

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*

Assistant Professor, Curriculum Studies Dept., Allame Tabatabai'i; university, Tehran, Iran

Adel Fatemi

Assistant Professor, Statistics Dept., Sanandaj Branch, Islamic Azad University, Sanandaj, Iran

Mostafa Rastgar Agah

Ph.D. in Assessment and Evaluation, Allame Tabatabai'i; university, Tehran, Iran

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), 63-80.
doi: 10.22054/jem.2022.56477.2097



اندازه‌گیری تربیتی

دوره ۱۲، شماره ۴۵، پاییز ۱۴۰۰، ۶۳-۸۰

jem.atu.ac.ir

DOI: 10.22054/jem.2022.56477.2097

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

غلامرضا یادگارزاده* | استادیار گروه مطالعات برنامه درسی، دانشگاه علامه طباطبائی، تهران، ایران

عادل فاطمی | استادیار گروه آمار، واحد سنندج، دانشگاه آزاد اسلامی، سنندج، ایران

مصطفی رستگار آگاه | دکتری سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران

چکیده

هدف از این پژوهش بررسی میزان واگرایی سؤال‌های آزمون کارشناسی ارشد و دکتری تعدادی از دروس اصلی آمار طی سال‌های ۱۳۹۷-۱۳۹۳ از سرفصل‌های مصوب شورای عالی برنامه‌ریزی بوده است. برای این منظور ابتدا بر اساس 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 سؤالات دفترچه آزمون‌های مذکور در بازه‌های زمانی مشخص در هر مبحث استخراج شد و فاصله این توزیع سؤالات 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 ($P=0/28$, $P=0/46$, $Chi-square=12.50$, $Value<0.01$). در نتیجه می‌توان گفت در میان دروس آمار در آزمون ورودی رشته‌های مهندسی کشاورزی و همچنین دروس تخصصی رشته آمار، بین سرفصل و سؤال‌های آزمون با وجود تغییرات، نسبتاً هماهنگی و تناسب رعایت شده است، به‌جز در دو ماده که میزان واگرایی و عدم تناسب معنی‌دار شده که این می‌تواند به دلیل عدم اطلاع طراحان سؤال از سرفصل‌های دروس مذکور باشد.

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

* 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 (Downing, 2006) لذا در فرایندهای سنجش و اندازه‌گیری و آزمون‌های ورودی بحث اطلاع‌طراحان آزمون‌ها از برنامه درسی و یا سرفصل دروس ضرورت دارد. برنامه‌های درسی^۱ به تعبیری قلب تمامی فعالیت‌های آموزشی هستند (Priestley & Philippou, 2019) و هر جا یادگیری در جریان است به‌نوعی برنامه‌های درسی حضور دارند. در تمامی نظام‌های آموزشی به‌منظور 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 (Biggs, 2003). در مقابل برخی معتقدند نباید آنچه در نظام‌های آموزشی می‌گذرد را با سنجش خروجی‌های مورد انتظار ارزشیابی و قضاوت کرد (Ord, 2004) زیرا نمی‌توان یادگیری را با یک آزمون و یا تکلیف مورد سنجش قرار داد. ورای این بحث وجدل 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 مورداستفاده قرار می‌گیرد آزمون‌های ورودی یا آزمون‌های خطیر و ملی است که از آن‌ها با عنوان غربالگر (Nagy, 2000) 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 داده شده سازگار باشد (Hargis, 2013). در استانداردهای آزمون‌روانی تربیتی^۲ تأکید شده است در سنجش آموخته‌ها باید آنچه مورد آزمون قرار می‌گیرد کاملاً مشخص و از پیش به اطلاع یادگیرندگان رسانده شود (Eignor, 2013). در واقع می‌توان گفت بین نتایج

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

2. Standards for Educational and Psychological Testing

یادگیری مورد انتظار و برنامه درسی باید تطابق وجود داشته باشد (Kumar, 2016) 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 (Nagy, 2000)، ترجمه یادگاززاده 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 تحت الشعاع قرار می‌دهد (Wiliam, 2001) و گاهی منجر به شکل‌گیری برنامه درسی پنهان می‌شود (یادگاززاده، 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 Annala و همکاران (2016) 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 کشورمان این امر می‌تواند توجه 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 بر اساس 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 سؤال اصلی مقاله حاضر بر میزان تطابق سؤال‌های آزمون‌های تحصیلات تکمیلی (کارشناسی ارشد و دکتری) با سرفصل دروس مصوب وزارت علوم، تحقیقات و فناوری است. روشن شدن ابعاد این سؤال 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 ثبت‌نام به اطلاع داوطلبان آزمون می‌رسد. همچنین در دفترچه ثبت‌نام داوطلبان به صورت روشن ذکر شده که سؤال‌های آزمون بر اساس سرفصل‌های مصوب شورای عالی برنامه‌ریزی طراحی خواهد شد (دفترچه ثبت‌نام آزمون تحصیلات تکمیلی سال ۱۳۹۸). در طراحی و تولید آزمون تحصیلات تکمیلی نیز معمولاً این موضوع به طراحان آزمون توصیه می‌شود و سؤال‌های خارج از سرفصل معمولاً غربال و اصلاح می‌شوند. بررسی سؤال‌های آزمون در رشته‌های مختلف نشان می‌دهد تطابق بین آزمون و سرفصل‌های دروس همیشه کامل نیست و معمولاً داوطلبان پس از آزمون و انتشار سؤال‌ها به آن اعتراض می‌کنند.

سرفصل‌های دروس در نظام آموزش عالی ما همواره موردنقد بوده‌اند و مسائلی مانند قدیمی بودن، عدم تطابق با نیازهای روز (کرمی و فتاحی، ۱۳۹۲)، عدم تطابق با تحولات علم و فناوری (علوی و همکاران، 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 بروندهای 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 خصوص 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 (Eignor, 2013؛ Jones et al., 1988؛ Husting et al., 2017). حتی برخی طرفداران این دیدگاه معتقدند کتاب یا منبع طرح سؤال نیز باید در اختیار آزمون‌شوندگان قرار داده شود. کسانی مانند Wiliam (2003)، Brookhart (2011) و فادیا ناصر (2007) معتقدند هماهنگی بین سنجش و برنامه درسی در سنجش‌های خطیر

ضرورت دارد و در سایر انواع سنجش که با هدف اندازه گیری میزان یادگیری انجام می شود خیلی ضرورت ندارد زیرا منجر به کاهش عملکرد فراگیران می شود و یادگیری طوطی وار را رواج می دهد.

در این میان موضوع مهم توجه به نوع نظام آموزشی و ساختار اجرای برنامه های درسی و سرفصل های دروس است که روی تطابق سنجش و برنامه درسی تأثیر می گذارد. در نظام آموزشی کشورمان به دلایلی این هماهنگی و تطابق باید فراهم باشد. این دلایل عبارت اند از سرنوشت ساز بودن نتیجه آزمون و اثر زیاد آن بر زندگی داوطلبان، تمرکز کامل نظام آموزش عالی و ضرورت اجرای هماهنگ سرفصل دروس در دانشگاه های 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 اعتراض های داوطلبان آزمون تحصیلات تکمیلی نشان می دهد نوعی واگرایی در سؤال های 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 برنامه های درسی به برخی دانشگاه ها و اختیار استادان در اجرای سرفصل دروس به صورت کاملاً میسر نمی شود. با توجه به آنچه گفته شد در پژوهش حاضر موضوع واگرایی سؤال های درس آمار از سرفصل دروس مصوب شورای عالی برنامه ریزی مورد مطالعه و بررسی قرار گرفته است.

روش

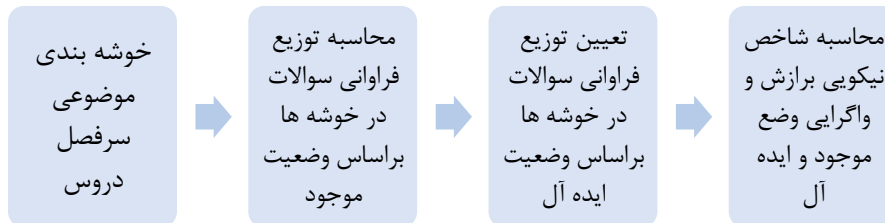
روش تحقیق مورداستفاده در این پژوهش تحلیل مستندات است. تحلیل مستندات به عنوان یک روش تحقیق کیفی عمدتاً با هدف کاوش مستندات منتشر شده (چاپی و الکترونیکی) و بررسی آن ها از ابعاد و زوایای مختلف است (Bowen, 2009). بر این اساس در این تحقیق ابتدا سرفصل دروس رشته کارشناسی و کارشناسی ارشد آمار (دروس آمار ریاضی، رگرسیون، نمونه گیری و مبانی احتمال در آزمون ارشد و دروس احتمال و استنباط آماری در آزمون دکتری) و همچنین دروس آمار و احتمالات آماری و طرح و تجزیه آزمایش های رشته مهندسی کشاورزی در سال های ۱۳۹۴ تا ۱۳۹۷ انتخاب و در گام نخست بر اساس موضوعات خوشه بندی شدند. پس از خوشه بندی و اختصاص کد به شاخه ها و زیرشاخه ها در گام دوم بر اساس آنچه در سنوات گذشته و در آزمون های برگزار شده توسط سازمان

سنجش آموزش کشور توزیع فراوانی سؤالات در شاخه‌ها و زیرشاخه‌ها محاسبه شد. در گام سوم و بر اساس نظر خبرگان حوزه آموزش که متولیان اصلی در تدوین سرفصل‌ها نیز هستند فراوانی سؤال‌ها به تفکیک زیر خوشه‌ها جهت پوشش صحیح سرفصل‌ها و البته بر اساس اهمیت موضوعات تحت عنوان توزیع فراوانی در وضعیت مورد انتظار و یا به عبارتی ایده‌آل (که توسط استادانی که دروس موردنظر 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 است وضعیت مطلوب نیز آمده است. جهت تطبیق و بررسی میزان واگرایی وضعیت موجود در سنوات گذشته و وضعیت ایده‌آل از آزمون نیکویی برازش بر اساس آماره خی‌دو و همچنین شاخص اطلاعاتی جفری که نسخه متقارن شاخص واگرایی کولبک-لایبلر^۱ است محاسبه شد. دلیل اینکه از آزمون خی‌دو و شاخص جفری استفاده شده این است که شاخص جفری میزان انحراف و یا واگرایی دو توزیع را بر اساس شاخص اطلاعاتی آنتروپی بسیار دقیق‌تر محاسبه می‌نماید، زیرا شاخص واگرایی اطلاعاتی جفری نشان‌دهنده میزان عدم انطباق اطلاعات موجود در دو توزیع است و از سویی حالت متقارن شاخص کولبک-لایبلر است. دلیل استفاده از شاخص خی‌دو نیز آزمون‌پذیری این شاخص به‌منظور تعیین معنی‌داری و عدم معنی‌داری میزان واگرایی است. به‌عبارت‌دیگر شاخص جفری میزان هدر رفت اطلاعات ناشی از پذیرش توزیع فراوانی سؤالات در حالت واقعی به‌جای جای توزیع مورد انتظار در سرفصل دروس مذکور و تدریس شده در دانشگاه‌ها است. البته این شاخص جایگزین آزمون خی‌دو نشده است، بلکه به‌عنوان مکمل در کنار آزمون خی‌دو آورده شده است؛ زیرا در روش خی‌دو به دلیل ادغام خانه‌ها با فراوانی کمتر از ۵ مورد عملاً شاخص قدرت تفکیک را از دست خواهد داد.

در شکل ۱ مراحل فوق به‌صورت خلاصه آمده است.

1. Kullback–Leibler divergence

شکل ۱. مراحل انجام تحقیق



از مزایای روش خرد دو که یک آزمون شناخته شده در آزمون نیکویی برازش است 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 سؤالات در برخی دروس و نیاز بیش از حد به ادغام خوشه هاست؛ بنابراین در بسیاری از موارد پژوهشگران ناچار به محاسبه شاخص واگرایی دو توزیع یعنی شاخص واگرایی (Jeffrey, 1946) شدند. این شاخص نسخه متقارن فاصله کولبک لایبلر است که بر اساس شاخص آنتروپی ساخته می شود و برای دو توزیع f و g برابر است با:

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

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 بی نظمی و یا به عبارتی میزان اطلاعات نهفته در توزیع احتمال f است. با توسعه مفاهیم آنتروپی شاخصی اطلاعاتی توأمان 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:

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

بررسی میزان واگرایی سرفصل دروس و سوال‌های ... | یادگاززاده و همکاران | ۷۱

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 بر اساس سنوات و دروس مقایسه شد.

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

نام درس	خوشه اصلی	زیر خوشه	مباحث
	مفاهیم پایه‌ای 120711	جامعه، نمونه، واحد نمونه‌گیری، چارچوب، سرشماری، مزایای نمونه‌گیری، مراحل نمونه‌گیری	۹۶
		با جایگذاری 1207121	نحوه و انتخاب نمونه محاسبه برآورد میانگین و واریانس و برآورد کننده آن ۹۳ ۹۴ ۹۶
روش‌های نمونه‌گیری 12071	نمونه‌گیری تصادفی ساده 120712	بدون جایگذاری 1207122	نحوه و انتخاب نمونه محاسبه برآورد میانگین و واریانس و برآورد کننده آن ۹۳ ۹۴ ۹۴ ۹۴ ۹۴ ۹۵ ۹۵ ۹۵ ۹۵ ۹۶ ۹۶ ۹۶
	برآورد نسبت‌ها 120713	محاسبه برآورد نسبت و میانگین و واریانس آن	۹۳ ۹۵

نام درس	خوشه اصلی	زیر خوشه	مباحث
120714	تعیین حجم نمونه	تعیین حجم نمونه به روش های فوق و روش نامساوی چپچف	۹۴ ۹۶
1207151	نمونه گیری طبقه بندی	روش نمونه گیری طبقه بندی	میانگین و واریانس برآورد کننده
120715	طبقه بندی		۹۳ ۹۳ ۹۳ ۹۳ ۹۴ ۹۴ ۹۴ ۹۵ ۹۵ ۹۵ ۹۵ ۹۵ ۹۶
1207152	مقایسه با روش تصادفی ساده	مقایسه با روش نمونه گیری تصادفی ساده	۹۳ ۹۴ ۹۵
120716	نمونه گیری با احتمال متغیر	کلیات و روش نمونه گیری با احتمال متغیر	۹۳

به عنوان نمونه در جدول ۱ خوشه بندی موضوعی و همچنین توزیع فراوانی سؤال های آزمون کارشناسی ارشد یکی از دروس رشته آمار، روش های نمونه گیری، به تفکیک چهار سال مشاهده می شود. همان طور که در جدول بالا مشاهده می شود برخی از موضوع های مورد تأکید در سرفصل (مانند روش نمونه گیری تصادفی بدون جایگذاری) دروس بیشتر مورد توجه طراحان آزمون بوده و برخی دیگر یا مورد نظر نبوده و یا فقط در یک سال (مانند نمونه گیری با احتمال متغیر) مورد آزمون قرار گرفته اند^۱. فعالیت مشابه برای ۷ درس آزمون کارشناسی ارشد و دکتری رشته آمار و همچنین دروس آمار آزمون کارشناسی ارشد رشته مهندسی کشاورزی انجام شد.

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

بررسی میزان واگرایی سرفصل دروس و سوال‌های ... | یادگاززاده و همکاران | ۷۳

جدول ۲. فراوانی مورد انتظار و فراوانی‌های مشاهده‌شده به تفکیک خوشه و یا زیر خوشه‌های درس روش‌های نمونه‌گیری آزمون کارشناسی ارشد رشته آمار

فراوانی مشاهده‌شده (وضعیت موجود)					
کد خوشه یا زیر خوشه	فراوانی مورد انتظار (وضعیت ایده آل)	سال ۱۳۹۳	سال ۱۳۹۴	سال ۱۳۹۵	سال ۱۳۹۶
۱۲۰۷۱۱	۱	۰	۰	۰	۱
۱۲۰۷۱۲۱	۲	۱	۱	۰	۱
۱۲۰۷۱۲۲	۲	۱	۴	۴	۳
۱۲۰۷۱۳	۱	۱	۰	۱	۰
۱۲۰۷۱۴	۱	۰	۱	۰	۱
۱۲۰۷۱۵۱	۲	۴	۳	۵	۲
۱۲۰۷۱۵۲	۱	۱	۱	۱	۰
۱۲۰۷۱۶	۰	۱	۰	۰	۰

با توجه به اطلاعات جداول فوق و به‌منظور بررسی انطباق فراوانی‌های مشاهده‌شده و فراوانی ایده‌آل و بررسی فرضیه واگرایی دو توزیع از آزمون نیکویی برازش شاخص‌خی‌دو استفاده گردید همچنین شاخص جفری نیز برای بررسی آنتروپی محاسبه شد.

جدول ۲. مقادیر آماره خی‌دو و شاخص جفری به همراه آزمون معنی‌داری واگرایی فراوانی سؤالات و وضعیت ایده‌آل درس روش‌های نمونه‌گیری آزمون کارشناسی ارشد رشته آمار

سال	۱۳۹۳	۱۳۹۴	۱۳۹۵	۱۳۹۶
شاخص خی‌دو	۱/۷۵	۰/۴۲	۱/۱۷	۰/۴۲
معنی‌داری	ns	ns	ns	ns
فاصله جفری	۰/۳۶	۰/۲۵	۰/۳۱	۰/۱۷

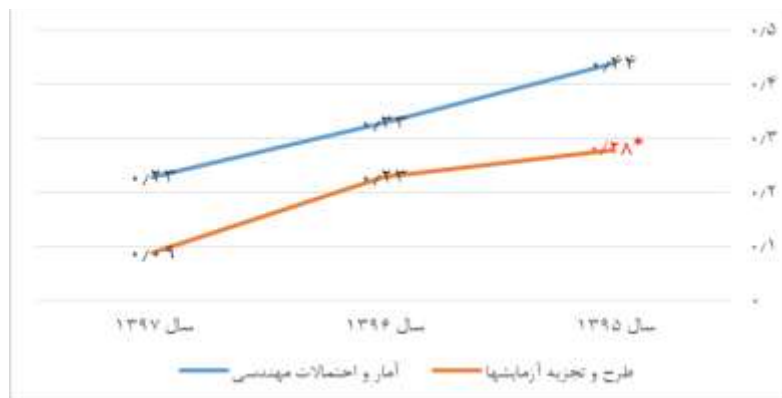
ns: عدم معنی‌داری ($P > 0.05$)

همان‌گونه که در جدول ۲ مشاهده می‌شود و بر اساس نتایج آزمون نیکویی برازش در تمام سنوات موردبررسی در سطح 0.05 تفاوت معنی‌داری بین وضعیت موجود و ایده‌آل مشاهده نمی‌شود.

نمودار ۱. شاخص خنی 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 آزمایش ها در آزمون ارشد مهندسی کشاورزی

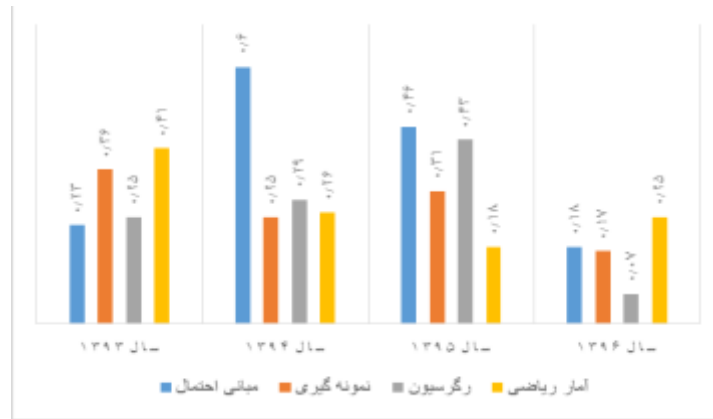


$(P < 0/05)$ *: معنی دار در سطح $0/05$

همان طور که مشاهده می شود نمودار ۲ نشان دهنده میزان واگرایی در دو درس از مواد آزمون کارشناسی ارشد مهندسی کشاورزی است. داده ها نشان می دهد با وجود روند نزولی این واگرایی در آزمون سال ۱۳۹۵ و در طراحی سؤالات درس طرح آزمایش ها شاهد یک واگرایی معنی دار در سطح $0/05$ هستیم ($P < 0/05$). علاوه بر مقادیر شاخص واگرایی جفری سطوح معنی داری بر اساس آزمون نیکویی برازش خنی دو و با ادغام سطوح برای رسیدن به تعداد کافی سؤالات در خوشه ها محاسبه شده است.

بررسی میزان واگرایی سرفصل دروس و سوال‌های ... | یادگارزاده و همکاران | ۷۵

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 هیچ‌یک از آن‌ها بر اساس آزمون نیکویی برازش و در سطح $0/05$ معنی‌دار نشده است. ($P > 0/05$)

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



** معنی‌دار در سطح $0/01$ ($P < 0/01$)

بر اساس 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 طوری که در سال ۱۳۹۶ این واگرایی در سطح $0/01$ معنی‌دار 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 (2014) 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 سؤالات 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 سؤالات در دو حالت ایده‌آل و آنچه در عمل رخ داده، است. از آنجایی که مبانی محاسبه این شاخص به آنتروپی یا همان شاخص اطلاع شانون بازمی‌گردد می‌توان گفت این شاخص معرف عدم انطباق اطلاعات در دو توزیع ایده‌آل و رویداد است. به همین جهت و البته از آنجایی که مانند حالت آزمون‌خی دو نیاز به ادغام خانه‌های با تعداد فراوانی کم نیست، این شاخص استفاده شد.

موضوع تناسب آزمون با محتوایی که آموزش داده می‌شود در اکثر مستندات حوزه سنجش مورد تأکید است و Downing در کتاب دست‌نامه طراحی آزمون‌های اثربخش (2006) تأکید می‌کند 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 آزمون نباشد. اگر حوزه محتوایی خوب تعریف نشده باشد یا دقیق مشخص نشده باشد، نمی‌توان 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 آنچه قرار است اندازه گرفته شود و استفاده از رویه‌های مناسب انتخاب نمونه‌های کافی و معرف از محتوا است. Stobart (2009) با تأکید بر موضوع روایی در سنجش و آزمون‌های ملی تأکید می‌کند در شرایطی که بین برنامه درسی ارائه شده و محتوای آزمون‌ها تطابق کم باشد روایی نیازمند بررسی و قضاوت خواهد

بود. همهٔ آزمون‌های پیشرفت تحصیلی و توانایی تأکید زیادی روی روایی محتوا برای تفسیر نمرات دارند (Kane, 2006).

روش‌های تعریف محتوا از نظر دقت و سخت‌گیری با توجه به، هدف آزمون، عواقب تصمیم‌هایی که بر اساس نتایج آزمون اتخاذ می‌شود و میزان قابل دفاع بودن هر تصمیمی که بر اساس نتایج آزمون اتخاذ می‌شود، بسیار متنوع هستند. برای آزمون‌های کمتر خطیر پیشرفت تحصیلی، روش‌های تعریف محتوا ممکن است بسیار ساده و روشن باشد، به‌عنوان مثال قضاوت غیررسمی (اما آگاهانه) مریان در خصوص محتوای مناسب برای آزمون کفایت خواهد کرد. برای آزمون‌های خطیر^۱، تعریف محتوا ممکن است با تحلیل وظیفه یا شغل چندساله شروع شود، میلیون‌ها دلار هزینه در بر داشته باشد و مستلزم به خدمت گرفتن افراد حرفه‌ای باشد (Downing, 2006).

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 سؤال‌های 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 درازمدت منجر به شکل‌گیری نوعی جوّی اعتمادی به محتوای سرفصل‌ها خواهد شد و دانشجویان به مطالعه منابعی که از آن‌ها در آزمون‌های ورودی سؤال طرح می‌شود گرایش پیدا می‌کنند. مشخص نبودن منابع خودبه‌خود 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)، ۴۸-۵۸.

- حسن‌زاده، حسام. (۱۳۹۸). بررسی تطبیقی سرفصل دروس دوره کارشناسی طراحی گرافیک در ایران و منتخبی از دانشگاه‌های جهان. *نشریه مطالعات تطبیقی هنر*، ۵(۹)، ۳۵-۴۵.
- دفترچه سؤال‌های آزمون دکتری سال‌های ۱۳۹۳ تا ۱۳۹۶ رشته‌های آمار و مهندسی کشاورزی (منتشرشده در سایت سازمان سنجش آموزش کشور).
- دفترچه سؤال‌های آزمون کارشناسی ارشد سال‌های ۱۳۹۳ تا ۱۳۹۶ رشته‌های آمار و مهندسی کشاورزی (منتشرشده در سایت سازمان سنجش آموزش کشور).
- رضایی، مژده. (۱۳۹۵). مطالعه تطبیقی سرفصل دروس رشته‌های حوزه فناوری اطلاعات سلامت. *در هفتمین کنگره پژوهشی سالیانه دانشجوئی*، ۲۳، اردیبهشت ۱۳۹۵، اردیبل - 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، امینیان رضوی، افسر و توراندخت بلند، آذین. (۱۳۹۱). بررسی سرفصل دروس و ارائه پیشنهادها اصلاحی دوره کارشناسی تربیت‌بدنی با گرایش مربیگری. *مدیریت ورزشی*، ۱۲(۱)، ۱۴۷-۱۶۲.
- صادقی شاهدانی، مهدی، صاحب‌هنر، حامد، عظیم زاده آرانی، محمد و حسینی دولت‌آبادی، سید مهدی. (۱۳۹۱). بررسی اثر شوک‌های پولی بر متغیرهای کلان اقتصادی با استفاده از روش BVAR: مطالعه موردی ایران. *فصلنامه علمی-پژوهشی مطالعات اقتصادی در ایران*، ۱(۴)، ۹۹-۱۰۰.
- علوی، سید محمد، کیوان پناه، شیوا و تاسی، یونس. (۱۳۹۶). بررسی سرفصل دروس انگلیسی برای اهداف آکادمیک بر اساس نیازسنجی. *JJEAP*، ۱(۶)، ۱-۱۶.
- کرمی، مرتضی و فتاحی، هدی. (۱۳۹۲). تغییر برنامه درسی آموزش عالی، (مورد: برنامه درسی دوره کارشناسی ارشد برنامه‌ریزی آموزشی). *مطالعات برنامه درسی آموزش عالی*، ۳(۷)، ۱۱۱-۱۳۸.
- غریب‌پور، افرا و توتونچی مقدم، مارال. (۱۳۹۴). بازنگری تطبیقی برنامه آموزشی پایه طراحی در دوره کارشناسی معماری. *نشریه هنرهای زیبا*، ۲۰(۴)، ۵۹-۷۲.
- متوسلی محمود و مزرعتی، محمد. (۱۳۹۰). پیش‌بینی و تحلیل سیاستی از تقاضای حامل‌های انرژی در ایران مدل‌های VAR, BVAR و پیشنهاد مدل‌های SBVAR. *مجله برنامه‌بودجه*، ۴۳(۴۴)، ۲۹-۷۶.

مهاجر ارومیه. الهه. (۱۳۹۱)، برآورد پارامترهای یک جمعیت متناهی با استفاده از توزیع پسین پولیا، پایان‌نامه کارشناسی ارشد، رشته آمار، دانشگاه شیراز، دی‌ماه ۱۳۹۱.

یادگازاده، 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., & Yezierski, 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)، ۶۳-۸۰. doi: 10.22054/jem.2022.56477.2097



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