازی هم‌صدک از مزایای مفروضه‌های هم توزیع‌های پارامتریک (هم‌وارسازی) و هم توزیع های ناپارامتریک (هم‌صدک) استفاده می‌کند؛ در هم‌ترازی هم‌صدک، چون مبنای هم‌ترازی آمارهای رتبه‌ای است، چولگی توزیع نمرات، رتبه‌بندی افراد را متاثر نمی‌سازد. اما باید مراقب بود هنگامی که برای هم‌وارسازی از روش‌های متفاوت بوت استراپینگ (به عنوان مثال تعداد نمونه‌های متفاوتی برای هم‌وارسازی) استفاده می‌شود ممکن است نتایج متفاوتی بدست آید. در این پژوهش نمونه‌های ۱۰ تایی استفاده شده. برای داده‌های آزمون‌های سازمان سنجش باید این مقادیر نمونه‌ای در یک مقدار معین تثبیت شود تا نمرات بدست آمده در دوره‌های مختلف آزمون تحت تاثیر این نمونه گیری‌های متفاوت قرار نگیرد. هم‌چنین پیش از اجرای هم‌ترازی بهتر است برای افزایش اعتبار فرایند هم‌ترازی، "مقادیر مربوط به غایبان" کنار گذاشته شود.

به هر حال، یکی از مشکلات هم‌ترازی نمرات پدیده آماری «بازگشت به میانگین» است. فرانسیس گالتون در قرن نوزدهم در مقاله‌ای که در این زمینه منتشر کرده بود، ادعا کرد که متوسط قد پسران دارای پدران قد بلند، کمتر از قد پدرانشان می‌شود و این در مورد پسرانی با پدرانی کوتاه قد بر عکس است. به عبارتی خصوصیات ارثی تمایل دارند به سوی میانگین برگشت کنند. کار گالتون در علم آمار به یکی از مهم‌ترین شاخص های علم یعنی همبستگی منجر شد که در فنون جدید برای روایی و پایایی آزمون‌ها بسیار به کار گرفته می‌شود (شولتز و شولتز، ۱۹۹۶، صص.۱۷۶-۱۸۰). این پدیده بازگشت به طرف میانگین در بسیاری از طرح‌های مختلف جمع‌آوری داده‌ها دیده شده. احتمالا در نمره کسانی که در سال دو بار آزمون بدهند نیز دیده خواهد شد؛ کسانی که در دور اول آزمون نمره بالایی گرفته‌اند نمره آنها در آزمون دوم به طور متوسط به طرف میانگین نمره جمعیت کاهش پیدا خواهد کرد. برعکس کسانی که در آزمون اول نمره پایینی گرفته‌اند در آزمون دوم به طور متوسط نمره آنها به طرف میانگین جمعیت افزایش پیدا خواهد کرد. در واقع به نظر می رسد هر فرد علاوه بر این‌که باید با دیگران رقابت کند با خودش نیز باید رقابت کند.

از دیگر مشکلات هنگامی که در سال تنها یک بار آزمون اجرا شود اگر به فرض سه رتبه اول (بهترین نمرات) پذیرفته شوند یعنی نمرات ۲۱، ۲۰ و ۱۹ حالا پس از اجرای هم‌ترازی باید نمرات ۲۱ از آزمون X، نمره ۲۰/۷۶۸ از آزمون Y و نمره ۲۰/۰۶۸ از آزمون Z انتخاب شوند و این همچنان رقابت را سخت تر خواهد کرد (به ستون و ردیف پایانی جدول ۹ نگاه کنید).

مشکل دیگر هنگامی پیش می‌آید که با کنار رفتن یک آزمون و جایگزین شدن یک آزمون دیگر نمرات آزمون‌هایی که پیشتر هم‌تراز شده‌اند دوباره تغییر می‌کند. اگر آزمون Z کنار برود بعد از دو سال مقادیر آزمون Y در هم‌ترازی های بعدی تغییر پیدا می‌کند. توجیه و اقنای آزمون‌شوندگان روی نمره این آزمون در دوره‌های متوالی ممکن است دشوار باشد: شش ماه پیش نمره یک فرد در Y برابر با ۲۰/۷۶۸ بوده است، با کنار رفتن آزمون Z و جایگزین شدن یک آزمون دیگر همین نمره ممکن است کاهش و یا افزایش پیدا کند و نیاز به کارنامه جدید باشد.

با توجه به گفته دست اندرکاران تهیه آزمون‌ها به دلیل محدودیت محتوای برخی از دروس آزمون‌ها، تهیه سوالات مشترک (لنگر) برای آنها امکان‌پذیر نیست. یا به دلیل انتشار آنها در جامعه قابلیت تهیه سوالات مشترک در هر فرم آزمون طی دورهای متوالی وجود ندارند. واقعیت این است که بدون سوالات لنگر فرایند هم‌ترازی نمرات آزمون‌ها با توابع هم‌ترازی که در جدول ۱ آمده به فرایند مرتبط سازی یا پیوند نمرات کاهش پیدا می‌کند؛ تنها به کمک سوالات لنگر است که می‌توان دشواری آزمون‌ها را تعدیل کرد و سپس مقیاس نمرات را به روی یکدیگر انتقال داد و هم‌تراز کرد.

References

Albano, A., & Albano, M. A. (2018). Package ‘equate’. Available at: https://cran.r-project.org/web/packages/equate/index.html

Albano، A. D. (2016). Equate: An R package for observed-score linking and equating. Journal of Statistical Software, 74, 1-36.

Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Educational Testing Service.

Angoff, W. H. (1971). Scales, norms, and equivalent scores. In RL Thorndike (Ed.), Educational measurement.

Bahmanabadi, S., Falsafinejad, M., Delavar, A., Farrokhi, N., & Minaei, A. (2020). Identification of Optimal Equating Method in Multidimensional Tests. *Educational Measurement and Evaluation Studies*, 10(30), 217-264. doi: 10.22034/emes.2020.44489

Brennan, R.L. (2001). Generalizability Theory, Iowa Testing Programs. University of Iowa. Springerverlag, New York.

Chen,F., Huang, H. & MacGregor, D.,(2009). EQUATING OR LINKING: BASIC CONCEPTS AND A CASE STUDY. Originally presented at CAL, Washington. Available at: https://faculty.ecnu.edu.cn/picture/article/220/0c/13/03357e474db0b2d5de11abaef0fb/793ecb9d-fe0b-4ff5-bd56-78148d7d4210.pdf.x

Dorans, N. J., Moses, T. P., & Eignor, D. R. (2010). Principles and practices of test score equating. *ETS Research Report Series*, *2010*(2), i-41.

Heh, V. K. (2007). Equating accuracy using small samples in the random groups design (Doctoral dissertation, Ohio University). Available at: https://etd.ohiolink.edu/apexprod/rws_etd/send_file/send?accession=ohiou1178299995&disposition=inline

Hendrickson, A. B., & Kolen, M. J. (2001). IRT Equating of the MCAT. MCAT Monograph.

Kim, S. H., & Cohen, A. S. (1998). A comparison of linking and concurrent calibration under item response theory. Applied psychological measurement, 22(2), 131-143.

Levine, R. Equating the score scales of alternative forms administered to samples of different ability 1955 Princeton. *NJ Educational Testing Service (ETS Research Bulletin No. 55-23)*. Available at: https://onlinelibrary.wiley.com/doi/epdf/10.1002/j.2333-8504.1955.tb00266.x

Liang, Z., Zhang, M., Huang, F., Kang, D., & Xu, L. (2021). Application Innovation of Educational Measurement Theory, Method, and Technology in China’s New College Entrance Examination Reform. Chinese/English Journal of Educational Measurement and Evaluation, 2(1), 3.

Livingston, S. A. (2014). Equating test scores (without IRT). Educational testing service.

Liu, J., & Low, A. C. (2007). An exploration of kernel equating using SAT® data: Equating to a similar population and to a distant population. *ETS Research Report Series*, *2007*(1), i-22.

MoghadamZade, A.(2015). Optimal Smoothing Method of Data in Test Equating: The Case of TOLIMO and Comprehensive Trial Tests of Iran Educational Testing Organization. Quarterly of Educational Measurement, 6(21), 261-287. doi: 10.22054/jem.2015.5736

Muraki, E., Hombo, C. M., & Lee, Y. W. (2000). Equating and linking of performance assessments. Applied Psychological Measurement, 24(4), 325-337. Available at: https://www.researchgate.net/publication/247742704_Equating_and_Linking_of_Performance_Assessments

Parsaeian, M., NaghiZadeh, S., Naderi, H. (2018). Selection the best Method of Equating Using Anchor-Test Design in Item Response Theory . Andishe-_ye-Amari. Avalable at: http://andisheyeamari.irstat.ir/article-۱-۵۰۴-fa.html

Rezvanifar, S., Falsafinejad, M., Delavar, A., (2016). eguating methods. *Quarterly of Educational Measurement*, 7(26), 1-33. doi: 10.22054/jem.2017.2737.1085

Ryan, J., & Brockmann, F. (2009). A Practitioner's Introduction to Equating with Primers on Classical Test Theory and Item Response Theory. Council of Chief State School Officers.

Schultz, D. P., & Schultz, S. E. (1996). A History of Modern Psychology [1969]. Translated by: Saif, A., Sharifi, H. P., Ali Abadi, K. & Najafi Zand, J. (2005). Dowran publication

Seufert, B. (2012).When, why, and how the business analyst should use linear regression. Available at: https://mobiledevmemo.com/when-why-and-how-you-should-use-linear-regression/

Shea, J. A., & Norcini, J. J. (1995). Equating. Licensure Testing: Purposes, Procedures, and Practices. Edited by Impara JC. Lincoln, NE, Buros Center for Testing, 253-287.

Supreme Council of Cultural Revolution (2021). Policies and criteria for organizing assessment and acceptance Applicants for admission to higher education. Resolution no. 3217: avalable at: https://sccr.ir/pro/3217/

Swaminathan, H. (---) .linking and equating of test scores. University of Connecticut. Available at: https://slidetodoc.com/linking-and-equating-of-test-scores-hariharan-swaminathan/

Tian, F. (2011). A comparison of equating/linking using the Stocking-Lord method and concurrent calibration with mixed-format tests in the non-equivalent groups common-item design under IRT. Unpublished doctoral dissertation, Boston College.

von Davier, M., & von Davier, A. A. (2004). A unified approach to IRT scale linking and scale transformations. ETS Research Report Series, 2004(1), i-21.

Zolfagharnasab, S., Khodaei, E., & Yadegarzadeh, G. (2013). Optimum Weighting to Entrance Subtests and Their Items to Make Composite Score. *Educational Measurement and Evaluation Studies*, 3(4), 79-104.