

لزوم سنجش و گزارش شاخص‌های اندازه اثر در پژوهش‌های روان‌شناختی

Necessity of Measuring and Reporting Effect Size Measures in Psychological Studies

دکتر سعید اکبری زردخانه

گروه روانشناسی، دانشگاه شهید بهشتی، تهران، ایران

محمد عطاری (نویسنده مسئول)

گروه روانشناسی، دانشگاه تهران، تهران، ایران . Mohammad.attari@yahoo.com

Dr. Saeed Akbari Zardkhane

Department of Psychology, Shahid Beheshti University, Tehran, Iran

Mohammad Atari (corresponding author)

Department of Psychology, University of Tehran, Tehran, Iran

Mohammad.attari@yahoo.com

Abstract

Dissatisfaction with tests of statistical significance in last decades has led to emergence of effect size measures. Effect size is generally a measure, incorporated into statistical analyses in order to emphasize the size of the difference rather than sample size. Assessing and reporting measures of effect size are of absolute importance in behavioral sciences and American Psychological Association encourages researchers to report them. This paper investigates the concept and pertaining calculations of effect size measures. Moreover, the necessity of reporting measures of effect size in research articles has been elaborated. A number of alternative measures of effect size were described. Finally, advice on the use of effect sizes in meta-analysis is summarized.

Keywords: effect size, sample size, statistical significance, meta-analysis, quantitative analysis, statistical testing.

چکیده

عدم رضایت پژوهشگران از آزمون‌های آماری در دهه‌های گذشته موجب به وجود آمدن شاخص‌های آماری با نام اندازه اثر گشته است. اندازه اثر شاخصی برای از بین بردن تأثیر حجم نمونه در سنجش معناداری آماری در پژوهش‌های کمی است. سنجش شاخص‌های اندازه اثر و گزارش آن‌ها در پژوهش‌های علوم رفتاری از اهمیت زیادی برخوردار بوده و انجمن روانشناسی آمریکا آن را توصیه می‌کند. این مقاله به صورت مقدماتی به بررسی مفهومی و محاسباتی شاخص‌های مختلف در سنجش اندازه اثر پرداخته و تعدادی از شاخص‌های اندازه اثر را توصیف می‌کند و همچنین لزوم گزارش آن‌ها را در مقالات پژوهشی در حیطه روانشناسی به منظور کاربرد در فراتحلیل شرح می‌دهد.

واژه‌های کلیدی: اندازه اثر، حجم نمونه، معناداری آماری، فراتحلیل، پژوهش کمی، آزمون آماری.

مقدمه

اکثر تحلیل‌های آماری مشهور (مانند تی استیودنت، ضریب همبستگی پیرسون، ضریب همبستگی اسپیرمن) و به دنبال آن‌ها، آزمون‌های معناداری این تحلیل‌ها حدود یک قرن پیش به وجود آمده‌اند. با این وجود نقد بر آزمون‌های معناداری این تحلیل‌ها تقریباً به اندازه خود تحلیل‌ها قدمت دارند. در دهه‌های اخیر بر تعداد این انتقادات افزوده شده است. اشمیت و هانتز^۱ (۱۹۹۷) معتقدند که آزمون‌های معناداری آماری رشد علمی را به تعویق می‌اندازد و هیچ‌گاه کمکی به این رشد علمی نمی‌کند. البته برخی طرفداران این حوزه پاسخ‌هایی را به چنین انتقادهایی داده‌اند (مانند رابینسون و واینر^۲، ۲۰۰۲).

انتقادات بر این حوزه پژوهشگران را به این سوال واداشته است که اگر آزمون‌های معناداری آماری نمی‌توانند کارایی مناسبی در پژوهش‌های کمی داشته باشند، پس چه چیزی می‌تواند جای آن‌ها را بگیرد؟ در دو دهه گذشته علوم اجتماعی و رفتاری به استفاده از معناداری کاربردی^۳ روی آورده‌اند. شروع این جنبش علمی شدیداً تحت تأثیر انتشار مقاله‌ای از طرف انجمن روانشناسی آمریکا در سال‌های ۱۹۹۶ و ۱۹۹۹ (ویلکینسن و انجمن روانشناسی آمریکا^۴، ۱۹۹۹) بوده است. در این مقالات دو گونه از شواهد معناداری کاربردی در این مقالات مورد بحث قرار گرفته است: اندازه اثر و فاصله اطمینان. در همین مقاله آمده است که: «گزارش و تحلیل اندازه اثر در کنار سایر شاخص‌های آماری برای یک پژوهش خوب لازم است».

با توجه به سیر تاریخی توضیح داده شده، به پژوهشگران حوزه‌های رفتاری پیشنهاد می‌کنند که علاوه بر آزمون معناداری آماری از شاخص‌های اندازه اثر نیز بهره گیرند (کوهن^۵، ۱۹۶۵). دلیل اصلی این پیشنهاد این است که آزمون‌های معناداری

^۱ -Schmidt & Hunter

^۲ -Robinson & Wainer

^۳ -Practical significance

^۴ -Wilkinson & APA task force on statistical inference

^۵ -Cohen

آماري به تنهایی «معناداری»^۱ با فحوا و مفهومی را تضمین نمی‌کنند، این آزمون‌ها تنها می‌توانند حاکی از این باشند که ارتباط مشاهده شده در یافته‌ها منتج از شانس نیست. هر چند آزمون‌های معناداری آماری باعث می‌شوند که پژوهشگر اختلاف‌های به ظاهر بزرگ بین جوامع آماری کوچک را به اشتباه معنادار تفسیر نکند ولی این آزمون‌ها از معنادار در نظر گرفتن اختلاف‌های بسیار کوچک در جوامع بزرگ نمی‌توانند جلوگیری کنند (کارور^۲، ۱۹۹۳). در جامعه‌های بزرگ اختلاف‌های کوچک می‌توانند از نظر آماری معنادار به دست آیند. این مورد برخی را بر این داشته که آزمون‌های آماری را به تنهایی مفید ندانند (کیرک^۳، ۱۹۹۶). این منتقدان معتقدند که بازه‌های معناداری^۴ و شاخص‌های اندازه اثر می‌بایست در تحلیل یافته‌ها مورد توجه ویژه قرار گیرند (تامپسون^۵، ۱۹۹۶). یکی از افرادی که از کاربرد و گزارش شاخص‌های اندازه اثر حمایت کامل می‌کند، تامپسون است (اولجنیک و آلگینا^۶، ۲۰۰۰). تامپسون یکی از منتقدان جمله بندی کتاب‌نامه جامع انجمن روانشناسی آمریکا در این رابطه است. در این کتاب‌نامه پژوهشگران به گزارش شاخص‌های اندازه اثر تشویق^۷ شده‌اند (انجمن روانشناسی آمریکا^۸، ۱۹۹۴). حال آن که وی معتقد است تمامی پژوهشگران می‌بایست^۹ این شاخص‌ها را گزارش کنند (اولجنیک و آلگینا، ۲۰۰۰). انجمن روانشناسی آمریکا در سال ۲۰۰۱ به پژوهشگران توصیه می‌کند که آن‌ها نباید فقط به خوانندگان اطلاعاتی در مورد معناداری آماری ارائه دهند، بلکه باید علاوه بر این، اطلاعاتی نیز در مورد اندازه اثر مشاهده شده ارائه نمایند (انجمن روانشناسی آمریکا، ۲۰۰۱).

^۱ -Significance

^۲ -Carver

^۳ -Kirk

^۴ -Confidence Intervals (CI)

^۵ -Thompson

^۶ -Olejnik & Algina

^۷ - Encouraged

^۸ -American Psychological Association

^۹ - Required

در سال‌های گذشته منابع مختلف در باب تاریخچه توسعه شاخص‌های اندازه اثر مقالات و کتاب‌هایی نوشته اند (مانند داویر، ۱۹۷۴؛ ریچاردسون، ۱۹۹۶) و برخی نیز شاخص‌های جدیدی ارائه کرده‌اند (مانند روزنتال و روبین^۱، ۱۹۸۲) و برخی نیز در باب چگونگی تجزیه و تحلیل این شاخص‌ها مطالبی منتشر کرده‌اند (مانند میچل و هارتمن^۲، ۱۹۸۱).

همچنین، هیئت‌های تحریریه مجلات معتبر جهان در حیطه‌های روان‌شناختی، نویسندگان مقالات را تشویق به استفاده از تکنیک‌های ساده، انعطاف پذیر و گرافیکی برای توضیح یافته‌های پژوهش می‌کنند (کوهن، ۱۹۹۴). معرفی مفهومی شاخص‌های اندازه اثر یکی از راه‌های ساده و مناسب برای تفسیر اطلاعات آماری مقالات است. هدف این مقاله بررسی مفهومی آزمون‌های معناداری آماری و سپس معرفی اندازه اثر به عنوان شاخصی برای از میان برداشتن کاستی‌های گزارش آزمون‌های آماری به تنهایی است. چند شاخص مشهورتر از شاخص‌های اندازه اثر نیز معرفی شده‌اند و کاربرد آن‌ها در یک مثال فرضی نشان داده شده است. همچنین افزایش تعداد پژوهشگرانی که این شاخص‌ها را گزارش می‌کنند، نیز بررسی شده است.

معناداری آماری

به طور کلی در پژوهش‌های علوم رفتاری^۳، برای به دست آوردن سطح معناداری آماری^۴ پژوهشگر باید به این سوال پاسخ دهد که آیا آماره^۵ آزمون آماری استفاده شده شده از حد معینی بیشتر است یا خیر. در روانشناسی معمولاً این حد مشخص را ۰,۰۵ یا ۰,۰۱ در نظر می‌گیرند (فرگوسن و تاکانه، ۱۳۹۲). این شرط برای کاهش احتمال ارتکاب خطای نوع اول^۶ است. در روش شناسی، خطای نوع اول، به معنای یافتن

^۱ - Rosenthal & Rubin

^۲ - Mitchell & Hartmann

^۳ - Behavioral sciences

^۴ - Statistical significance

^۵ - Statistic

^۶ - Type I error

مقادیر معنادار برای یک آزمون آماری است وقتی که در حقیقت چنین رابطه‌ای وجود نداشته باشد. این شرط ($P < 0.05$) پژوهشگر را در برابر ارتکاب به خطای نوع دوم^۱ حفاظت نمی‌کند. خطای نوع دوم، به معنای عدم یافتن رابطه‌ای معنادار است هنگامی که در حقیقت رابطه‌ای معنادار وجود داشته باشد. خطای نوع دوم به بحث توان آزمون آماری مربوط است (فراهانی و همکاران، ۱۳۸۷). معناداری آماری تابعی از عوامل مختلف است از جمله: بزرگی اثر^۲ (اندازه اختلاف موجود)، بزرگی حجم نمونه^۳، پایایی^۴ تأثیر (آیا مداخله برای همه آزمودنی‌ها به یک اندازه تأثیر گذار است؟) و پایایی پایایی ابزار سنجش (بکر^۵، ۲۰۰۰). برخی پژوهشگران معناداری آماری را به معنای پایان فرآیند آماری پژوهش در نظر می‌گیرند. در واقع آن‌ها تنها به در نظر گرفتن ارزش P^6 بسنده می‌کنند. حال آنکه مقادیر کوچک معناداری (مانند $P < 0.001$) می‌تواند می‌تواند تنها به دلیل بزرگی حجم نمونه یا پایایی خیلی بالای ابزار سنجش و یا ترکیبی از آن‌ها باشد (اولجنیک، ۲۰۰۰). به عنوان مثال همبستگی‌های بسیار کوچک در حجم‌های بزرگ نمونه می‌توانند دارای معناداری آماری باشند. این مسئله در جداول ارزش‌های بحرانی آماره‌های مختلف نیز تایید می‌گردد (فرگوسن و تاکانه، ۱۳۹۲). در بسیاری پژوهش‌ها صرفاً تعداد زیاد شرکت کنندگان سطح معناداری آماری مناسبی را فراهم کرده و پژوهشگر را قانع می‌سازد. در حالی که هر پژوهشگر خبره‌ای به صورت شهودی می‌داند که اندازه اختلافی که یک مداخله می‌تواند ایجاد کند دارای اهمیت اساسی است و نه صرفاً حجم نمونه یا پایایی ابزار سنجش. پیشگامان روش فراتحلیل^۷ همواره به دنبال تعیین اندازه اثر یک مداخله بوده‌اند. ولی معناداری آماری که خود آلوده به حداقل چهار عامل ذکر شده است، نمی‌تواند شاخص مناسبی برای عامل اول

^۱ -Type II error

^۲ - Effect magnitude

^۳ - Sample size

^۴ -Reliability

^۵ -Becker

^۶ -P-value

^۷ -Meta-analysis

(بزرگی اثر) به تنهایی باشد. لیکن این پیشگامان شاخص‌هایی را پیشنهاد داده‌اند که تنها به بررسی اندازه اثر پرداخته و در واقع تأثیر عوامل دیگر را خنثی می‌کند. لازم به ذکر است که سطح معناداری و اندازه اثر هر کدام اطلاعات جداگانه‌ای به پژوهشگر می‌دهند. با استفاده از آزمون معناداری تعیین می‌گردد که آیا نتیجه خاصی به علت شانس رخ داده است یا خیر؛ در حالی که از طریق بررسی اندازه اثر می‌توان متوجه شد که متغیر مستقل تا چه اندازه بر متغیر وابسته اثر گذاشته است (بکر، ۲۰۰۰).

اندازه اثر

گاهی به دلیل در دسترس نبودن نمونه‌های بزرگ یا سختی در فرآیند نمونه‌گیری یا مواردی از این قبیل، معناداری آماری برای قبول یک فرضیه آماری حاصل نمی‌شود. با این حال، پژوهشگر شهود^۱ حدس می‌زند که اختلاف معناداری وجود داشته که آزمون آماری^۲ به تنهایی نتوانسته آن را نشان دهد. وجود چنین مواردی در علوم رفتاری، باعث شده تا شاخص‌هایی تنها برای اندازه‌گیری اندازه اثر معرفی شوند (بکر، ۲۰۰۰). به طور مثال فرض کنید جدول زیر نشان دهنده نمرات دو گروه مختلف در سیاهه افسردگی بک^۳ است. این سیاهه در سال ۱۹۶۲ توسط بک ساخته شده و دارای ۲۱ گویه است. هر ماده بیان‌کننده حالتی در فرد است و علایم افسردگی را مشخص می‌کند. معمولاً نمرات بالاتر از ۱۸ به عنوان نشانه افسردگی تلقی می‌شود (منصور، ۱۳۷۶).

^۱ -Intuitively

^۲ -Statistical test

^۳ - Beck Depression Inventory (BDI)

جدول ۱. نمرات دو گروه فرضی در سیاهه افسردگی بک

نمرات گروه ۱	نمرات گروه ۲
۳	۵
۵	۷
۷	۱۱
۸	۹
۶	۷
۴	۸

آماره t برای گروه‌های مستقل در این دو گروه $t = ۲$ بوده و در سطح ۹۵٪ اطمینان از نظر آماری معنادار نیست. با این وجود از نظر بالینی^۱ تفاوت نسبتاً واضحی را می‌توان مشاهده کرد. در چنین شرایطی (یا حالتی که تفاوت آشکاری دیده نمی‌شود ولی تعدد شرکت کنندگان باعث معناداری آماری گشته است) لزوم استفاده از شاخص‌های اندازه اثر معنایی روشن‌تر پیدا می‌کند. امروزه درصد نویسندگان ایرانی که تمایل به گزارش شاخص‌های اندازه اثر در پژوهش‌ها دارند، رو به افزایش است.

شاخص‌های متفاوتی برای اندازه گیری اندازه اثر در آزمون‌های متفاوت وجود دارد. در سال ۱۹۹۶ کیرک ۴۱ نوع شاخص اندازه اثر را شناسایی کرد (کیرک، ۱۹۹۶). با توجه به افزایش توجه پژوهشگران نسبت به شاخص‌های اندازه اثر، این تعداد در سال‌های آتی بدون شک افزایش خواهد یافت. جدول (۱) تنها چند شاخص پرکاربردتر و مشهورتر را برای اندازه گیری اندازه اثر نشان می‌دهد.

شاخص d برای محاسبه اندازه اثر بین دو گروه مورد استفاده قرار می‌گیرد. این شاخص عبارت است از حاصل تقسیم اختلاف میانگین دو گروه بر انحراف معیار کل^۲ دو گروه. این شاخص برای اولین بار توسط کوهن معرفی شده و جز مشهورترین

^۱ - Clinical^۲ - Pooled

شاخص‌های اندازه اثر به شمار می‌آید. شاخص f شبیه به d بوده با این تفاوت که برای ارزیابی اندازه اثر بین میانگین‌های بیش از دو گروه مورد استفاده قرار می‌گیرد. در تحلیل واریانس یکراهه می‌توان این شاخص را گزارش کرد. شاخص η^2 (مجذور ای‌تا) نیز برای تحلیل واریانس یکراهه مورد استفاده است و از نظر تاریخی قدیمی‌تر از f است. این دو شاخص را می‌توان به سادگی از طریق فرمول (۱) به یکدیگر تبدیل کرد. شاخص r همان ضریب همبستگی است. در واقع می‌توان این طور گفت که در آزمون‌های معناداری همبستگی می‌توان از خود r به عنوان شاخصی از اندازه اثر استفاده کرد. استفاده از r معمولاً ساده است. برای آزمون خی دو می‌توان از شاخص اندازه اثر امگا (ω) بهره برد. گلاس نیز شاخص دلتا (Δ) را برای مقایسه میانگین گروه آزمایشی و گروه کنترل معرفی کرد. دلتا از حاصل تقسیم اختلاف میانگین گروه آزمایشی و کنترل بر انحراف استاندارد گروه کنترل به دست می‌آید. البته از این شاخص برای گروه‌های مستقل نیز استفاده شده است. ارزیابی مقدی هر یک از این شاخص‌ها نیز در جدول (۱) آمده است.

فرمول (۱)

$$f = \sqrt{\frac{\eta^2}{1 - \eta^2}}$$

جدول ۲. انواع مشهورتر شاخص‌های اندازه اثر به همراه شرایط استفاده، فرمول و ارزیابی آن‌ها

شاخص	شرایط استفاده	فرمول	ارزیابی
d	اختلاف دو میانگین		کوچک ۰,۲
			متوسط ۰,۵
			بزرگ ۰,۸
f	تحلیل واریانس یک راهه		کوچک ۰,۱
			متوسط ۰,۲۵
			بزرگ ۰,۴
R	تحلیل واریانس یک راهه		کوچک ۰,۰۱
			متوسط ۰,۰۶
			بزرگ ۰,۱۴
	همبستگی گشتاوری		کوچک ۰,۱
			متوسط ۰,۳
			بزرگ ۰,۵
ω	آزمون خی-دو		کوچک ۰,۱
			متوسط ۰,۳
			بزرگ ۰,۵
Δ	اختلاف دو میانگین		کوچک ۰,۲
			متوسط ۰,۵
			بزرگ ۰,۸

بدین ترتیب برای مثال ذکر شده می‌توان از شاخص d^1 برای ارزیابی مقدار اندازه اثر بهره برد (کولمن^۱، ۲۰۰۹):

^۱ - Cohen's d

$$d = \frac{M_1 - M_2}{\sigma_p} \rightarrow d = \frac{2.04 - 1.87}{2.23} \rightarrow d = 1.04$$

این اندازه اثر بزرگ‌تر از ۰٫۸ بوده و طبق جدول ۲ حاکی از تفاوت قابل توجه بین دو گروه است. این تفاوت با استفاده از آماره t مقداری غیر معنادار به دست داده بود. با نگاهی به مقدار اندازه اثر به دست آمده مشخص می‌گردد که عدم معناداری آماری تنها به دلیل کوچک بودن نمونه‌ها (۱۲ نفر) است. در واقع اندازه اثر تفاوت M_1 و M_2 را بر حسب انحراف استاندارد جامعه به ما می‌دهد؛ به عبارت دیگر تعداد واحدهای انحراف معیاری را که دو میانگین جامعه با هم تفاوت دارند نشان می‌دهد (سرمد، ۱۳۸۴). هرچه اندازه اثر بزرگ‌تر باشد، درجه حضور پدیده بیشتر است. روزنتال معتقد است که اگر اندازه اثر محاسبه نشود، احتمال بروز خطای نوع اول افزایش می‌یابد (روزنتال، ۱۹۹۱).

باید توجه داشت که سطح معناداری آماری و اندازه اثر اطلاعات جداگانه‌ای به پژوهشگر می‌دهند. گزارش تنها یکی از آن‌ها می‌تواند منبعی برای خطا باشد. از طریق اندازه اثر متوجه می‌شویم که متغیر مستقل تا چه اندازه بر متغیر وابسته اثر گذاشته است (فراهانی و عریضی، ۱۳۸۴). تامپسون (۲۰۰۱) معتقد است که اگر در تعبیر اندازه اثر به گزارش اثرهای «کوچک»، «متوسط» و «بزرگ» اکتفا کنیم (مانند گزارش $P < 0.05$ در گزارش‌های معناداری آماری)، دچار اشتباهی بزرگ شده‌ایم. در واقع این اهمیت موضوع پژوهش و ابعاد متفاوت آن است که تعبیر نتایج اندازه اثر را مشخص می‌کند. به طور مثال اگر پژوهشی در مورد جان انسان‌ها باشد، مقادیر کوچک اندازه اثر می‌تواند معانی متفاوتی داشته باشند و تعبیر درست آن‌ها با سایر پژوهش‌ها متفاوت است.

روزنتال و دما تو شاخص‌های مختلف اندازه اثر را به دو خانواده اصلی تقسیم کرده‌اند: شاخص‌های خانواده d و شاخص‌های خانواده r . اعضای خانواده d عمدتاً شاخص‌هایی برای اندازه گیری تفاوت‌های استاندارد بین میانگین‌ها است و

شاخص‌های خانواده r برآوردهایی از گشتاور حاصل ضربی همبستگی پیرسون را نشان می‌دهد (روزنتال و دماتو^۱، ۲۰۰۱). هر دو برآورد d و r را می‌توان به یکدیگر تبدیل کرد و در نهایت فراتحلیل گران باید تصمیم بگیرند که از کدام شاخص اندازه اثر بهره ببرند. برخی نویسندگان کار با خانواده r را مناسب‌تر بیان کرده‌اند (کدیور و همکاران، ۱۳۸۹). به طور مثال فریدمن فرمول زیر را برای محاسبه d با داشتن متغیر r ارائه کرده است:

فرمول (۲)

$$d = \frac{2r}{[(1-r^2)^{0.5}]}$$

گزارش این شاخص در یک پژوهش علاوه بر اینکه می‌تواند در فراتحلیل به کار آید، می‌تواند نقطه شروعی برای پژوهش‌های تأییدی مشابه با نمونه‌های بزرگ‌تر باشد. فراتحلیل روشی است برای ترکیب کمی اطلاعات حاصل از چند پژوهش و در نتیجه در پی کشف روابط تازه‌ای است که از مطالعات جداگانه و انفرادی قابل حصول نیست (عابدی، ۱۳۸۴). از طرفی برای انجام فراتحلیل باید نتایج آماری تمام مطالعات را به اندازه اثر تبدیل کرد (ایزانلو، ۱۳۹۰). این موضوع بر اهمیت درک مفهوم اندازه اثر و آشنایی با محاسبه آن می‌افزاید.

مقدار p یک متغیر تصادفی است که از یک نمونه به نمونه‌ای دیگر متفاوت است. در نتیجه مقایسه این مقادیر بین پژوهش‌های مختلف مناسب نیست (گود و هاردین^۲، ۲۰۱۲). دقیقاً به همین دلیل است که در فراتحلیل از مقادیر معناداری آماری استفاده نشده و مقادیر اندازه اثر مقایسه می‌گردند. در ادامه مبانی نظری فراتحلیل به صورت اجمالی مورد بررسی قرار گرفته است.

مبانی نظری فراتحلیل

تقریباً از سال ۱۹۳۰ میلادی دانشمندان به بازنگری پیشینه تحقیق توجه نشان دادند. در همین سال‌ها برای اولین بار ثورندایک به منظور مطالعه تغییرات پایایی، نتایج ۳۶

^۱-Rosenthal & Dematteo

^۲-Good & Hardin

مطالعه جداگانه را که به بررسی آزمون بینه پرداخته بودند، گردآورد و میانگین ضرایب پایایی را محاسبه کرد. همچنین در همین دوره فیشر و پیرسون به طور جداگانه به خلاصه کردن روابط بین متغیرها در پژوهش‌های گوناگون پرداختند و روش‌هایی را برای ترکیب احتمالات ارائه دادند. در دهه ۱۹۶۰ میلادی روزنتال از طریق توسعه روش‌هایی که قبلاً ایجاد شده بودند، شروع به ترکیب و مقایسه نتایج مطالعات جداگانه نمود. در نهایت گلاس در سال ۱۹۷۶ در مقاله‌ای اصطلاح فراتحلیل را به کار برد (ازکیا و توکلی، ۱۳۸۵).

گلاس و همکاران برای توصیف اصطلاح فراتحلیل به انواع تحلیل‌ها اشاره کرده‌اند: (۱) تجزیه و تحلیل اولیه: وقتی پژوهشگر خود داده‌ها را جمع آوری کرده است و برای حصول نتایج، آن‌ها را تجزیه و تحلیل نماید. (۲) تجزیه و تحلیل ثانویه: وقتی پژوهشگر، داده‌های مربوط به مطالعات قبلی را برای پاسخگویی به سوالات جدید مورد تجزیه و تحلیل مجدد قرار می‌دهد. (۳) فراتحلیل: وقتی که پژوهشگر نتایج تعدادی از مطالعات اولیه و ثانویه را برای پاسخگویی به سوال خود ترکیب می‌کند و نتیجه جدیدی به دست می‌آورد (به نقل از کدیور و همکاران، ۱۳۸۹).

فراتحلیل انواع و تکنیک‌های متفاوتی دارد و به لحاظ روش شناختی بسیار متکثر است (کدیور و همکاران، ۱۳۸۹). با این حال فراتحلیل گران از هر رویکرد روش شناختی که بتواند صحت و دقت فراتحلیل را افزایش دهد، استقبال می‌کند (هانتز و اشمیت، ۲۰۰۴).

اندازه اثر اصلی‌ترین و معروف‌ترین اصطلاح مورد استفاده در روش فراتحلیل است که روزنتال و دمتو از آن با عنوان «شاه مهره» حیطه فراتحلیل یاد کرده‌اند (به نقل از کدیور و همکاران، ۱۳۸۹). کوهن معتقد است که اندازه اثر باید به یکی از دو صورت زیر تعریف شود: «حدی که پدیده مورد نظر در جمعیت موجود است» یا «حدی که فرضیه صفر نادرست است» (کوهن، ۱۹۸۸). همچنین فراتحلیل به حل تناقض‌ها در پژوهش‌های روان‌شناختی و تربیتی و دیگر علوم رفتاری و رسیدن به نتایج دقیق‌تر کمک می‌کند (دلاور، ۱۳۹۲).

هانتز و اشمیت (۲۰۰۴) معتقدند که تمامی فراتحلیل ها را می توان به کمک ضرایب همبستگی به عنوان شاخصی از اندازه اثر انجام داد. نویسندگان مختلف روش های متفاوتی برای تبدیل آزمون های مختلف پارامتریک و ناپارامتریک به ضرایب همبستگی ارائه کرده اند.

هر چند برخی محققان درباره کارایی فراتحلیل نقدهایی ارائه کرده اند ولی اکثریت جامعه علمی حاضر این روش را به عنوان روشی مفید و کارا می پذیرد (بکر، ۲۰۰۰). چنانچه در پژوهش های مورد بررسی برای فراتحلیل، اندازه اثر از سوی پژوهشگران گزارش نشده باشد می توان از جدول خطوط راهنمای کوهن استفاده نمود (فراهانی و عریضی، ۱۳۸۴).

در برخی مقالات نیز چگونگی به دست آوردن اندازه اثر برای مقالات پیشین و مورد نیاز برای فراتحلیل توضیح داده شده است (مانند تالهایمر و کوک، ۲۰۰۲). یادآوری این مطلب مهم است که فراتحلیل پاسخ قطعی و نهایی برای درک مسائل تحقیقی نیست و مانند سایر روش های تحلیل داده ها همان طور که گرین و هال می گویند: یاری گر اندیشه است، نه جایگزین آن (ولف، ۱۹۸۶ به نقل از دلاور، ۱۳۹۲).

نتیجه گیری

در پژوهش های علوم رفتاری پژوهشگر می بایست علاوه بر محاسبه معناداری آماری به محاسبه و گزارش شاخص مناسب اندازه اثر همت گمارد. از آنجایی که آزمون های معناداری در علوم رفتاری با درصد بالایی از خطای نوع دوم همراه است، به منظور کاهش خطای نوع دوم یا ارتقای توان آزمون علاوه بر راهبردهایی که روش شناسی پژوهش در اختیار پژوهشگران قرار می دهد، به برآورد اندازه اثر و استفاده از آن در تصمیم گیری در مورد رد یا قبول فرض صفر پرداخته می شود. انجمن

روانشناسی آمریکا^۱ گزارش این شاخص را در پژوهش‌ها پیشنهاد می‌کند. ارائه چنین شاخص‌هایی در واقع کاستی‌های مربوط به حجم نمونه را برطرف کرده و قابلیت استفاده مقاله را در فراتحلیل‌ها افزایش می‌دهد. این شاخص‌ها همچنین می‌توانند شروعی برای پژوهش‌های بعدی با حجم نمونه‌های بزرگ‌تر باشند. امروزه مقالات پژوهشی در ایران تمایل به گزارش شاخص‌های اندازه اثر پیدا کرده‌اند و تعداد مقالاتی که این شاخص‌ها را گزارش می‌کنند رو به افزایش است. با توجه به اینکه گزارش این شاخص‌ها عرصه را برای انجام فراتحلیل در حیطه‌های مختلف فراهم می‌کند، می‌توان انتظار داشت که میزان فراتحلیل‌های معتبر در آینده با استفاده از این گزارش‌ها افزایش یابد.

منابع

- ازکیا، مصطفی و محمود توکلی (۱۳۸۵)، فراتحلیل مطالعات رضایت شغلی در سازمان‌های آموزشی، فصلنامه علوم اجتماعی، ۲۷، ۱-۲۶.
- ایزانلو، بلال و حبیبی، مجتبی (۱۳۹۰)، کاربرد فراتحلیل در تحقیقات علوم اجتماعی و رفتاری: مروری بر مزایا، تنگناها و روش‌شناسی، تحقیقات علوم رفتاری، ۹، ۱، ۷۰-۸۳.
- دلاور، علی (۱۳۹۲)، مبانی نظری و عملی پژوهش در علوم انسانی و اجتماعی. تهران: رشد.
- سرم، زهره (۱۳۸۴)، آمار استنباطی: گزیده‌ای از تحلیل‌های آماری تک متغیری، انتشارات سمت، تهران.
- عابدی، احمد. حمیدرضا عریضی و علیرضا شواخی (۱۳۸۴)، فراتحلیل عوامل مؤثر بر افزایش کاربست یافته‌های پژوهشی در وزارت آموزش و پرورش. فصلنامه نوآوری آموزشی، ۱۲، ۴، ۱۰۹-۱۳۳.
- منصور، محمود (۱۳۷۶)، روانشناسی بالینی: آزمایش بالینی و فرآیند تشخیص، انتشارات بعثت، تهران.
- فراهانی، حجت‌الله و عریضی، حمیدرضا (۱۳۸۴)، روش‌های پیشرفته پژوهش در علوم انسانی، انتشارات جهاد دانشگاهی، اصفهان.

فراهانی، حجت الله؛ عریضی، حمیدرضا؛ رحیمی نژاد، عباس؛ قمرانی، امیر (۱۳۸۷)، بررسی پاره‌ای از خطاهای رایج آماری در پژوهش‌های رفتاری، فصلنامه حوزه و دانشگاه روش شناسی علوم انسانی، سال ۱۴، شماره ۵۵، ۸۵-۹۴.

فرگوسن، جرج. ا. و تاکانه، یوشیو (۱۳۹۲)، تحلیل آماری در روانشناسی و علوم تربیتی، ترجمه علی دلاور و سیامک نقشبندی، نشر ارسباران، تهران.
کدیور، پروین و همکاران (۱۳۸۹)، فراتحلیل مطالعات مربوط به عملکرد ریاضی، فصلنامه اندازه گیری تربیتی، ۷۵-۹۵.

American Psychological Association. (۱۹۹۴). *Publication manual of the American Psychological Association* (۴th edition), Washington, DC: Author.

American Psychological Association, (۲۰۰۱). *Publication manual of the American psychological association* (۵th edition), Washington D.C., American psychological association, Boston, MA: McGraw hill.

Becker, L.A. (۲۰۰۰). *Effect size (ES)*, web.uccs.edu/lbecker.

Carver, R. P. (۱۹۹۳). The case against statistical significance testing revisited. *Journal of Experimental Education*, ۶۱, ۲۸۷-۲۹۲.

Cohen, J. (۱۹۶۵). *Some statistical issues in psychological research*. In B. B. Wolman (Ed.), *Handbook of clinical psychology* (pp. ۹۵-۱۲۱). New York: Academic Press.

Cohen, J. (۱۹۸۸). *Statistical power analysis for the behavioral sciences* (second ed). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J. (۱۹۹۴). The earth is round ($p < .۰۵$). *American Psychologist*, ۴۹, ۹۹۷-۱۰۰۳.

Colman, A. M. (۲۰۰۹). *Oxford dictionary of psychology*, New York, Oxford University Press.

Good, P. I., & Hardin, J. W. (۲۰۱۲). *Common errors in statistics (and how to avoid them)*. John Wiley & Sons.

Hunter, J. E., & Schmidt, F. L. (۲۰۰۴). *Methods of meta-analysis: Thousand oaks*, CA: Sage publications.

Kirk, R. E. (۱۹۹۶). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, ۵۶, ۷۴۶-۷۵۹.

- Dwyer, J. H. (۱۹۷۴). Analysis of variance and the magnitude of effects. *Psychological Bulletin*, ۸۱, ۷۳۱-۷۳۷.
- Mitchell, C., & Hartmann, D. P. (۱۹۸۱). A cautionary note on the use of omega squared to evaluate the effectiveness of behavioral treatments. *Behavioral Assessment*, ۳, ۹۳-۱۰۰.
- Olejnik, S., & Algina, J. (۲۰۰۰). Measures of effect size for comparative studies: applications, interpretations, and limitations. *Contemporary educational psychology*, ۲۵, ۲۴۱-۲۸۶.
- Richardson, J. T. E. (۱۹۹۶). Measures of effect size. *Behavioral Research Methods, Instruments & Computers*, ۲۸, ۱۲-۲۲.
- Rosenthal, R., & Rubin, D. B. (۱۹۸۲). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, ۷۴, ۱۶۶-۱۶۹.
- Rosenthal, R. L., & Rosnow, R. (۱۹۹۱). Essentials of behavioral research: methods and data. *Research methods*, ۵, ۱۲۵-۱۳۲.
- Rosenthal, R., & Dematteo, M. R. (۲۰۰۱). Meta-analysis: Recent developments in Quantitative Methods for Literature Reviews. *Annual Review of Psychology* ۵۵, ۵۹- ۸۲.
- Thalheimer, W., & Cook, S. (۲۰۰۲). *How to calculate effect sizes from published research articles: A simplified methodology*. Retrieved November ۳۱, ۲۰۰۲ from http://work-learning.com/effect_sizes.htm
- Thompson, B. (۱۹۹۶). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher*, ۲۵, ۲۶-۳۰.
- Schmidt, F. L., & Hunter, J. E. (۱۹۹۷). *Eight common but false objections to the discontinuation of significance testing in the analysis of research data*. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (EDs.), what if there were no significance tests? (pp. ۳۷-۶۴). Mahwah, NJ: Erlbaum.
- Robinson, D. H., & wainer, H. (۲۰۰۲). On the past and future of null hypothesis significance testing. *Journal of wildlife management*, ۶۶, ۲۶۳-۲۷۱.

- Wilkinson, L., & APA Task Force on Statistical Inference. (۱۹۹۹). Statistical methods in psychology journals: Guidelines and explanations. *American psychologist*, ۵۴, ۵۹۴-۶۰۴.
- Thompson, B. (۲۰۰۱). Significance, effect size, stepwise methods, and other issues: strong arguments move the field. *Journal of experimental education*, ۷۰, ۸۰-۹۳.