

## Evaluation of the divergence of syllabus and questions using the goodness-of-fit test and the Jeffrey index (Case study: Graduation test statistics course questions 1393-1393)

Gholamreza  
Yadegarzadeh\*

Allame Tabatabai'i; university, Tehran

Adel Fatemi

Islamic Azad University-Sanandaj Branch

Mostafa Rastgar Agah

psychology and educational sciences  
department, Allame Tabatabai'i; university,  
Tehran

### Abstract

The purpose of this study is to investigate the divergence of the questions of the master's and doctoral exams of a number of statistical courses during the years 1397-1393 from the topics approved by the Supreme Planning Council. For this purpose, the main topics and sub-topics (with the title of main cluster, sub-cluster and topics) were first identified based on the course titles approved by the Ministry of Science, Research and Technology, on which the master's and doctoral entrance exams are designed. In the next step, the frequency distribution of the mentioned test booklet questions in specific time intervals in each topic was extracted and the distance of this distribution of questions in the clusters with the ideal and expected distribution of specialized experts was calculated and tested. The Chi-square test was used to examine the divergence and the Jeffrey index was used to evaluate the good fit of the two distributions as well as the information distance. The sample studied in this study is a case study consisting of a total of eight courses, two courses "Statistics and Experimental Design" in the master's degree exam in agricultural engineering and also four specialized master's degree courses in statistics (mathematical statistics, regression, sampling and basics of probability) And two specialized doctoral courses in this field (probability and statistical inference). The results show that out of eight non-specialized courses in only one course (experimental design in the field of agricultural engineering) and of course in 2016 the rate of divergence was significant (P-Value = 0.28, Chi-square = 8.88 , Jeffrey). Also, among the specialized courses in the field of statistics, the degree of divergence from the ideal situation for one of the courses (probability course of the doctoral exam in statistics in 1396) is significant (0.46P-Value <0.01, Chi-square = 12.50, Jeffrey =).

**Keywords:** Divergence, course titles, master's and doctoral exams, goodness-of-fit test, Jeffrey divergence index

\* Corresponding Author: yadegarzadeh@gmail.com

**How to Cite:** yadegarzadeh, G., fatemi, A., & Rastgar Agah, M. (2021). Evaluation of the divergence of syllabus and questions using the goodness-of-fit test and the Jeffrey index (Case study: Graduation test statistics course questions 1393-1393). *Quarterly of Educational Measurement*, 12(45), 65-83.  
doi: 10.22054/jem.2022.56477.2097

## بررسی میزان واگرایی سرفصل دروس و سوال‌های آزمون با استفاده از آزمون نیکویی برازش و شاخص جفری (نمونه موردی: سوال‌های درس آمار آزمون تحصیلات تکمیلی ۱۳۹۷-۱۳۹۳)

غلامرضا یادگارزاده\*

استادیار دانشگاه علامه طباطبائی

عادل فاطمی

استادیار دانشگاه آزاد اسلامی واحد سنج

مصطفی رستگاره آگاه

دانشگاه علامه طباطبائی

### چکیده

هدف از این پژوهش بررسی میزان واگرایی سوال‌های آزمون کارشناسی ارشد و دکتری تعدادی از دروس اصلی آمار طی سال‌های ۱۳۹۳-۱۳۹۷ از سرفصل‌های مصوب شورای عالی برنامه‌ریزی است. برای این منظور ابتدا براساس سرفصل‌های دروس مصوب وزارت علوم، تحقیقات و فناوری، که آزمون‌های ورودی کارشناسی ارشد و دکتری بر اساس آنها طراحی می‌شوند، موضوع‌ها و زیر موضوع‌های اصلی (با عنوان خوشه اصلی، زیرخوشه و مباحث) مشخص شد. در مرحله بعد توزیع فراوانی سوالات دفترچه آزمون‌های مذکور در بازه‌های زمانی مشخص در هر مبحث استخراج شد و فاصله این توزیع سوالات در خوشه‌ها با توزیع ایده‌آل و مورد انتظار خبرگان رشته تخصصی محاسبه و آزمون شده است. برای بررسی واگرایی، از آزمون "خی‌دو" و برای بررسی نیکویی برازش دو توزیع و همچنین فاصله اطلاعاتی از "شاخص جفری" استفاده شده است. نمونه مورد بررسی در این تحقیق یک نمونه موردی شامل مجموعاً هشت درس، دو درس "آمار و طرح آزمایش‌ها" در آزمون کارشناسی ارشد رشته‌های مهندسی کشاورزی و همچنین چهار درس تخصصی کارشناسی ارشد رشته آمار (آمار ریاضی، رگرسیون، نمونه‌گیری و مبانی احتمال) و دو درس تخصصی دکتری این رشته (احتمال و استنباط آماری) است. نتایج نشان می‌دهد که از میان هشت درس غیر تخصصی تنها در یکی از دروس (طرح آزمایش‌ها در رشته مهندسی کشاورزی) و البته در سال ۱۳۹۵ میزان واگرایی معنی‌دار بوده است ( $P=0/28$ ). همچنین در بین دروس تخصصی رشته آمار نیز میزان واگرایی از وضعیت ایده‌آل برای یکی از دروس (درس احتمال آزمون دکتری آمار سال ۱۳۹۶) معنی‌دار است ( $P=0/46$ ).  $P=0/01$ ,  $Chi-square= 12/50$ ,  $Jeffrey$ ). در نتیجه می‌توان گفت در میان دروس آمار در آزمون ورودی رشته‌های مهندسی کشاورزی و همچنین دروس تخصصی رشته آمار، بین سرفصل و سوال‌های آزمون با وجود تغییرات، نسبتاً هماهنگی و تناسب رعایت شده است، بجز در دو ماده که میزان واگرایی و عدم تناسب معنی‌دار شده که این می‌تواند به دلیل عدم اطلاع طراحان سوال از سرفصل‌های دروس مذکور باشد.

کلیدواژه‌ها: واگرایی، سرفصل دروس، آزمون کارشناسی ارشد و دکتری تخصصی، آزمون نیکویی

برازش، شاخص واگرایی جفری

\* نویسنده مسئول: yadegarzadeh@gmail.com

## مقدمه

بر اساس استانداردهای سنجش و اندازه‌گیری تربیتی بین آنچه تدریس می‌شود (برنامه درسی) و آنچه مورد آزمون قرار می‌گیرد بایستی تناسب و هماهنگی وجود داشته باشد (داونینگ<sup>۱</sup>، ۲۰۰۶) لذا در فرایندهای سنجش و اندازه‌گیری و آزمون‌های ورودی بحث اطلاع طراحان آزمون‌ها از برنامه درسی و یا سرفصل دروس ضرورت دارد. برنامه‌های درسی<sup>۲</sup> به تعبیری قلب تمامی فعالیت‌های آموزشی هستند (پریستلی<sup>۳</sup> و فیلیپو<sup>۴</sup>، ۲۰۱۹) و هر جا یادگیری در جریان است به نوعی برنامه‌های درسی حضور دارند. در تمامی نظام‌های آموزشی به منظور پاسخگویی، تعیین میزان تحقق اهداف، تضمین کیفیت، ورود افراد به مقاطع بالاتر و یا دفاع از عملکرد سیستم، یادگیری فراگیران مورد سنجش قرار می‌گیرد و همانطور که اشاره شد باید بین اهداف یا بروندهای یادگیری و برنامه‌های درسی تطابق وجود داشته باشد (بیگز<sup>۵</sup>، ۲۰۰۳). در مقابل برخی معتقدند نباید آنچه در نظام‌های آموزشی می‌گذرد را با سنجش خروجی‌های مورد انتظار ارزشیابی و قضاوت کرد (أرد<sup>۶</sup>، ۲۰۰۴) زیرا نمی‌توان یادگیری را با یک آزمون و یا تکلیف مورد سنجش قرار داد. ورای این بحث و جدل نظری در اکثر نظام‌های آموزشی از مکانیسم سنجش و آزمون برای قضاوت در خصوص عملکرد عناصر نظام آموزشی (که یادگیرنده نیز یکی از این عناصر است) و تصمیم‌گیری در خصوص وارد شدن آنها به مقاطع بالاتر استفاده می‌شود. طیف وسیعی از آزمون‌ها وجود دارد که هر یک دارای هدف و کاربرد خاص خود هستند. یکی از این نوع آزمون‌ها که در اکثر نظام‌های آموزشی نیز مورد استفاده قرار می‌گیرد آزمون‌های ورودی یا آزمون‌های خطیر و ملی است که از آنها با عنوان غربالگر (ناگی<sup>۷</sup>، ۲۰۰۰) نیز یاد می‌شود. هدف اصلی این آزمون‌ها عمدتاً کمک به تصمیم‌گیری مورد ورود افراد به مقاطع تحصیلی بالاتر است.

مرور مستندات پژوهشی نشان می‌دهد سنجش و یا ارزشیابی پیشرفت تحصیلی (خصوصاً در مواردی که نتیجه آن برای افراد مهم است) باید با آنچه در قالب برنامه درسی آموزش

1. Downing, S

۲. در نظام آموزش عالی کشورمان از آن تعبیر به سرفصل می‌شود.

3. Priestley, M

4. Philippou, S

5. Biggs, J

6. Ord, J

7. Nagy, P

داده شده سازگار باشد (هارگیس<sup>۱</sup>، ۲۰۱۳). در استانداردهای آزمودن روانی تربیتی<sup>۲</sup> تأکید شده است در سنجش آموخته‌ها باید آنچه مورد آزمون قرار می‌گیرد کاملاً مشخص و از پیش به اطلاع یادگیرندگان رسانده شود (اینور<sup>۳</sup>، ۲۰۱۳). در واقع می‌توان گفت بین نتایج یادگیری مورد انتظار و برنامه درسی باید تطابق وجود داشته باشد (کومار<sup>۴</sup>، ۲۰۱۶) که به نوبه خود تطابق سنجش و برنامه درسی را به دنبال دارد. این موضوع در سنجش خطیر و آزمون‌هایی که با هدف غربالگری به انجام می‌رسد اهمیت بیشتری دارد (ناگی، ۲۰۰۰/ ترجمه یادگارزاده و همکاران، ۱۳۸۶) زیرا نتایج این آزمون‌ها در تعیین سرنوشت افراد نقش زیادی دارند. بررسی‌ها نشان می‌دهد در سیستم‌های آموزشی متمرکز که دارای آزمون‌های ملی و یا منطقه‌ای هستند برنامه‌های درسی تحت تأثیر آزمون‌ها قرار می‌گیرند و برخی معتقدند اهداف یادگیری را تحت‌الشعاع قرار می‌دهد (ویلیام<sup>۵</sup>، ۲۰۰۱) و گاهی منجر به شکل‌گیری برنامه درسی پنهان می‌شود (یادگارزاده، ۱۳۹۹).

بین آنچه در کلاس درس می‌گذرد (چه در قالب برنامه و چه در قالب سرفصل) و آنچه مورد سنجش قرار می‌گیرد رابطه مستقیم وجود دارد. در آموزش عالی همانطور که آنالا<sup>۶</sup>، لیندن<sup>۷</sup> و مک‌کینن<sup>۸</sup> (۲۰۱۶) تأکید می‌کنند محتوایی که آموزش داده می‌شود در قالب برنامه درسی و یا سرفصل دروس بروز و ظهور پیدا می‌کنند. در نظام آموزش عالی کشورمان اساس کار آموزش سرفصل دروس است و برنامه‌های آموزشی در دانشگاهها بر ریل دروس اجرا می‌شود و برخی دروس به عنوان درس‌های اصلی در آزمون‌های ورودی مقطع بعدی مورد آزمون قرار می‌گیرند. در آزمون‌های ورودی در برخی از کشورها منابع درسی مشخصی که تدریس شده‌اند مورد آزمون قرار می‌گیرند اما در آزمون‌های تحصیلات تکمیلی کشورمان ملاک سرفصل دروس است و این امر ایجاب می‌کند بین محتوای آزمون و سرفصل دروس هماهنگی وجود داشته باشد. با توجه به تمرکز نظام برنامه‌ریزی درسی در آموزش عالی

- 
1. Hargis, C. H
  2. Standards for Educational and Psychological Testing
  3. Eignor, D. R
  4. Kumar, P. M
  5. Wiliam, D
  6. Annala, J
  7. Lindén, J
  8. Mäkinen, M

کشورمان این امر می‌تواند توجیه داشته باشند اما با واگذاری اختیار تدوین سرفصل دروس به دانشگاهها بحث تطابق آزمون‌های ورودی با سرفصل‌ها دچار چالش اساسی خواهد شد. مشکلی که معمولاً داوطلبان آزمون‌های کارشناسی ارشد و دکتری که براساس سرفصل طراحی می‌شوند از آن گلایه دارند ناهماهنگی و یا عدم تناسب بین محتوای آزمون‌ها با سرفصل دروس است. بر این اساس و با توجه به آنچه گفته شد سوال اصلی مقاله حاضر بر میزان تطابق سوال‌های آزمون‌های تحصیلات تکمیلی (کارشناسی ارشد و دکتری) با سرفصل دروس مصوب وزارت علوم، تحقیقات و فناوری است. روشن شدن ابعاد این سوال مستلزم توجه به دو موضوع مهم در نظام آموزش عالی است: آزمون‌های تحصیلات تکمیلی دانشگاهها و مراکز آموزش عالی و سرفصل‌های دروس. بر اساس بند "ث" ماده ۲ قانون "سنجش و پذیرش دانشجو در دوره‌های تحصیلات تکمیلی در دانشگاهها و مراکز آموزش عالی کشور"، تعیین دروس آزمون کارشناسی ارشد بر عهده شورای سنجش و پذیرش تحصیلات تکمیلی گذاشته شده که هر ساله در قالب دفترچه ثبت نام به اطلاع داوطلبان آزمون می‌رسد. همچنین در دفترچه ثبت نام داوطلبان به صورت روشن ذکر شده که سوال‌های آزمون براساس سرفصل‌های مصوب شورای عالی برنامه‌ریزی طراحی خواهد شد (دفترچه ثبت نام آزمون تحصیلات تکمیلی سال ۱۳۹۸). در طراحی و تولید آزمون تحصیلات تکمیلی نیز معمولاً این موضوع به طراحان آزمون توصیه می‌شود و سوال‌های خارج از سرفصل معمولاً غربال و اصلاح می‌شوند. بررسی سوال‌های آزمون در رشته‌های مختلف نشان می‌دهد تطابق بین آزمون و سرفصل‌های دروس همیشه کامل نیست و معمولاً داوطلبان پس از آزمون و انتشار سوال‌ها به آن اعتراض می‌کنند.

سرفصل‌های دروس در نظام آموزش عالی ما همواره مورد نقد بوده‌اند و مسائلی مانند قدیمی بودن، عدم تطابق با نیازهای روز (کرمی و فتاحی، ۱۳۹۲)، عدم تطابق با تحولات علم و فناوری (علوی، کیوان‌پناه، تاسی، ۱۳۹۶)، عدم تطابق بین واحدهای نظری و عملی (شب‌خیز و همکاران، ۱۳۹۱)، عدم توجه به پیش‌نیاز دروس و ارتباط افقی دروس (بهرام‌بگی، ۱۳۹۱)، نبود نگاه تخصصی و دقیق به موضوعها خصوصاً در هنر (حسن‌زاده، ۱۳۹۸)، دقیق نبودن اهداف و بروندادهای یادگیری، عدم تطابق با برنامه‌ها و اسناد بالادستی و عدم توجه به اصول تدوین برنامه‌های درسی (رضایی، ۱۳۹۵؛ غریب‌پور و توتونچی مقدم، ۱۳۹۴؛ کرمی و فتاحی، ۱۳۹۲) در پژوهش‌های مختلف مورد تأکید قرار گرفته است. در

خصوصاً موضوع این پژوهش تاکنون مطالعه‌ای مستند نشده و تحقیقات انجام شده عمدتاً متمرکز بر بررسی سرفصل دروس در رشته‌های مختلف است.

در خصوص تطابق برنامه درسی (سرفصل) با سنجش (آزمون) دیدگاهها در دو طیف طبقه‌بندی می‌شود طیف اول کسانی هستند که معتقدند باید تطابق و هماهنگی بین این دو باید وجود داشته باشد (اینور، ۲۰۱۳؛ جونز<sup>۱</sup>، سوترن<sup>۲</sup> و بریگهام<sup>۳</sup>، ۱۹۸۸؛ هاستینگ<sup>۴</sup>، هارشمن<sup>۵</sup> و یزیرسکی<sup>۶</sup>، ۲۰۱۷). حتی برخی طرفداران این دیدگاه معتقدند کتاب یا منبع طرح سوال نیز باید در اختیار آزمون شوندگان قرار داده شود. کسانی مانند ویلیام<sup>۷</sup> (۲۰۰۳)، بروکهارت<sup>۸</sup> (۲۰۱۱) و فادیا ناصر (۲۰۰۷) معتقدند هماهنگی بین سنجش و برنامه درسی در سنجش‌های خطیر ضرورت دارد و در سایر انواع سنجش که با هدف اندازه‌گیری میزان یادگیری انجام می‌شود خیلی ضرورت ندارد زیرا منجر به کاهش عملکرد فراگیران می‌شود و یادگیری طوطی‌وار را رواج می‌دهد.

در این میان موضوع مهم توجه به نوع نظام آموزشی و ساختار اجرای برنامه‌های درسی و سرفصل‌های دروس است که روی تطابق سنجش و برنامه درسی تأثیر می‌گذارد. در نظام آموزشی کشورمان به دلایلی این هماهنگی و تطابق باید فراهم باشد. این دلایل عبارتند از سرنوشت‌ساز بودن نتیجه آزمون و اثر زیاد آن بر زندگی داوطلبان، تمرکز کامل نظام آموزش عالی و ضرورت اجرای هماهنگ سرفصل دروس در دانشگاه‌های مختلف کشور، محدود بودن شیوه ورود به آموزش عالی (صرفاً از طریق آزمون)، تأکید اسناد بالادستی بر رعایت سرفصل دروس چه در فرایند آموزش و چه در فرایند سنجش و حاکمیت سیستم از بالا به پایین در نظام آموزش عالی. بررسی اعتراض‌های داوطلبان آزمون تحصیلات تکمیلی نشان می‌دهد نوعی واگرایی در سوال‌های آزمون وجود دارد و طراحان آزمون کاملاً به سرفصل دروس وفادار نیستند. این موضوع به دلایلی از قبیل قدیمی بودن سرفصل، ناهماهنگی با تحولات علم و فناوری، واگذاری اختیار تدوین برنامه‌های درسی به برخی

- 
1. Jones, E. D
  2. Southern, W. T
  3. Brigham, F. J
  4. Husting, C
  5. Harshman, J
  6. Yeziarski, E
  7. Wiliam, D
  8. Brookhart, S. M

دانشگاهها و اختیار استادان در اجرای سرفصل دروس به صورت کاملاً میسر نمی‌شود. با توجه به آنچه گفته شد در پژوهش حاضر موضوع واگرایی سوال‌های درس آمار از سرفصل دروس مصوب شورای عالی برنامه‌ریزی مورد مطالعه و بررسی قرار گرفته است.

## روش

روش تحقیق مورد استفاده در این پژوهش تحلیل مستندات است. تحلیل مستندات به عنوان یک روش تحقیق کیفی عمدتاً با هدف کاوش مستندات منتشر شده (چاپی و الکترونیکی) و بررسی آنها از ابعاد و زوایای مختلف است (بوون<sup>۱</sup>، ۲۰۰۹). بر این اساس در این تحقیق ابتدا سرفصل دروس رشته کارشناسی و کارشناسی ارشد آمار (دروس آمار یاضی، رگرسیون، نمونه‌گیری و مبانی احتمال در آزمون ارشد و دروس احتمال و استنباط آماری در آزمون دکتری) و همچنین دورس آمار و احتمالات آماری و طرح و تجزیه آزمایش‌های رشته مهندسی کشاورزی در سال‌های ۱۳۹۴ تا ۱۳۹۷ انتخاب و در گام نخست براساس موضوعات خوشه‌بندی شدند. پس از خوشه‌بندی و اختصاص کد به شاخه‌ها و زیرشاخه‌ها در گام دوم براساس آنچه در سنوات گذشته و در آزمون‌های برگزار شده توسط سازمان سنجش آموزش کشور توزیع فراوانی سوالات در شاخه‌ها و زیرشاخه‌ها محاسبه شد. در گام سوم و براساس نظر خبرگان حوزه آموزش که متولیان اصلی در تدوین سرفصل‌ها نیز هستند فراوانی سوال‌ها به تفکیک زیرخوشه‌ها جهت پوشش صحیح سرفصل‌ها و البته براساس اهمیت موضوعات تحت عنوان توزیع فراوانی در وضعیت مورد انتظار و یا به عبارتی ایده‌آل (که توسط استادانی که دروس مورد نظر را حداقل ۳ ترم در دوره‌های کارشناسی و کارشناسی ارشد تدریس می‌کنند تعیین شد. در هر درس ۳ استاد در این فرایند مشارکت داشتند.) در جدول ثانویه و در مقابل جدول نخست که وضعیت موجود می‌باشد وضعیت مطلوب نیز آمده است. جهت تطبیق و بررسی میزان واگرایی وضعیت موجود در سنوات گذشته و وضعیت ایده‌آل از آزمون نیکویی برازش براساس آماره خی دو و همچنین شاخص اطلاعاتی جفری که نسخه متقارن شاخص واگرایی کولبک-لایبلر<sup>۲</sup> می‌باشد محاسبه شد. دلیل اینکه از آزمون خی دو و شاخص جفری استفاده شده این است که شاخص جفری میزان انحراف و یا واگرایی دو توزیع را براساس شاخص اطلاعاتی آنتروپی بسیار دقیق تر محاسبه

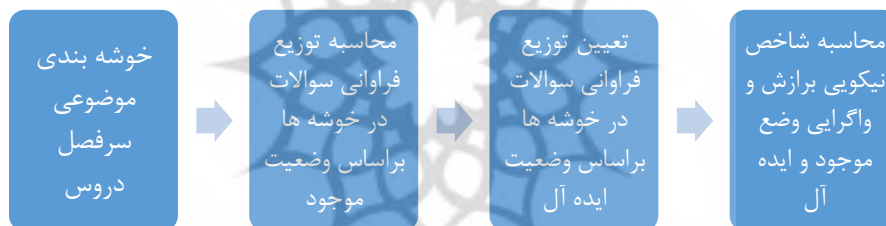
1. Bowen, G. A

2. Kullback-Leibler divergence

می نماید، زیرا شاخص واگرایی اطلاعاتی جفری نشان دهنده میزان عدم انطباق اطلاعات موجود در دو توزیع است و از سویی حالت متقارن شاخص کولبک-لایبلر است. دلیل استفاده از شاخص خی دو نیز آزمون پذیری این شاخص به منظور تعیین معنی داری و عدم معنی داری میزان واگرایی است. به عبارت دیگر شاخص جفری میزان هدر رفت اطلاعات ناشی از پذیرش توزیع فراوانی سوالات در حالت واقعی به جای توزیع مورد انتظار در سرفصل دروس مذکور و تدریس شده در دانشگاه ها است. البته این شاخص جایگزین آزمون خی دو نشده است، بلکه به عنوان مکمل در کنار آزمون خی دو آورده شده است. زیرا در روش خی دو به دلیل ادغام خانه ها با فراوانی کمتر از ۵ مورد عملاً شاخص قدرت تفکیک را از دست خواهد داد.

در شکل ۱ مراحل فوق بصورت خلاصه آمده است.

شکل ۱. مراحل انجام تحقیق



از مزایای روش خی دو که یک آزمون شناخته شده در آزمون نیکیویی برازش می باشد محاسبه سطح معنی داری در پذیرش و یا رد فرضیه واگرایی وضع موجود و ایده آل است. اما یکی از اصلی ترین معایب آن در این تحقیق کم بودن تعداد سوالات در برخی دروس و نیاز بیش از حد به ادغام خوشه ها است. بنابراین در بسیاری از موارد پژوهشگران ناچار به محاسبه شاخص واگرایی دو توزیع یعنی شاخص واگرایی جفری<sup>۱</sup> (جفری، ۱۹۴۶) شدند. این شاخص نسخه متقارن فاصله کولبک لایبلر می باشد که براساس شاخص آنتروپی ساخته می شود و برای دو توزیع  $f$  و  $g$  برابر است با:

1. Jeffrey's divergence (JD)



$$J(f, g) = \sum f(x) \log\left(\frac{f(x)}{g(x)}\right) + \sum g(x) \log\left(\frac{g(x)}{f(x)}\right)$$

فرمول فوق تکامل فرمول آنتروپی طی مراحل زیر است:

ابتدا شاخص آنتروپی یک توزیع آماری با تابع چگالی  $f$  عبارت است از:

$$H(f) = - \sum f(x) \log(f(x))$$

این شاخص نشان دهنده میزان بی‌نظمی و یا به عبارتی میزان اطلاعات نهفته در توزیع احتمال  $f$  است. با توسعه مفاهیم آنتروپی شاخصی اطلاعاتی توآمان از دو توزیع  $f$  و  $g$  به صورت زیر توسط کولبک و لایبلر ارائه گردید، این شاخص که شاخص واگرایی کولبک لایبلر نامیده شد که نشان دهنده میزان عدم تقارن اطلاعاتی دو توزیع است:

$$K(f, g) = \sum f(x) \log\left(\frac{f(x)}{g(x)}\right)$$

این شاخص اگرچه با توجه به ساختار آنتروپی شاخصی بسیار مناسب در ارزیابی واگرایی دو توزیع آماری بود اما کماکان نیازمند تکامل بیشتری بود، دلیل آن به عدم تقارن شاخص کولبک-لایبلر بازمی‌گشت، زیرا:

$$K(f, g) \neq K(g, f)$$

برای رفع این مشکل جفری تغییری بسیار ساده به شکل زیر انجام داد:

$$J(f, g) = K(f, g) + K(g, f)$$

کاملاً مشخص است که این تغییر کوچک در شاخص باعث شده است ضمن حفظ ویژگیهای شاخص پیشین در شاخص جدید ویژگی تقارن نیز اعمال شود.

پس از اتمام مراحل فوق و محاسبه واگرایی‌ها نتایج براساس سنوات و دروس مقایسه

شد.

جدول ۱. نمونه خوشه‌بندی موضوعی و فراوانی سوالات طرح شده در هر بخش مربوط به درس روش‌های نمونه‌گیری آزمون کارشناسی ارشد رشته آمار

نام درس	خوشه اصلی	زیر خوشه	مباحث
روش‌های نمونه‌گیری	مفاهیم پایه‌ای 120711		جامعه، نمونه، واحد نمونه‌گیری، چارچوب، سرشماری، مزایای نمونه‌گیری، مراحل نمونه‌گیری <b>۹۶</b>
		با جایگذاری 1207121	نحوه و انتخاب نمونه محاسبه برآورد میانگین و واریانس و برآوردکننده آن <b>۹۳ ۹۴ ۹۶</b>
	بدون جایگذاری 1207122	نمونه‌گیری تصادفی ساده 120712	نحوه و انتخاب نمونه محاسبه برآورد میانگین و واریانس و برآوردکننده آن <b>۹۳ ۹۴ ۹۴ ۹۴ ۹۴ ۹۵ ۹۵ ۹۵ ۹۵ ۹۶ ۹۶ ۹۶</b>
		برآورد نسبت‌ها 120713	محاسبه برآورد نسبت و میانگین و واریانس آن <b>۹۳ ۹۵</b>
		تعیین حجم نمونه 120714	تعیین حجم نمونه به روش‌های فوق و روش نامساوی چیچف <b>۹۴ ۹۶</b>
	نمونه‌گیری طبقه‌بندی	روش نمونه‌گیری طبقه‌بندی 1207151	تخصیص متناسب، بهینه، نیمن، تعیین برآورد میانگین و واریانس برآوردکننده <b>۹۳ ۹۳ ۹۳ ۹۳ ۹۴ ۹۴ ۹۴ ۹۵ ۹۵ ۹۵ ۹۵ ۹۵ ۹۶</b>
		120715	مقایسه با روش مقایسه با روش نمونه‌گیری تصادفی ساده <b>۹۳ ۹۴ ۹۵</b>

نام درس	خوشه اصلی	زیر خوشه	مباحث
	نمونه‌گیری با احتمال متغیر	کلیات و روش نمونه‌گیری با احتمال متغیر	
	120716		۹۳

به عنوان نمونه در جدول ۱ خوشه‌بندی موضوعی و همچنین توزیع فراوانی سوال‌های آزمون کارشناسی ارشد یکی از دروس رشته آمار، روش‌های نمونه‌گیری، به تفکیک چهار سال مشاهده می‌شود. همانطور که در جدول بالا مشاهده می‌شود برخی از موضوع‌های مورد تأکید در سرفصل (مانند ورش نمونه‌گیری تصادفی بدون جایگذاری) دروس بیشتر مورد توجه طراحان آزمون بوده و برخی دیگر یا مورد نظر نبوده و یا فقط در یک سال (مانند نمونه‌گیری با احتمال متغیر) مورد آزمون قرار گرفته‌اند. فعالیت مشابه برای ۷ درس آزمون کارشناسی ارشد و دکتری رشته آمار و همچنین دروس آمار آزمون کارشناسی ارشد رشته مهندسی کشاورزی انجام شد.

جدول ۲. فراوانی مورد انتظار و فراوانی‌های مشاهده شده به تفکیک خوشه و یا زیرخوشه‌های درس روش‌های نمونه‌گیری آزمون کارشناسی ارشد رشته آمار

کد خوشه یا زیر خوشه	فراوانی مورد انتظار (وضعیت ایده آل)	فراوانی مشاهده شده (وضعیت موجود)			
		سال ۱۳۹۳	سال ۱۳۹۴	سال ۱۳۹۵	سال ۱۳۹۶
120711	1	0	0	0	1
1207121	2	1	1	0	1
1207122	2	1	4	4	3
120713	1	1	0	1	0
120714	1	0	1	0	1
1207151	2	4	3	5	2
1207152	1	1	1	1	0
120716	0	1	0	0	0

با توجه به اطلاعات جدول فوق و به منظور بررسی انطباق فراوانی‌های مشاهده شده و فراوانی ایده آل و بررسی فرضیه واگرایی دو توزیع از آزمون نیکویی برازش شاخص خی دو استفاده گردید همچنین شاخص جفری نیز برای بررسی آنتروپی محاسبه شد.

۱. مبنای این مقایسه دفترچه سوال‌های آزمون کارشناسی ارشد و دکتری در سال‌های ۱۳۹۳ تا ۱۳۹۶ و سرفصل کارشناسی (مصوب ۱۳۸۸/۰۲/۲۶) و کارشناسی ارشد آمار ریاضی (مصوب ۱۳۷۴/۱۰/۱۰) بوده است.

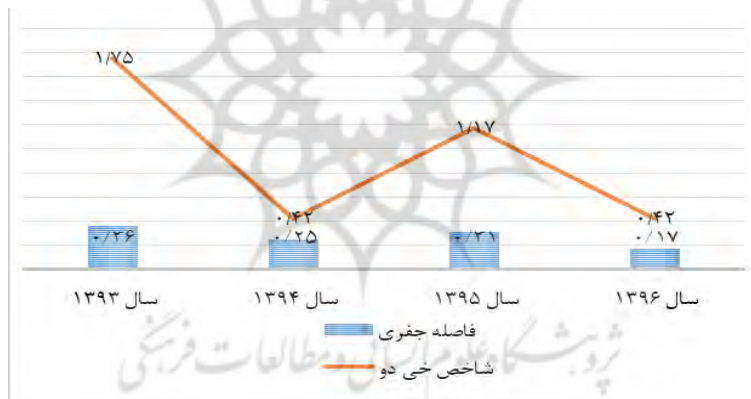
جدول ۲. مقادیر آماره خی دو و شاخص جفری به همراه آزمون معنی داری واگرایی فروانی سوالات و وضعیت ایده آل درس روش های نمونه گیری آزمون کارشناسی ارشد رشته آمار

سال	۱۳۹۳	۱۳۹۴	۱۳۹۵	۱۳۹۶
شاخص خی دو	۱/۷۵	۰/۴۲	۱/۱۷	۰/۴۲
معنی داری	ns	ns	ns	ns
فاصله جفری	۰/۳۶	۰/۲۵	۰/۳۱	۰/۱۷

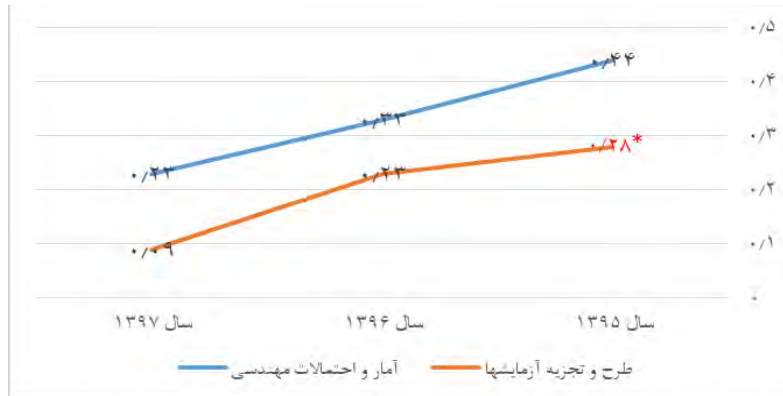
ns: عدم معنی داری ( $P > 0.05$ )

همانگونه که در جدول ۲ مشاهده می شود و براساس نتایج آزمون نیکویی برازش در تمام سنوات مورد بررسی در سطح  $0.05$  تفاوت معنی داری بین وضعیت موجود و ایده آل مشاهده نمی شود.

نمودار ۱. شاخص خی دو و فاصله جفری مربوط به میزان واگرایی درس روش های نمونه گیری آزمون ارشد آمار



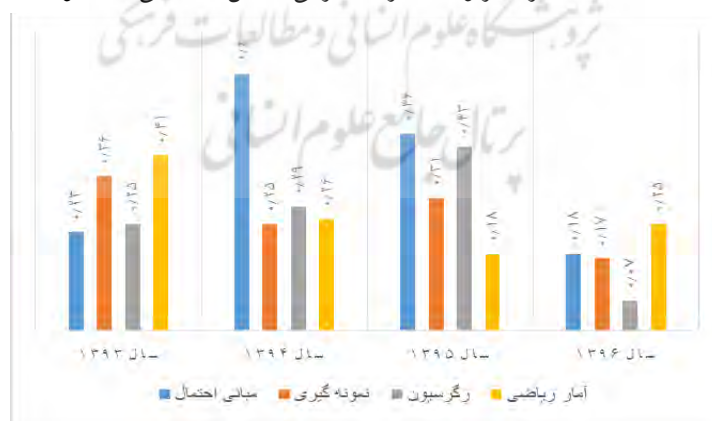
نمودار ۲. مقایسه فاصله جفری مربوط به میزان واگرایی دروس آمار و احتمالات مهندسی و طرح و تجزیه آزمایش‌ها در آزمون ارشد مهندسی کشاورزی



( $P < 0.05$ ) (\*: معنی دار در سطح ۰/۰۵)

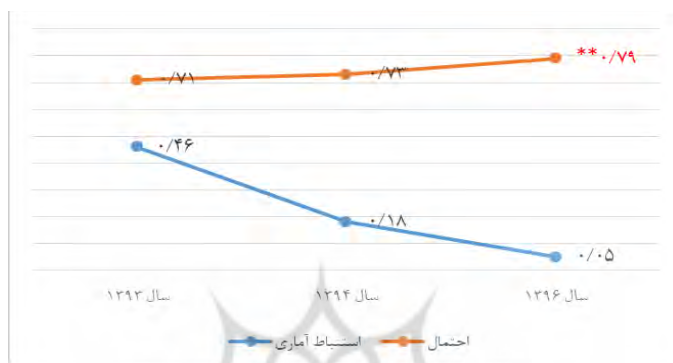
همانطور که مشاهده می‌شود نمودار ۲ نشان دهنده میزان واگرایی در دو درس از مواد آزمون کارشناسی ارشد مهندسی کشاورزی است. داده‌ها نشان می‌دهد با وجود روند نزولی این واگرایی در آزمون سال ۱۳۹۵ و در طراحی سوالات درس طرح آزمایشها شاهد یک واگرایی معنی دار در سطح ۰/۰۵ هستیم ( $P < 0.05$ ). علاوه بر مقادیر شاخص واگرایی جفری سطوح معنی داری براساس آزمون نیکویی برازش خنثی دو و با ادغام سطوح برای رسیدن به تعداد کافی سوالات در خوشه‌ها محاسبه شده است.

نمودار ۳. مقایسه فاصله جفری مربوط به میزان واگرایی دروس تخصصی در آزمون ارشد آمار



در آزمون کارشناسی ارشد آمار همانگونه که مشاهده می‌شود ضمن کاهش واگرایی طی سالیان اخیر هیچ یک از آنها براساس آزمون نیکویی برازش و در سطح ۰/۰۵ معنی دار نشده است. ( $P > 0/05$ )

نمودار ۴. مقایسه فاصله جفری مربوط به میزان واگرایی دروس تخصصی در آزمون دکتری آمار



\*\* معنی دار در سطح ۰/۰۱ ( $P < 0/01$ )

براساس نتایج مربوط به نمودار ۴ مشاهده می‌شود که برخلاف درس استنباط آماری که روند مناسبی داشته است ماده آزمون درس احتمال روند واگرایی افزایشی داشته به طوریکه در سال ۱۳۹۶ این واگرایی در سطح ۰/۰۱ معنی دار شده است. ( $P < 0/01$ )

### بحث و نتیجه گیری

همانطور که گفته شد در آزمون‌هایی که با هدف قضاوت در مورد عملکرد آموزشی افراد و یا ورود آنها به مقاطع بالاتر صورت می‌گیرد، تطابق بین برنامه درسی (سرفصل) و یا درس ارائه شده و آزمون و یا سنجش پیشرفت تحصیلی ضرورت دارد. در اکثر متون حوزه طراحی آزمون از جمله استانداردهای آزمون روانی تربیتی (۲۰۱۴) نیز این موضوع مورد تأکید قرار گرفته است. در پژوهش حاضر به کمک آزمون نیکویی برازش و همچنین شاخص جفری میزان واگرایی بین وضعیت موجود توزیع سوالات در بین مباحث ذیل سرفصل دروس و آنچه از دید صاحب‌نظران و طراحان سؤال‌های آزمون کارشناسی ارشد و دکتری وضعیت ایده‌آل بود محاسبه و آزمون شد. نتایج نشان داد در بعضی از دروس و برخی سال‌ها این واگرایی معنی بوده است. شاخص واگرایی جفری که حالت متقارن شاخص کولبک-لایبلر است، بیانگر میزان انحراف توزیع سوالات در دو حالت ایده‌آل و آنچه که در عمل رخ داده

است، می باشد. از آنجایی که مبانی محاسبه این شاخص به آنتروپی یا همان شاخص اطلاع شانون باز می گردد می توان گفت این شاخص معرف عدم انطباق اطلاعات در دو توزیع ایده آل و رویداد است. به همین جهت و البته از آنجایی که مانند حالت آزمون خی دو نیاز به ادغام خانه های با تعداد فراوانی کم نیست، این شاخص استفاده شد.

موضوع تناسب آزمون با محتوایی که آموزش داده می شود در اکثر مستندات حوزه سنجش مورد تأکید است و داویننگ<sup>۱</sup> در کتاب دستنامه طراحی آزمون‌های اثربخش (۲۰۰۶) تأکید می کند یکی از اولین پرسش‌هایی که در گام‌های ابتدایی طراحی آزمون باید پاسخ داده شود این است: چه محتوایی باید مورد آزمون قرار گیرد؟ در اولین گام طراحی یک آزمون اثربخش، شاید هیچ کاری مهم‌تر از مشخص کردن حوزه محتوایی برای آزمودن نباشد. اگر حوزه محتوایی خوب تعریف نشده باشد یا دقیق مشخص نشده باشد، نمی توان انتظار داشت سایر فعالیت‌های طراحی آزمون بتوانند این نقص را جبران کنند. اعتبار نتایج نمرات آزمون‌های پیشرفت تحصیلی به شدت وابسته به کفایت و قابل دفاع بودن روش‌های مورد استفاده در تعریف محتوا، آشکار بودن آن چه قرار است اندازه گرفته شود و استفاده از رویه‌های مناسب انتخاب نمونه‌های کافی و معرف از محتوا می باشد. استوبارت<sup>۲</sup> (۲۰۰۹) با تأکید بر موضوع روایی در سنجش و آزمون‌های ملی تأکید می کند در شرایطی که بین برنامه درسی ارائه شده و محتوای آزمون‌ها تطابق کم باشد روایی نیازمند بررسی و قضاوت خواهد بود. همه آزمون‌های پیشرفت تحصیلی و توانایی تأکید زیادی روی روایی محتوا برای تفسیر نمرات دارند (کان<sup>۳</sup>، ۲۰۰۶).

روش‌های تعریف محتوا از نظر دقت و سخت‌گیری با توجه به، هدف آزمون، عواقب تصمیم‌هایی که بر اساس نتایج آزمون اتخاذ می شود، و میزان قابل دفاع بودن هر تصمیمی که بر اساس نتایج آزمون اتخاذ می شود، بسیار متنوع هستند. برای آزمون‌های کمتر خطیر پیشرفت تحصیلی، روش‌های تعریف محتوا ممکن است بسیار ساده و روشن باشد، به عنوان مثال قضاوت غیررسمی (اما آگاهانه) مربیان در خصوص محتوای مناسب برای آزمون کفایت خواهد کرد. برای آزمون‌های خطیر<sup>۴</sup>، تعریف محتوا ممکن است با تحلیل وظیفه یا

---

1. Downing, S  
2. Stobart, G  
3. Kane, M  
4. High-stakes

شغل چند ساله شروع شود، میلیون‌ها دلار هزینه در بر داشته باشد، و مستلزم به خدمت گرفتن افراد حرفه‌ای باشد (داونینگ، ۲۰۰۶).

در نظام آموزش عالی کشورمان به دلیل اینکه نتایج آزمون‌های ملی اثر بسیار زیادی بر افراد دارد ضروری است تطابق بین سرفصل دروس و محتوای آزمون بسیار زیاد باشد. ذکر این نکته نیز ضروری است که عدم معرفی منابع آزمون در آزمون‌های کارشناسی ارشد و دکتری موضوعی قابل تأمل و توجه در آزمون‌های ورودی کارشناسی ارشد و دکتری است. با توجه به انتشار سوال‌های آزمون در هر دو آزمون کارشناسی ارشد و دکتری و رصد منابع آزمون‌ها توسط داوطلبان، واگرایی محتوای آزمون با آنچه در دانشگاهها تدریس شده در دراز مدت منجر به شکل‌گیری نوعی جوّی اعتمادی به محتوای سرفصل‌ها خواهد شد و دانشجویان به مطالعه منابعی که از آنها در آزمون‌های ورودی سوال طرح می‌شود گرایش پیدا می‌کنند. مشخص نبودن منابع خود به خود منجر به نوعی واگرایی در ارتباط بین سرفصل و محتوای آزمون و در نتیجه بی‌اثر شدن آزمون‌ها در کلاس درس خواهد شد. با توجه به یافته‌های این پژوهش ضرورت دارد موضوع میزان تطابق بین سرفصل درس و محتوای آزمون مورد رصد مداوم قرار گیرد. این امر منجر به ارتقای ویژگی‌های فنی آزمون‌های خواهد شد.

## منابع

- بهرام‌بیگی، مه‌ری (۱۳۹۱). ویژگی‌های مطلوب در تدوین سرفصل درسی دانشگاهی با نقد بر سرفصل‌های درسی رشته زبان فرانسه. پژوهش و نگارش کتب دانشگاهی، ۱۶(۲۵)، ۵۸-۴۸.
- حسن‌زاده، حسام (۱۳۹۸). بررسی تطبیقی سرفصل دروس دوره کارشناسی طراحی گرافیک در ایران و منتخبی از دانشگاه‌های جهان. نشریه مطالعات تطبیقی هنر، شماره ۹.
- دفترچه سوال‌های آزمون دکتری سال‌های ۱۳۹۳ تا ۱۳۹۶ رشته‌های آمار و مهندسی کشاورزی (منتشر شده در سایت سازمان سنجش آموزش کشور).
- دفترچه سوال‌های آزمون کارشناسی ارشد سال‌های ۱۳۹۳ تا ۱۳۹۶ رشته‌های آمار و مهندسی کشاورزی (منتشر شده در سایت سازمان سنجش آموزش کشور).
- رضائی، مژده (۱۳۹۵). مطالعه تطبیقی سرفصل دروس رشته‌های حوزه فناوری اطلاعات سلامت. در: هفتمین کنگره پژوهشی سالیانه دانشجویی، ۲۳ اردیبهشت ۱۳۹۵، اردبیل - ایران.
- سرفصل و برنامه آموزشی دوره کارشناسی رشته آمار (مصوب ۱۳۸۸/۰۲/۲۶).



سرفصل و برنامه آموزشی دوره کارشناسی ارشد رشته آمار ریاضی (مصوب ۱۳۷۴/۱۰/۱۰).  
شب‌خیز، فاطمه؛ جعفری حجین، افسر؛ امینیان رضوی، توراندخت؛ بلند، اذین (۱۳۹۱). بررسی  
سرفصل دروس و ارائه پیشنهادها برای اصلاحی دوره کارشناسی تربیت بدنی با گرایش  
مربیگری. مدیریت ورزشی، شماره ۱۲، صفحه ۱۶۲-۱۴۷.  
صادقی شاهدانی. مهدی، صاحب هنر. حامد، عظیم زاده آرانی. محمد، حسینی دولت آبادی. سید  
مهدی (۱۳۹۱)، بررسی اثر شوک های پولی بر متغیر های کلان اقتصادی با استفاده از  
روش BVAR: مطالعه موردی ایران، فصلنامه علمی-پژوهشی مطالعات اقتصادی در  
ایران، سال اول، شماره ۴، ص ۹۹-۱۰۰.  
علوی، سیدمحمد؛ کیوان پناه، شیوا و تاسی، یونس (۱۳۹۶). بررسی سرفصل دروس انگلیسی  
برای اهداف آکادمیک بر اساس نیازسنجی. IJEAP، دوره ۶، شماره ۱، صفحه ۱-۱۶.  
کرمی، مرتضی و فتاحی، هدی (۱۳۹۲). تغییر برنامه درسی آموزش عالی: (مورد: برنامه درسی  
دوره کارشناسی ارشد برنامه ریزی آموزشی). مطالعات برنامه درسی آموزش عالی، دوره  
۳، شماره ۷، بهار و تابستان ۱۳۹۲، صفحه ۱۱۱-۱۳۸.  
غریب پور، افرا و توتونچی مقدم، مارال (۱۳۹۴). بازنگری تطبیقی برنامه آموزشی پایه طراحی در  
دوره کارشناسی معماری. نشریه هنرهای زیبا، وره ۲۰، شماره ۴، صفحه ۷۲-۵۹.  
متوسلی. محمود، مزرعتی. محمد (۱۳۹۰)، پیش بینی و تحلیل سیاستی از تقاضای حامل های  
انرژی در ایران مدل های VAR, BVAR و پیشنهاد مدل های SBVAR، مجله برنامه  
و بودجه، شماره ۴۳ و ۴۴، ص ۲۹-۷۶.  
مهاجر ارومیه. الهه (۱۳۹۱)، برآورد پارامترهای یک جمعیت متناهی با استفاده از توزیع پسین  
پولیا، پایان نامه کارشناسی ارشد رشته آمار، دانشگاه شیراز، دی ماه ۱۳۹۱.  
یادگارزاده، غلامرضا (۱۳۹۹). بررسی و مطالعه اثر آزمون سراسری بر برنامه درسی دوره متوسطه.  
طرح پژوهشی آموزش و پرورش شهرستان‌های استان تهران.

## References

- Abrego, L. and Osterholm, P (2008), "External Linkages and Economic Growth in Colombia: Insights from A Bayesian VAR Model", IMF Working Paper WP/08/46.
- Amisano, G. and Serati, M (2004), " Time Varying Parameters BVAR Models for Inflation Forecasting", Research Unit of the Bank of Italy.
- Annala, J., Lindén, J., & Mäkinen, M. (2015). Curriculum in higher education research. In *Researching higher education* (pp. 171-189). Routledge.
- Biggs, J. (2003). Aligning teaching and assessment to curriculum objectives. *Imaginative Curriculum Project, LTSN Generic Centre, 12*.
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative research journal, 9*(2), 27-40.

- Borissov, B (1997), "BVAR Modeling in the Presence of Outliers", AMasters thesis, University of Toledo, Department of Economics.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3-12.
- Doan, T. Litterman, R and Sims, C (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions", *Econometric Reviews*, No. 3. PP: 1-100.
- Downing, S (2006). Twelve Steps for Effective Test Development in Downing & Haladyna (Eds.) *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Eignor, D. R. (2013). The standards for educational and psychological testing.
- Gower. B. (1997), "Scientific Method: An Historical & Philosophical Introduction.", university of Guelph. Routledge. December 12.
- Green. W.H. (1993), "*Econometric Analysis* (2nd ed)", Macmillan Publishing Company.
- Hargis, C. H. (2013). *Curriculum-based assessment: A primer*. Charles C Thomas Publisher.
- Husting, C., Harshman, J., & Yeziarski, E. (2017). Using teacher action research in high school chemistry to develop novel assessment tools. *J. Teach. Action Res*, 3(2), 14-29.
- Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007), 453-461.
- Jones, E. D., Southern, W. T., & Brigham, F. J. (1998). Curriculum-based assessment: Testing what is taught and teaching what is tested. *Intervention in School and Clinic*, 33(4), 239-249.
- Kane, Michael (2006), Content-Related Validity Evidence in Test Development, In Downing & Haladyna (Eds.) *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Kumar, P. M. (2016). Curriculum Enrichment for Learning Outcomes in Higher Education. *International Journal of Scientific Research and Modern Education (IJSRME) ISSN (Online)*, 2455-5630.
- Ord, J. (2004). Curriculum Debate: The Youth Work Curriculum as Process not as Outcome and Output to Aid Accountability. *Youth and Policy*, 85.
- Priestley, M., & Philippou, S. (2019). Curriculum is—or should be—at the heart of educational practice. *The Curriculum Journal*, 30:1, 1-7
- Stobart, G. (2009). Determining validity in national curriculum assessments. *Educational Research*, 51(2), 161-179.
- Wiliam, D. (2003). National curriculum assessment: how to make it better. *Research Papers in Education*, 18(2), 129-136.
- Wiliam, D. (2001). An overview of the relationship between assessment and the curriculum. *Curriculum and assessment*, 1, 165-181.

**استناد به این مقاله:** یادگارزاده، غلامرضا، فاطمی، عادل و رستگار آگاه، مصطفی. (۱۴۰۰). بررسی میزان واگرایی سرفصل دروس و سوال های آزمون با استفاده از آزمون نیکویی برازش و شاخص جفری (نمونه موردی: سوال های درس آمار آزمون تحصیلات تکمیلی ۱۳۹۷-۱۳۹۳). فصلنامه اندازه گیری تربیتی ۱۲(۴۵)، ۶۵-۸۲. doi: 10.22054/jem.2022.56477.2097



Educational Measurement is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.