

## فرا تحلیل روایی آزمون‌های مبتنی بر نقشه مفهومی<sup>۱</sup>

صادق صیادی\*  
جواد مصرآبادی\*\*  
رامین حبیبی کلیبر\*\*\*  
ابوالفضل فرید\*\*\*\*

### چکیده

هدف از اجرای پژوهش حاضر، فراتحلیل پژوهش‌هایی بود که روایی آزمون‌های مبتنی بر نقشه مفهومی را بررسی کرده بودند. روش‌های نمونه‌گیری پژوهش، هدفمند و گلوله برفی بودند. پس از حذف پژوهش‌های نامربوط، تعداد ۲۹ پژوهش به‌عنوان نمونه انتخاب شدند. این پژوهش‌ها شامل مقاله‌ها، پایان‌نامه‌ها و طرح‌های پژوهشی بود که در مجله‌های مختلف چاپ شده یا در دانشگاه‌ها و مؤسسات معتبر داخلی و خارجی اجرا شده بودند. از این پژوهش‌ها تعداد ۱۵۳ اندازه اثر به دست آمد. بعد از بررسی ملاک‌های ورود و خروج و حذف داده‌های پرت، تعداد ۱۴۳ اندازه اثر باقی ماند که با نرم‌افزار CMA ویرایش ۲ تحلیل شدند. نتایج فراتحلیل نشان داد که آزمون‌های مبتنی بر نقشه مفهومی دارای اندازه اثر ترکیبی ۰/۴۸۶ بوده که بر طبق ملاک کوهن برای تحقیقات همبستگی، اندازه اثر بزرگی محسوب می‌شود. نتایج به دست آمده از فراتحلیل حاضر نشان داد که آزمون‌های مبتنی بر نقشه مفهومی از روایی مناسبی برخوردارند و می‌توان از این آزمون‌ها برای هدف‌های ارزیابی در حیطه‌های مختلف استفاده کرد.

واژگان کلیدی: فراتحلیل، آزمون‌های مبتنی بر نقشه مفهومی، روایی

۱. این مقاله برگرفته از پایان‌نامه دکتری با عنوان «بررسی شاخص‌های روایی، پایایی، بار شناختی و نظام‌های نمره‌گذاری آزمون‌های مبتنی بر نقشه مفهومی» است.

\*دانشجوی دکتری رشته روان‌شناسی تربیتی، دانشگاه شهید مدنی آذربایجان، تبریز

\*\*استاد رشته روان‌شناسی تربیتی، گروه علوم تربیتی، دانشگاه شهید مدنی آذربایجان (نویسنده مسئول):

(mesrabadi@azaruniv.ac.ir)

\*\*\*دانشیار، رشته روان‌شناسی تربیتی، گروه علوم تربیتی، دانشگاه شهید مدنی آذربایجان، تبریز

\*\*\*\*استادیار، رشته روان‌شناسی تربیتی، گروه علوم تربیتی، دانشگاه شهید مدنی آذربایجان، تبریز

## مقدمه

ساختار دانش به شیوه‌ای اشاره دارد که افراد مفاهیم و ایده‌های یک حیطه مشخص را در حافظه بلندمدت خود به هم ربط داده و سازمان‌دهی می‌کنند. ساختار دانش، فرایندی مهم است به این دلیل که موجب سیالی در فعالیت‌های شناختی می‌شود و عملکرد حل مسئله سطح بالا توسط افراد خبره به آن نسبت داده می‌شود (کلاریانا<sup>۱</sup>، ۲۰۰۹). بنابراین به‌عنوان مؤلفه‌ای مهم اما ارزیابی نشده در فرایند کسب علم و دانش در نظر گرفته می‌شود (یین، وانیدس، رویز-پریمو، آیالا و شیولسون<sup>۲</sup>، ۲۰۰۵). واضح است که ارزیابی در جریان یادگیری، نقشی حیاتی دارد؛ بنابراین اگر بتوان پیامدهای واقعی یادگیری را تعیین کرد، روند تدریس و یادگیری به شیوه‌ای بسیار بهتر و مؤثرتر برنامه‌ریزی خواهد شد (اینگک<sup>۳</sup>، ۲۰۰۹). همچنین ارزیابی دانش به ما اجازه می‌دهد تا پیامدهای یادگیری را اندازه‌گیری کرده و مشخص کنیم که آیا فرایندهای آموزشی موفقیت‌آمیز بوده‌اند یا خیر؟ (استراوتمین<sup>۴</sup>، ۲۰۱۲). بنابراین ارزیابی ساختار دانش باید یک قسمت اساسی از ارزیابی منظم نتایج یادگیری باشد (کلاریانا، ۲۰۰۹).

رایج‌ترین روش ارزیابی نتایج یادگیری در کلاس‌های درس آزمون‌های چندگزینه‌ای هستند (آرنسون<sup>۵</sup>، ۲۰۰۵). بیشتر روش‌های سنتی ارزیابی مانند آزمون‌های چندگزینه‌ای، صحیح / غلط، پر کردن جاهای خالی و ... بر یادآوری ایده‌ها و مفاهیم پراکنده و مجزا تمرکز می‌کنند (شوندیمن<sup>۶</sup>، ۲۰۱۵). این روش‌های سنتی معمولاً کارایی لازم در زمینه ارزیابی را ندارند؛ برای مثال، آزمون‌های چندگزینه‌ای معمولاً روابط پیچیده بین مفاهیم را نشان نمی‌دهند. افزون بر این، آنها به میزان زیادی پاسخ‌های شاگردان را محدود و پی بردن به دلایل زیربنایی پاسخ‌ها را نیز مشکل می‌کنند. درنهایت اینکه با آزمون‌های چندگزینه‌ای می‌توان مفاهیم مهمی را که طراحان آزمون قصد ارزیابی آنها را دارند، مورد ارزیابی قرار داد اما برای ردیابی کج‌فهمی‌های<sup>۶</sup> شاگردان ابزار مفیدی نیستند. بنابراین، این آزمون‌ها در تشخیص مؤثر اطلاعات

---

1. Clariana

2. Yin, Vanides, Ruiz-Primo, Ayala & Shavelson

3. İnceç

4. Arneson

5. Schwendimann

6. Misconceptions

ارزشمند درباره جنبه‌های اساسی ساختار دانش آزمودنی‌ها شکست می‌خورند (لیو، لین و تسای<sup>۱</sup>، ۲۰۰۹).

ارزیابی ایدئال، به‌عنوان ارزیابی توصیف می‌شود که عینی و پایا بوده، تأثیر محیط (بافت) روی پاسخ‌ها را به حداقل برساند و مهم‌تر از همه اطلاعاتی درباره ماهیت ساختاری دانش یک فرد (یادگیرنده) فراهم آورد. تجربه عملی از ارزیابی به ما می‌گوید که اشکال آزمون‌های عینی (مانند سؤال‌های چندگزینه‌ای، آیتم‌های تکمیل کردنی، آزمون‌های کوتاه‌پاسخ و ...) اگرچه ممکن است که عینی بوده و برای ارزیابی آسان باشند، اما پاسخ دادن به آنها اغلب مبتنی بر حافظه است. چنین آزمون‌هایی تمایل دارند تا عوامل اصلی دانش را ارزیابی کنند اما در تشخیص ساختار دانش شاگردان با مشکل روبرو هستند (لیپینگ، یان، بوکینگ و زنگیائو<sup>۲</sup>، ۲۰۰۹). بر این اساس، پژوهشگران ابداع راهبردهای ارزیابی ابتکاری برای پیشبرد یادگیری فردی شاگردان و بینش نسبت به ارزیابی دوره‌های آموزشی و تحصیلی را ضروری دانسته‌اند (واتسون، پلکی، نویس و راجرز<sup>۳</sup>، ۲۰۱۶). در همین باره، پژوهشگران حوزه آموزش و پرورش در جهت کشف اشکال مؤثر ارزیابی به دلیل نارضایتی از روش‌های ارزیابی سنتی تلاش کرده‌اند. یکی از این ابزارهای ارزیابی جانشین، نقشه‌های مفهومی است (لیو<sup>۴</sup>، ۲۰۱۳).

پژوهشگران زیادی در سطح جهان به استفاده از نقشه‌های مفهومی در زمینه ارزیابی تأکید کرده‌اند (رویز-پریمو<sup>۵</sup>، ۲۰۰۰؛ یین و شیولسون<sup>۶</sup>، ۲۰۰۸؛ مصرآبادی، ۱۳۹۰؛ وودووزوف و راود<sup>۷</sup>، ۲۰۱۵؛ تزنگ<sup>۸</sup>، ۲۰۱۴). نقشه‌های مفهومی در مقایسه با اشکال سنتی ارزیابی، چندین مزیت عمده دارند: ۱- برخلاف اشکال ارزیابی مبتنی بر یادآوری، نقشه‌های مفهومی اشکال زایایی از ارزیابی هستند که می‌توانند درک نسبی شاگردان را نیز آشکار سازند؛ ۲- برای درک و به‌کارگیری مفاهیم، باید مفاهیم را به

1. Liu, Lin, & Tsai

2. Liping, Yan, Boqing, & Zengyao

3. Watson, Pelkey, Noyes, & Rodgers

4. Liu

5. Ruiz-Primo

6. Yin & Shavelson

7. Vodovozov & Raud

8. Tzeng

مفاهیم موجود ربط داد، چراکه ارتباط بین مفاهیم ویژگی اصلی و ضروری دانش است. برخلاف اشکال سنتی ارزیابی که بر یادآوری مفاهیم و ایده‌های مجزا و پراکنده تمرکز می‌کنند، نقشه‌های مفهومی ارتباطات بین مفاهیم و ایده‌ها را نیز نشان می‌دهند؛ ۳- افراد خبره و یادگیرندگان موفق معمولاً چارچوب منسجم و کاملاً یکپارچه‌ای از مفاهیم مرتبط را توسعه می‌دهند. نقشه‌های مفهومی می‌توانند سازمان‌دهی دانش شاگردان را از راه نشان دادن ارتباطات، خوشه‌های مفاهیم، سطوح سلسله‌مراتب و ارتباطات عرضی بین مفاهیم به تصویر کشند. نقشه‌های مفهومی می‌توانند به جای حفظ کردن ایده‌ها و مفاهیم پراکنده توسط شاگردان، یادگیری آنها را در جهت درک مفهومی تسریع بخشند؛ ۵- نتایج پژوهش‌ها نشان‌دهنده آن است که نقشه‌های مفهومی برخلاف اشکال سنتی ارزیابی می‌توانند اشکال متفاوت دانش را ارزیابی کنند (شوندیمن، ۲۰۱۵). تعداد زیادی از پژوهش‌ها نشان می‌دهند که تکنیک‌های ارزیابی مبتنی بر نقشه مفهومی می‌توانند سطوح شناختی سطح بالای شاگردان و مفهومی‌سازی درک شاگردان را ارزیابی کنند. آنها همچنین ویژگی‌های یادگیری معنی‌دار را نیز (به جهت داشتن رویکرد سلسله‌مراتبی و بصری‌سازی ساختار یافته) دارند. و درنهایت آنها می‌توانند به صورت مستقیم ساختارهای دانش ذهنی شاگردان را نشان دهند (نواک و گووین<sup>۱</sup>، ۱۹۸۴؛ به نقل از لیبینگ، یان، بوکینگ و زنگیانو، ۲۰۰۹).

با وجود تمام آنچه گفته شد استفاده از آزمون‌های مبتنی بر نقشه مفهومی در مقایسه با آزمون‌های مرسوم از پیچیدگی زیادی برخوردار است؛ چراکه این آزمون‌ها پراکندگی و تنوع زیادی در تمام حیطه‌ها (مثلاً تکمیل یا ترسیم؛ شیوه‌های نمره‌دهی و ...) دارند (مصرآبادی، ۱۳۹۰). در همین باره، رویز-پریمو و شیولسون (۱۹۹۶) خاطرنشان ساختند پیش از اینکه بتوان از تکنیک‌های مختلف نقشه‌های مفهومی برای هدف‌های ارزیابی استفاده کرد باید اطلاعات موثقی را درباره روایی و پایایی آنها فراهم آورد (بین و شیولسون، ۲۰۰۸). در پژوهش‌های علوم رفتاری و اجتماعی، روایی اهمیت بسیاری دارد. اگر پژوهشگر نتواند ابزار اندازه‌گیری را که از لحاظ خصیصه مورد نظر روایی کافی دارد تهیه کند، پژوهش وی بی‌ارزش خواهد بود زیرا برای آنکه نتایج حاصل از اندازه‌گیری معنای مشخصی داشته باشد، باید میزان روایی ابزار سنجش از پیش معلوم باشد. بسیاری از انتقادهایی که به اندازه‌گیری‌های روانی،

<sup>1</sup>. Novak & Gowin

تربیتی و اجتماعی، از سوی متخصصان و سایر افراد، وارد شده متمرکز بر روایی ابزار اندازه‌گیری است (هومن، ۱۳۹۰). بنابراین روایی نقشه‌های مفهومی موضوع بسیار مهمی است (لیو و همکاران، ۲۰۰۹).

روایی اساساً پاسخگویی به این پرسش است که: آیا ابزار ارزیابی آنچه را که قصد اندازه‌گیری آن را دارد، اندازه می‌گیرد؟ (جانسون و سوینگبی<sup>۱</sup>، ۲۰۰۷). روایی انواع گوناگونی دارد که روایی محتوایی، ملاکی و افتراقی<sup>۲</sup> از آن جمله‌اند. روایی محتوایی به این مطلب اشاره دارد که نمونه سؤال‌های مورد استفاده در یک آزمون تا چه حد می‌تواند معرف کل سؤال‌های ممکن باشد که می‌توان از محتوای مورد نظر تهیه کرد (سیف، ۱۳۸۹). روایی ملاکی، میزان همبستگی بین نمره‌های به دست آمده از یک مقیاس جدید است با نمره‌های به دست آمده از یک مقیاس ملاک (مرجع) که مقیاسی معتبر است. اگر نمره‌های مقیاس ملاک هم‌زمان با مقیاس موجود گردآوری شود هدف برآورد روایی هم‌زمان<sup>۳</sup> است؛ اما زمانی که نمره‌های مقیاس ملاک بعداً گردآوری شود، هدف برآورد روایی پیش‌بین<sup>۴</sup> است (پلومر، ۲۰۰۸). روایی افتراقی (تفکیکی یا تمیزی) نیز توانایی یک آزمون در جدا کردن گروه‌های مختلف افراد است (مصرآبادی، ۱۳۸۸). مک‌میلان<sup>۵</sup> (۲۰۰۴) روایی را ویژگی تعریف کرده است که به مناسب بودن استنباط‌ها، کاربردها، صحت، قابلیت اعتماد، یا مشروعیت ادعاها یا استنباط‌هایی که بر مبنای امتیازات به دست آمده به عمل می‌آید، اشاره دارد. به این معنی که هر گاه استنباط‌هایی درباره ساختار دانش یا ساختارهای دانش طرح‌واره‌ای یادگیرندگان به عمل می‌آید، روش‌های ارزیابی مبتنی بر نقشه مفهومی باید نمره‌هایی ارائه کنند که استنباط‌های دقیق و درستی درباره ساختارهای دانش طرح‌واره‌ای یادگیرندگان را فراهم آورند (سیپرین‌ساس<sup>۶</sup>، ۲۰۱۰). روایی در ارزیابی نقشه‌های مفهومی کیفیت استنباط‌هایی را منعکس می‌کند که می‌توان در ارتباط با دانش یادگیرنده از فعالیت ترسیم نقشه مفهومی به دست آورد (کلاین، بروستر و فل<sup>۷</sup>،

1. Jonsson & Svingby

2. Content validity, Criterion validity & Discriminative validity

3. Concurrent validity

4. Predictive validity

5. McMillan

6. Ciprian Sas

7. Cline, Brewster & Fell

۲۰۱۰). در همین رابطه، پژوهشگران حیطه ارزیابی از نقشه‌های مفهومی به‌عنوان نشانگرهای دانش استفاده می‌کنند که این دانش با کیفیت گزاره، کیفیت مجموع گزاره‌ها و کمیت گزاره‌ها اندازه‌گیری می‌شود. کیفیت گزاره، روایی هر گزاره را نشان می‌دهد که دارای کمترین ساختار پیچیدگی نشان داده شده در یک نقشه مفهومی است، در حالی که کیفیت مجموع، استنباط‌های گسترده‌تری راجع به درک و فهم شاگردان را نشان می‌دهد که با استفاده از چندین گزاره به دست می‌آید. کمیت گزاره‌ها نیز مقدار اطلاعات پردازش شده توسط یادگیرنده را در طول فرایند نقشه کشی مفهومی نشان می‌دهد (لیم<sup>۱</sup>، ۲۰۰۸).

عوامل بسیاری روایی نقشه‌های مفهومی را تحت تأثیر قرار می‌دهند که برخی از آنها عبارت‌اند از: الف- نوع تکلیف ترسیم نقشه مفهومی ارائه شده که ممکن است محدودیت‌هایی را بر فعالیت‌های نقشه‌کشی یادگیرندگان تحمیل کند (مثلاً ساخت یا تکمیل نقشه‌های مفهومی)؛ ب- شیوه‌ای که یادگیرندگان نقشه‌ها را تولید می‌کنند (مداد - کاغذی یا کامپیوتری)؛ ج- روش امتیازدهی استفاده شده برای ارزیابی کردن نقشه‌ها و د- آزمودنی‌هایی که تکالیف نقشه مفهومی را انجام می‌دهند (دانش‌آموزان، دانشجویان و ...). محدودیتی که هر یک از این مؤلفه‌ها روی یادگیرندگان یا نمره‌گذاران نقشه‌های مفهومی ایجاد می‌کند روی روایی آزمون‌های مبتنی بر نقشه مفهومی تأثیر می‌گذارد. علاوه بر موارد ذکر شده، خود روش‌های برآورد روایی آزمون‌های مبتنی بر نقشه مفهومی نیز از پراکندگی زیادی برخوردارند چراکه روایی می‌تواند از راه روایی محتوایی، ملاکی، افتراقی و ... برآورد شود.

پژوهش‌های بسیاری در سطح جهان، روایی آزمون‌های مبتنی بر نقشه مفهومی را بررسی کرده‌اند هر چند که در ایران تعداد این پژوهش‌ها بسیار اندک است. از جمله پژوهش‌های خارجی در این زمینه می‌توان به پژوهش رویز-پریمو، شولتز، لی و شیولسون<sup>۲</sup> (۲۰۰۱)؛ مک‌کلور، سوناک و سوئن<sup>۳</sup> (۱۹۹۹)؛ رویز-پریمو (۲۰۰۰)؛ کول، کلاریانا و صالحی<sup>۴</sup> (۲۰۰۵)؛ پلومر (۲۰۰۸)؛ ریسکا، سویکا، مولیتس، رانیکا و

1. Lim

2. Ruiz-Primo, Schultz, Li & Shavelson

3. McClure, Sonak & Suen

4. Coul, Clariana & Salehi

سویارد<sup>۱</sup> (۲۰۱۵)؛ اکین، اولوسینار ساگیر و سالتان<sup>۲</sup>، (۲۰۱۶) اشاره کرد. از پژوهش‌های داخلی فقط دو پژوهش در این زمینه یافت شد که شامل مصرآبادی (۱۳۸۸) و عبادالله‌وند (۱۳۹۵) بود. با توجه به یافته‌های نسبتاً متنوع و متفاوت پژوهش‌های انجام گرفته در حیطه بررسی روایی آزمون‌های مبتنی بر نقشه مفهومی (به علت تنوع در روش‌های نمره‌گذاری، جامعه و نمونه متفاوت، متفاوت بودن ساختار تکالیف مانند تکمیل یا ساخت، متفاوت بودن شیوه تهیه نقشه‌های مفهومی، روش‌های متفاوت برآورد روایی و ...) دستیابی به نتیجه‌ای کلی و روشن در این زمینه به نظر ضروری می‌رسد. بنابراین به منظور منسجم‌سازی و یکپارچه کردن نتایج این پژوهش‌ها و کم کردن ابهام در این حیطه، پژوهش حاضر با هدف انجام یک فرا تحلیل مدون انجام گرفت. از سوی دیگر، جستجوهای صورت گرفته در پایگاه‌های داخل و خارج از کشور نشان‌دهنده آن است که در زمینه بررسی روایی آزمون‌های مبتنی بر نقشه مفهومی تاکنون فرا تحلیل مدونی انجام نگرفته است. بنابراین در پژوهش حاضر با ترکیب نتایج پژوهش‌های مختلف صورت گرفته در این زمینه به این سؤال اساسی پاسخ داده می‌شود که آیا آزمون‌های مبتنی بر نقشه مفهومی از روایی مناسبی برخوردارند؟ به عبارت دیگر هدف اساسی از اجرای پژوهش این است که مشخص کند که آیا آزمون‌های مختلف مبتنی بر نقشه مفهومی هنگام استفاده شدن به عنوان ابزار ارزیابی از روایی مطلوب و مناسبی که مشخصه ضروری هر ابزار ارزیابی است برخوردار هستند یا خیر؟

پژوهش‌های حوزه روان‌شناسی، علوم تربیتی و علوم رفتاری گاهی به نتایج متناقض می‌انجامند. تفاوت در تدابیر آزمایشی، موقعیت‌ها، ابزارهای اندازه‌گیری و روش‌های پژوهش موجب می‌شود که پژوهشگر به آسانی نتواند یافته‌ها را مقایسه کند. حتی تکرار متوالی پژوهش نیز ممکن است بی‌فایده باشد. فرا تحلیل نتایج مطالعات مختلف را به مقیاسی مشترک تبدیل کرده و با روش‌های آماری رابطه بین ویژگی‌های مطالعات و یافته‌ها را بررسی می‌کند. مرور نظام‌مند پژوهش‌های پیشین موجب سازمان‌دهی و بهره‌برداری بیشتر از پژوهش‌ها و ایجاد ارتباط بین پژوهش‌های گذشته و آینده می‌شود (گلاس<sup>۳</sup>، ۱۹۸۳؛ به نقل از رعایی، دلاور و فرخی، ۱۳۸۹). همچنین

1. Reiska, Soika, Möllits, Rannikmäe & Soobard

2. Ekin, Ulucinar Sagir, Saltan

3. Glass

نتایج فراتحلیل می‌تواند در زمینه وجود ناسازگاری‌ها، تعارضات یا تناقضات در نتایج پژوهش‌ها، تعیین اندازه اثر کلی، تشخیص متغیرهای تعدیل‌کننده احتمالی، تشخیص حوزه‌های پژوهشی مسئله‌دار و تشخیص حوزه‌هایی که پژوهشگران گذشته به آن توجه نکرده‌اند نیز راهگشا باشد (ایزانلو و حبیبی، ۱۳۹۰).

### روش‌شناسی

فرضیه اصلی این پژوهش، ترکیب و مقایسه یافته‌های متفاوت و مجزای پژوهش‌های انجام شده در زمینه روایی آزمون‌های مبتنی بر نقشه مفهومی و به دست آوردن نتیجه‌ای منسجم از میان نتایج پراکنده و متنوع پژوهش‌ها با استفاده از روش‌های آماری نیرومند بود. همچنین یافتن اندازه اثر واقعی و میانگین و انحراف معیار اندازه اثرها و آگاهی از میزان سوگیری انتشار و ناهمگنی در پژوهش‌های اولیه از دیگر هدف‌های این پژوهش بود. بنابراین در پژوهش حاضر با توجه به ماهیت پژوهش از روش فراتحلیل استفاده شد. روش فراتحلیل به تحلیل‌های آماری اشاره می‌کند که برای ترکیب داده‌های یک مجموعه از مطالعات انجام می‌گیرد (برنشتاین، هجز، هیگینز و روزشتاین<sup>۱</sup>، ۲۰۰۹).

### جامعه آماری، حجم نمونه، روش نمونه‌گیری

ایده بنیادینی که فراتحلیل بر آن استوار است این است که واحد تجزیه و تحلیل از مطالعه اخذ می‌شود نه از آزمودنی. بر این اساس جامعه آماری پژوهش حاضر، همه پایان‌نامه‌های کارشناسی ارشد و دکتری، پژوهش‌های اجرا شده در دستگاه‌های اجرایی یا مؤسسات و پژوهش‌های چاپ شده در مجلات علمی پژوهشی معتبر داخل و خارج از کشور بودند که روایی نقشه‌های مفهومی را بررسی کرده و در فاصله سال‌های ۱۹۹۶ تا ۲۰۱۷ اجرا شده بودند. این پژوهش‌ها در پایگاه‌های تحقیقاتی مختلف<sup>۲</sup> منتشر شده بودند. حجم نمونه نیز شامل ۲۹ پژوهش بود که به روش نمونه‌گیری هدفمند<sup>۳</sup> و گلوله برفی<sup>۴</sup> انتخاب شدند. در نمونه‌گیری هدفمند پژوهشگران بر اساس

1. Borenstein, Hedges, Higgins & Rothstein

2. Springer, Taylor & Francis, Science Direct, ResearchGate, John Wiley & Sons, Sage, Blackwell Science, Wiley Periodicals, Elsevier Science, proQuest

3. Purposive

4. Snow ball



قضاوت خود پژوهش‌هایی را به‌عنوان مطالعات اولیه انتخاب کردند که در راستای هدف‌های فراتحلیل بودند. پس از انتخاب این نمونه‌ها پژوهشگر با مراجعه به منابع این مطالعات، دیگر پژوهش‌های مربوطه را به روش گلوله برفی انتخاب کرد. برای انتخاب پژوهش‌های اولیه از چارچوب نمونه‌گیری، نخست کلیدواژه‌های معتبر (مانند روایی، شاخص‌های روان‌سنجی، ابزارهای ارزیابی جدید، نقشه مفهومی، آزمون‌های مبتنی بر نقشه مفهومی) بر اساس مرور پیشینه پژوهشی به‌منظور استفاده در جستجوی پژوهش‌های اولیه، تعیین و سپس بر اساس ملاک‌های ورود و خروج، پژوهش‌های مورد نظر شناسایی و انتخاب شدند. اطلاعات کامل مربوط به این پژوهش‌ها در جدول (۱) ارائه شده است. برای تعیین نمونه آماری پژوهش حاضر سه چارچوب نمونه‌گیری تعریف شد:

الف- مقاله‌هایی که در مجلات و همایش‌های معتبر داخلی و خارجی قابل دستیابی بودند. از این مجلات می‌توان به science direct, sage, springer اشاره کرد. لازم به ذکر است که از مجلات و همایش‌های داخلی پژوهشی یافت نشد.

ب- پایان‌نامه‌های دانشجویی در سطح کارشناسی ارشد و دکتری در داخل کشور که در کتابخانه دانشگاه‌ها قابل دسترسی بودند. گفتنی است تنها یک پایان‌نامه کارشناسی ارشد در دانشگاه شهید مدنی در این زمینه در کشور صورت گرفته بود و در سایر دانشگاه‌های کشور، پژوهشی در این حیطه به دست نیامد. در خارج از کشور نیز پایان‌نامه‌هایی که در سایت proquest (قابل دسترسی در سایت کتابخانه دانشگاه شهید مدنی آذربایجان) منتشر شده بودند به‌عنوان نمونه انتخاب شدند.

پ- طرح‌های پژوهشی.

### ابزار گردآوری اطلاعات

در مطالعات فراتحلیل از آنجا که چارچوب و نحوه استخراج مطالب از قبل کاملاً طراحی می‌شود، بنابراین کدبرداری به‌منظور استخراج اطلاعات از پژوهش‌های اولیه با استفاده از فرم‌های از قبل طراحی شده، انجام می‌گیرد. در پژوهش حاضر نیز بر اساس معیارهای مورد نظر فراتحلیل و نیز هدف‌های پژوهش از فرم گردآوری اطلاعات تحقیقات اولیه مصرآبادی (۱۳۹۵) استفاده شد که امکان گردآوری اطلاعات کتاب‌شناختی، روش‌شناختی و داده‌های موردنیاز در محاسبه اندازه اثر را فراهم می‌کند.

### ملاک‌های ورود و خروج پژوهش‌ها به فراتحلیل:

در پژوهش حاضر به منظور انتخاب نمونه از پژوهش‌های اولیه، ملاک‌هایی برای ورود و خروج پژوهش‌ها در نظر گرفته شد. ملاک‌های ورود پژوهش‌ها به فراتحلیل عبارت بودند از:

۱. پژوهش‌ها می‌بایست بین سال‌های ۱۹۹۶ تا ۲۰۱۷ انجام شده بودند (۲۰ سال اخیر).
  ۲. پژوهش‌ها می‌بایست به بررسی روایی نقشه‌های مفهومی پرداخته بودند.
  ۳. پژوهش‌ها می‌بایست اندازه اثر یا اطلاعات لازم برای محاسبه آن را ارائه کرده بودند.
  ۴. پایان‌نامه‌ها می‌بایست در سطح کارشناسی ارشد یا دکتری اجرا شده بودند.
- ملاک‌های خروج پژوهش‌ها از فراتحلیل نیز عبارت بودند از:
۱. پژوهش‌هایی که یکی از اطلاعات لازم برای محاسبه اندازه اثر را گزارش نداده بودند.
  ۲. مقاله‌های مشابه که با عناوین مختلف در دو نشریه چاپ شده بودند.
  ۳. پژوهش‌هایی که دارای ضعف‌های روش‌شناختی جدی (مثلاً روش نمونه‌گیری نامناسب) بودند.
  ۴. پژوهش‌هایی که با استفاده از ابزارها و روش‌های معتبر و دارای ویژگی‌های روان‌سنجی مطلوب اجرا نشده بودند.
  ۵. پژوهش‌هایی که دارای نقص اطلاعات کتاب‌شناختی بودند (مثلاً پژوهش‌هایی که نام پدیدآورندگان، نام فصلنامه یا مرکز اجرای پژوهش را ذکر نکرده بودند).
- لازم به ذکر است که در پژوهش حاضر فقط روایی ملاکی و افتراقی بررسی شد چراکه تنها تعداد ۴ اندازه اثر مربوط به روایی محتوایی بود که این ۴ اندازه اثر نیز در جریان تحلیل حساسیت از پژوهش خارج شدند.

### روش آماری

در تجزیه و تحلیل داده‌های گردآوری شده از نرم‌افزار CMA ویرایش ۲ استفاده شد. نرم‌افزار CMA یک برنامه پردازش آماری است که داده‌ها را بر اساس روش

فرا تحلیل پردازش می‌کند. برای بررسی و تحلیل داده‌ها از اندازه اثر ترکیبی با دو مدل اثرات ثابت و تصادفی، نمودار قیفی<sup>۱</sup> به منظور انجام تحلیل حساسیت و شاخص‌های  $Q$  کوکران و آماره  $I^2$  به منظور بررسی ناهمگنی در پژوهش‌ها استفاده شد. همچنین پس از شناسایی پژوهش‌ها و اعتبارسنجی آنها، برای تحلیل نتایج از روش ترکیب آزمون‌ها و محاسبه اندازه اثر گروه ۲ استفاده شد. اندازه‌های اثر گروه ۲ همان ضرایب همبستگی هستند چراکه خود ضریب همبستگی یک آماره استاندارد شده است. به این معنی که ضرایب همبستگی پژوهش‌های مختلف با هم قابل مقایسه هستند و نیازی به محاسبه شاخص دیگری نیست.

---

1. Funnel Plot

## یافته‌ها

جدول (۱) اطلاعات عمومی مربوط به پژوهش‌های اولیه

پژوهشگران	شاخص آماری استخراج شده از پژوهش	سال نشر	نوع پژوهش	محل چاپ
مصرآبادی	همبستگی، شاخص t	۱۳۸۸	طرح پژوهشی	دانشگاه شهید مدنی آذربایجان
عبادالله‌وند	همبستگی	۱۳۹۵	پایان‌نامه کارشناسی ارشد	دانشگاه شهید مدنی آذربایجان
Ingec	همبستگی	۲۰۰۹	مقاله	International Journal of Science Education
McClure, sonak & suen	همبستگی	۱۹۹۹	مقاله	Journal of Research In Science Teaching
Yin, Vanides, Ruiz- Primo, Ayala & Shavelson	همبستگی	۲۰۰۵	مقاله	Journal of Research In Science Teaching
Plummer	همبستگی	۲۰۰۸	پایان‌نامه دکتری	Brigham Young University
Stoddart, Abrams, Gasper & Canaday	همبستگی	۲۰۰۰	مقاله	International Journal of Science Education
Coul, Clariana & Salehi	همبستگی	۲۰۰۵	مقاله	Journal of Educational Computing Research
Nicoll, Francisc & Nakhleh	همبستگی	۲۰۰۱	مقاله	International Journal of Science Education
Crvalho	میانگین، انحراف معیار و حجم نمونه دو گروه - همبستگی	۲۰۱۰	پایان‌نامه کارشناسی ارشد	San Jose State University
Reiska, Soika, Möllits, Rannikmäe & Soobard	همبستگی	۲۰۱۵	مقاله	Journal of Social and Behavioral Sciences
Ekin, Ulucinar Sagir & Saltan	همبستگی	۲۰۱۶	مقاله	Participatory Educational Research (PER)
Tsai, Lin & Yuan	همبستگی	۲۰۰۱	مقاله	Journal of Computer Assisted Learning
Wu, Hwang, Milrad, Ke & Huang	شاخص t	۲۰۱۲	مقاله	British Journal of Educational Technology
Kassab & Hussain	همبستگی	۲۰۱۰	مقاله	Journal of Medical Teacher

محل چاپ	نوع پژوهش	سال نشر	شاخص آماری استخراج شده از پژوهش	پژوهشگران
Poster presented at 18 <sup>th</sup> International Conference on Chemical, İstanbul, Turkey.	مقاله	۲۰۰۴	همبستگی	Kaya & Kilic
Journal of Research In Science Teaching	مقاله	۲۰۰۱	همبستگی	Ruiz-Primo, Schultz, Li & Shavelson
Journal of Educational Technology & Society	مقاله	۲۰۰۵	همبستگی	Chang, Sung, Chang & Lin
Pennsylvania State University	پایان‌نامه دکتری	۲۰۰۸	همبستگی	Lim
Journal of Research in Science Teaching	مقاله	۱۹۹۸	همبستگی	Rice, Ryan & Samson
International Journal of Science Education	مقاله	۱۹۹۶	همبستگی	Liu & Hinchey
Journal of School Science and Mathematics	مقاله	۲۰۰۲	همبستگی	Rye & Rubba
Journal of Informetrics	مقاله	۲۰۱۵	همبستگی	Sasson, Ravid & Pliskin
Journal of Educational Assessment	مقاله	۲۰۱۷	همبستگی	Won, Krabbe, Ley, Treagust & Fischer
Journal of Education	مقاله	۲۰۱۱	همبستگی	Eroglu & Kelecioglu
Journal of Educational and Psychological Measurement	مقاله	۲۰۰۱	همبستگی	Schau, Mattern, Zeilik, Teague & Weber
Fifth International Conference on Concept Mapping Valletta, Malta	مقاله	۲۰۱۲	همبستگی	Ley, Krabbe & Fischer
Journal of Educational Studies	مقاله	۲۰۱۲	همبستگی	Conradty & Bogner
CRESST/University of California, Los Angeles	طرح پژوهشی	۲۰۰۲	همبستگی	Klein, Chung, Osmundson & Herl

در جدول (۱) اطلاعات عمومی مربوط به ۲۹ پژوهش اولیه ارائه شده است که در پژوهش حاضر مورد تحلیل قرار گرفتند. در تمامی این پژوهش‌ها نقشه‌های مفهومی بررسی روایی شده بودند.

جدول (۲) اندازه اثرهای حاصل از پژوهش‌های اولیه پیش از تحلیل حساسیت

اندازه اثر	مطالعه	اندازه اثر	مطالعه
۱/.	ریسکا و همکاران (۲۰۱۵)	*./۳۵۴	مصرآبادی
*./۲۴۵	ریسکا و همکاران (۲۰۱۵)	*./۴۴۸	مصرآبادی
*./۲۴۵	ریسکا و همکاران (۲۰۱۵)	*./۹۷۳	مصرآبادی
*./۱۹۲	ریسکا و همکاران (۲۰۱۵)	*./۴۲۴	مصرآبادی
*./۲۹۹	ریسکا و همکاران (۲۰۱۵)	*./۸۲۹	مصرآبادی
*./۳۴۳	ریسکا و همکاران (۲۰۱۵)	*./۷۰۹	مصرآبادی
*./۵۶۵	اکین و همکاران (۲۰۱۶)	*./۵۳۶	مصرآبادی
./۲۷۸	اکین و همکاران (۲۰۱۶)	۱/۲۲۱	■ مصرآبادی
*./۱۱۶	اکین و همکاران (۲۰۱۶)	./۹۷۳	■ مصرآبادی
./۲۲۸	اکین و همکاران (۲۰۱۶)	*./۳۲۹	مصرآبادی
./۲۷۷	تسای و همکاران (۲۰۰۱)	*./۴۷۱	مصرآبادی
*./۱۵۷	■ تسای و همکاران (۲۰۰۱)	*./۷۰۲	مصرآبادی
*./۳۵۵	وو و همکاران <sup>۱</sup> (۲۰۱۲)	*./۳۹	مصرآبادی
./۰۷۸	کساب و حسین (۲۰۱۰)	*./۵۰۶	مصرآبادی
*./۲۴۵	کساب و حسین (۲۰۱۰)	*./۴۷۹	مصرآبادی
./۲۱۴	کساب و حسین (۲۰۱۰)	*./۲۰۷	مصرآبادی
./۱۶۸	کساب و حسین (۲۰۱۰)	*./۳۷۷	ایننگ (۲۰۰۹)
*./۲۹۵	کساب و حسین (۲۰۱۰)	*./۳۶۵	ایننگ (۲۰۰۹)
./۰۶۹	کساب و حسین (۲۰۱۰)	./۲۲۴	ایننگ (۲۰۰۹)
./۱۰۷	کساب و حسین (۲۰۱۰)	*./۳۵۴	ایننگ (۲۰۰۹)
./۱۳۶	کساب و حسین (۲۰۱۰)	*./۲۵۵	ایننگ (۲۰۰۹)
*./۱۹۴	کساب و حسین (۲۰۱۰)	*./۴۰۸	مک‌کلور و همکاران (۱۹۹۹)
./۱۶۵	کساب و حسین (۲۰۱۰)	*./۴۸۲	مک‌کلور و همکاران (۱۹۹۹)
*./۵۲	کایا و کیلیک (۲۰۰۴)	*./۳۹۹	مک‌کلور و همکاران (۱۹۹۹)
*./۳۸۸	رویز-پریمو و همکاران (۲۰۰۱)	*./۷۰۶	مک‌کلور و همکاران (۱۹۹۹)
*./۴	رویز-پریمو و همکاران (۲۰۰۱)	*./۳۹۲	مک‌کلور و همکاران (۱۹۹۹)
*./۷۷۵	رویز-پریمو و همکاران (۲۰۰۱)	./۱۹۵	مک‌کلور و همکاران (۱۹۹۹)
*./۷۹۳	رویز-پریمو و همکاران (۲۰۰۱)	./۲۶۵	بین و همکاران (۲۰۰۵)
*./۵۶۳	رویز-پریمو و همکاران (۲۰۰۱)	*./۶۲	بین و همکاران (۲۰۰۵)
*./۴۷۲	رویز-پریمو و همکاران (۲۰۰۱)	*./۷۹۳	پلومر (۲۰۰۸)
*./۴۹۷	رویز-پریمو و همکاران (۲۰۰۱)	*./۷۴۱	پلومر (۲۰۰۸)

<sup>۱</sup>. Wu et al

اندازه اثر	مطالعه	اندازه اثر	مطالعه
*.۷۴۱	چانگ و همکاران (۲۰۰۵)	*.۷۲۵	پلومر (۲۰۰۸)
*.۵۲۳	چانگ و همکاران (۲۰۰۵)	*.۷۲۵	پلومر (۲۰۰۸)
*.۴۵۱	ساسون و همکاران (۲۰۱۵)	*.۹۷۳	پلومر (۲۰۰۸)
*.۲۵۵	ساسون و همکاران (۲۰۱۵)	*.۱/۲۷	پلومر (۲۰۰۸)
*.۱/۰۰۳	ساسون و همکاران (۲۰۱۵)	*.۹۷۳	پلومر (۲۰۰۸)
*.۶/۰۶	ساسون و همکاران (۲۰۱۵)	*.۷۷۵	پلومر (۲۰۰۸)
*.۳۳۷	ساسون و همکاران (۲۰۱۵)	*.۲/۱۸۵	■ استودارت و همکاران (۲۰۰۰)
*.۱/۳۷۱	ساسون و همکاران (۲۰۱۵)	.۳۵۴	کول و همکاران (۲۰۰۵)
*.۱/۸۴۲	■ ساسون و همکاران (۲۰۱۵)	- .۰۷	کول و همکاران (۲۰۰۵)
*.۱/۷۳	■ ساسون و همکاران (۲۰۱۵)	.۴	کول و همکاران (۲۰۰۵)
*.۱/۸۲۲	■ ساسون و همکاران (۲۰۱۵)	*.۷۰۹	کول و همکاران (۲۰۰۵)
*.۱/۵۴۵	■ ساسون و همکاران (۲۰۱۵)	.۴۱۲	کول و همکاران (۲۰۰۵)
*.۱/۵۸۹	■ ساسون و همکاران (۲۰۱۵)	*.۶۴۸	کول و همکاران (۲۰۰۵)
*.۸۳۸	لیم (۲۰۰۸)	*.۶/۰۴	کول و همکاران (۲۰۰۵)
*.۹۲۴	لیم (۲۰۰۸)	.۲۳۴	کول و همکاران (۲۰۰۵)
*.۱/۳۷۶	عبادالله‌وند (۱۳۹۵)	*.۴۹۷	کول و همکاران (۲۰۰۵)
*.۱/۲۲۱	عبادالله‌وند (۱۳۹۵)	.۳۳۲	کول و همکاران (۲۰۰۵)
*.۳۲۵	لی و همکاران (۲۰۱۲)	.۰۶	کول و همکاران (۲۰۰۵)
*.۳۴۴	لی و همکاران (۲۰۱۲)	*.۴۹۷	کول و همکاران (۲۰۰۵)
.۱۸	لی و همکاران (۲۰۱۲)	*.۹۵	■ نیکول و همکاران (۲۰۰۱)
.۰۹۸	لی و همکاران (۲۰۱۲)	*.۹۱۴	کراوالیو (۲۰۱۰)
.۰۶۷	لی و همکاران (۲۰۱۲)	*.۱/۲۰۶	کراوالیو (۲۰۱۰)
- .۱/۰۳	لی و همکاران (۲۰۱۲)	*.۱/۳۸۷	کراوالیو (۲۰۱۰)
*.۲/۰۳	کنرادتی و باگنر (۲۰۱۲)	*.۹۱۲	کراوالیو (۲۰۱۰)
*.۱/۳۱	کنرادتی و باگنر (۲۰۱۲)	*.۶۶۲	کراوالیو (۲۰۱۰)
*.۲/۱۳	کنرادتی و باگنر (۲۰۱۲)	*.۵۱	کراوالیو (۲۰۱۰)
*.۱/۲۱	کنرادتی و باگنر (۲۰۱۲)	*.۴۱۲	کراوالیو (۲۰۱۰)
*.۲/۷۷	کنرادتی و باگنر (۲۰۱۲)	*.۴۷۲	کراوالیو (۲۰۱۰)
*.۲/۲۴	کنرادتی و باگنر (۲۰۱۲)	*.۹۵	شانو و همکاران (۲۰۰۱)
*.۳۱	کنرادتی و باگنر (۲۰۱۲)	*.۱/۰۲	شانو و همکاران (۲۰۰۱)
*.۲/۰۳	کنرادتی و باگنر (۲۰۱۲)	*.۹۷۳	شانو و همکاران (۲۰۰۱)
*.۱/۳۷۶	رایس و همکاران (۲۰۰۸)	*.۱/۰۴۵	شانو و همکاران (۲۰۰۱)
*.۱/۱۵۷	رایس و همکاران (۲۰۰۸)	*.۸۸۷	شانو و همکاران (۲۰۰۱)
*.۱/۲۲۱	رایس و همکاران (۲۰۰۸)	*.۷۹۳	شانو و همکاران (۲۰۰۱)

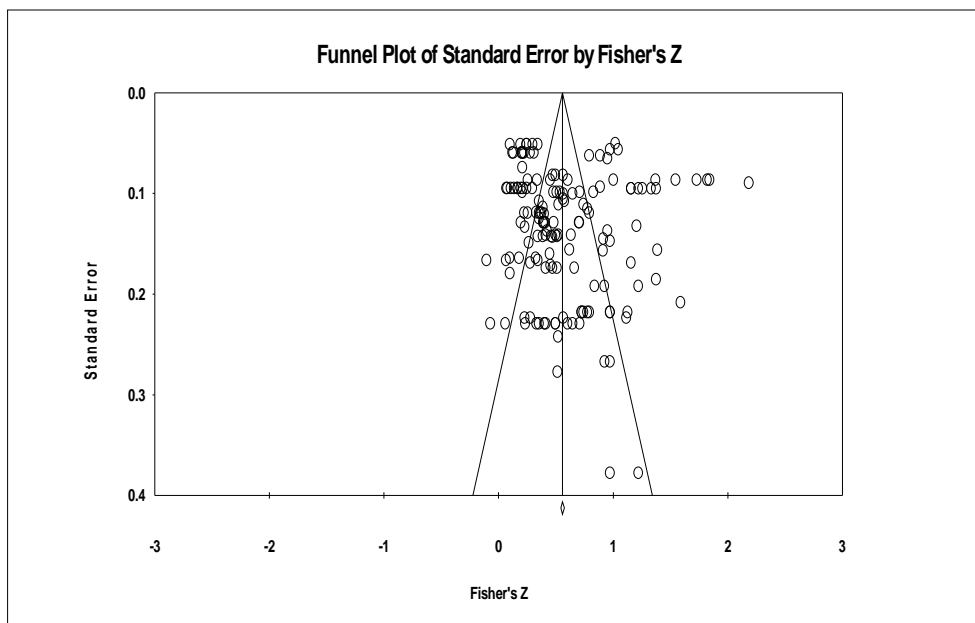
اندازه اثر	مطالعه	اندازه اثر	مطالعه
*۱/۱۵۷	رایس و همکاران (۲۰۰۸)	*/۸۸۷	شانو و همکاران (۲۰۰۱)
*۱/۲۵۶	رایس و همکاران (۲۰۰۸)	*/۵۶۳	شانو و همکاران (۲۰۰۱)
*۱/۳۳۳	رایس و همکاران (۲۰۰۸)	*/۵۷۶	شانو و همکاران (۲۰۰۱)
*/۴۹۷	کلاین و همکاران (۲۰۰۲)	*/۵۱۵	رای و رویا (۲۰۰۲)
*/۵۲۳	کلاین و همکاران (۲۰۰۲)	*/۹۲۷	رای و رویا (۲۰۰۲)
*/۶۳۳	کلاین و همکاران (۲۰۰۲)	*/۹۷۳	رای و رویا (۲۰۰۲)
*/۴۱۲	کلاین و همکاران (۲۰۰۲)	*/۲۰۹	ون و همکاران (۲۰۱۷)
*/۳۸۸	کلاین و همکاران (۲۰۰۲)	*/۴۵۵	لیو و هینچی (۱۹۹۶)
*/۳۴۳	کلاین و همکاران (۲۰۰۲)	/۱	لیو و هینچی (۱۹۹۶)
*/۵۱	کلاین و همکاران (۲۰۰۲)	/۲۳	لیو و هینچی (۱۹۹۶)
*/۴۶	کلاین و همکاران (۲۰۰۲)	*/۶۴۸	اروغلو و کلسیوغلو (۲۰۱۱)
		*/۵۶۳	اروغلو و کلسیوغلو (۲۰۱۱)

■ مطالعاتی که دارای تورش انتشار هستند و در ادامه فراتحلیل از جریان تحلیل خارج شدند  
\* اندازه اثر معنی‌دار

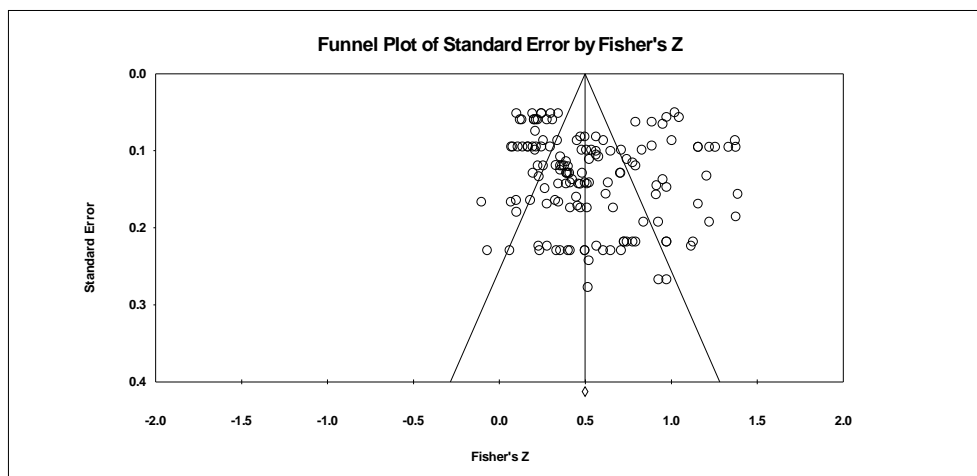
همان‌گونه که در جدول (۲) دیده می‌شود از پژوهش‌های اولیه تعداد ۱۵۳ اندازه اثر به دست آمد. دلیل زیاد بودن اندازه‌های اثر نسبت به تعداد پژوهش‌ها، وجود متغیرهای تعدیل‌کننده مختلف مانند روش برآورد روایی، نوع انجام تکالیف نقشه مفهومی (مثلاً ساخت نقشه یا تکمیل نقشه)، روش‌های مختلف نمره‌دهی نقشه‌های مفهومی (مثلاً رابطه‌ای، ساختاری، شباهت) و ... بود. جدول (۲) اندازه‌های ترکیبی پژوهش‌های مربوط به روایی نقشه‌های مفهومی را پیش از تحلیل حساسیت نشان می‌دهد. جدول (۲) نشان می‌دهد که تعداد ۱۲۴ اندازه اثر معنی‌دار و تعداد ۲۹ اندازه اثر غیر معنی‌دار هستند که اندازه‌های اثر معنی‌دار با علامت \* نشان داده شده‌اند. پس از انجام تحلیل حساسیت مطالعات دارای سوگیری انتشار<sup>۱</sup> (تورش انتشار) از جریان تحلیل خارج شدند (۱۰ اندازه اثر) که با علامت ■ نشان داده شده‌اند.

<sup>۱</sup>. Publication Bias





شکل (۱) نمودار کیفی سوگیری انتشار پیش از انجام تحلیل حساسیت



شکل (۲) نمودار کیفی سوگیری انتشار پس از انجام تحلیل حساسیت

با توجه به اینکه یکی از پیش فرض‌های اصلی فراتحلیل نبود سوگیری انتشار است پیش از بررسی نتایج، ابتدا یافته‌های مربوط به بررسی این پیش فرض ارائه می‌شود.

سوگیری انتشار به چاپ نشدن پژوهش‌های مرتبط با موضوع فراتحلیل مربوط می‌شود که دارای یافته‌های غیر معنادار هستند. در پژوهش حاضر برای بررسی این پیش‌فرض از نمودار کیفی<sup>۱</sup> (شکل ۱ و ۲) استفاده شد. در این نمودار، محور افقی نشان‌دهنده مقادیر اندازه‌های اثر پژوهش‌های اولیه و محور عمودی نشان‌دهنده خطای معیار آنهاست. سوگیری انتشار بر اساس نمودار کیفی زمانی قابل تشخیص است که نقاط در اطراف نمودار به شکل متقارن پراکنده نشده باشند که این امر ناشی از مقادیر بسیار بزرگ اندازه‌های اثر و نیز خطای معیار بزرگ آنهاست (شکل ۱). در شکل (۱) مشخص است که تعدادی از اندازه‌های اثر نامتعرف و پرت هستند و این اندازه‌ها نمودار را نامتقارن کرده‌اند. با حذف تدریجی ۱۰ اندازه اثر، نمودار کیفی شکل (۲) به دست آمد که نسبت به نمودار کیفی شکل (۱) متقارن‌تر است. پس از حذف ۱۰ اندازه اثر اولیه تعداد ۱۴۳ اندازه اثر باقی ماند و در تحلیل‌های بعدی از این تعداد اندازه اثر استفاده شد.

جدول (۳) اندازه‌های اثر ترکیبی اثرات ثابت و تصادفی مربوط به روایی نقشه‌های مفهومی

مقدار P	مقدار Z	فاصله اطمینان ۹۰٪		اندازه اثر ترکیبی	تعداد اندازه اثر	مدل
		حد بالا	حد پایین			
.۰۰۰۱	۵۸/۶۱۴	.۴۷۱	.۴۴۴	.۴۵۸	۱۴۳	ثابت
.۰۰۰۱	۱۷/۲۷۹	.۵۳۱	.۴۳۹	.۴۸۶	۱۴۳	تصادفی

با توجه به اینکه هدف اصلی هر فراتحلیل ترکیب شاخص‌های عددی پژوهش‌های اولیه در قالب یک شاخص کلی است، در جدول (۳) اندازه‌های اثر ترکیبی یا خلاصه بر اساس دو مدل ثابت و تصادفی برای ۱۴۳ اندازه اثر ارائه شده است. بیشتر فراتحلیل‌ها مبتنی بر دو مدل آماری اثر ثابت و اثرات تصادفی هستند. در مدل اثر ثابت فرض می‌شود که یک اندازه اثر واقعی وجود دارد که زیربنای همه تحلیل‌هاست و همه تفاوت‌های اندازه‌های اثر مشاهده شده در پژوهش‌های اولیه ناشی از خطای نمونه‌گیری است. در مقابل، در مدل اثرات تصادفی فرض می‌شود که

<sup>۱</sup>. Funnel Plot

اندازه اثر واقعی از پژوهشی به پژوهش دیگر در حال تغییر است. یکی از علل اصلی این تغییر وجود متغیرهای مداخله‌کننده در رابطه بین متغیر مستقل و متغیر وابسته است (برنشتاین و همکاران، ۲۰۰۹). همان‌گونه که مشاهده می‌شود در این فراتحلیل مقادیر اندازه‌های اثر ترکیبی حاصل از ۱۴۳ اندازه اثر برای مدل ثابت برابر با ۰/۴۵۶ و برای مدل تصادفی برابر با ۰/۴۸۶ است که هر دوی این اندازه‌های اثر از لحاظ آماری معنی‌دار هستند (۰/۰۰۱ < p). برای اینکه مدل نهایی فراتحلیل حاضر مشخص شود باید یک مجموعه تحلیل‌های ناهمگنی برای اطمینان از وجود متغیرهای تعدیل‌کننده انجام گیرد. در صورت وجود ناهمگنی در اندازه‌های اثر پژوهش‌های اولیه مدل تصادفی انتخاب می‌شود و فرض می‌شود که در جامعه آماری ماهیت روابط متغیر مستقل و وابسته، تحت تأثیر متغیرهای تعدیل‌کننده تغییر می‌یابد.

جدول (۴) شاخص‌های ناهمگنی اندازه‌های اثر در پژوهش‌های اولیه

معنی‌داری	درجه آزادی	مجذور I	Q کوکران
۰/۰۰۰۱	۱۴۲	۹۱/۸۳۲	۱۷۳۸/۴۳۳

در جدول (۴) نتایج بررسی ناهمگنی اندازه‌های اثر در پژوهش‌های اولیه بر اساس شاخص Q کوکران ارائه شده است. مقدار شاخص Q برابر با ۱۷۳۸/۴۳۳ است که از لحاظ آماری معنی‌دار است (۰/۰۱ < P) که نشانگر تفاوت واقعی بین اندازه‌های اثر پژوهش‌های اولیه است. با توجه به محدودیت شاخص Q از لحاظ معنی‌داری (هر چه تعداد اندازه‌های اثر بیشتر شود توان آزمون برای رد همگنی بیشتر می‌شود) فراتحلیلگران استفاده از مجذور I را پیشنهاد کرده‌اند (برنشتاین و همکاران، ۲۰۰۹). این شاخص دارای مقداری از صفر تا ۱۰۰ درصد است که مقدار ناهمگنی را به صورت درصد نشان می‌دهد. نتایج مجذور I نشان می‌دهد که بالای ۹۱ درصد از پراکنش موجود در نتایج پژوهش‌های اولیه، واقعی و ناشی از وجود متغیرهای تعدیل‌کننده است که نشان‌دهنده وجود ناهمگنی در پژوهش‌های اولیه است. بر اساس هر دو شاخص ناهمگنی، مشخص شد که متغیرهای تعدیل‌کننده در بررسی روایی نقشه‌های مفهومی تأثیر معنی‌داری دارند بنابراین مدل تصادفی به‌عنوان مدل فراتحلیل انتخاب شد و اندازه اثر ترکیبی همان مقدار ۰/۴۸۶ در نظر گرفته شد.

با مشخص شدن نقش متغیرهای تعدیل‌کننده در ادامه به تحلیل‌های بیشتر در مورد نقش و شدت تعامل متغیرهای تعدیل‌کننده روش برآورد روایی (ملاکی و افتراقی)؛ شیوه انجام تکالیف نقشه مفهومی توسط یادگیرندگان (تکمیل نقشه در مقابل ساخت نقشه)؛ نوع آزمودنی‌های مورد استفاده در پژوهش‌ها (دانش‌آموز یا دانشجو)؛ و شیوه تهیه نقشه‌های مفهومی (کامپیوتری یا مداد-کاغذی) اقدام شد. گفتنی است در این تحلیل‌ها نیز نتایج بر طبق مدل‌های تصادفی گزارش شده است؛ چراکه تحلیل‌های ناهمگنی نشان داد که اندازه‌های اثر داخل خود این متغیرهای تعدیل‌کننده هم ناهمگن هستند. با توجه به وجود ناهمگنی در داخل پژوهش‌های اولیه فقط نتایج مربوط به مدل تصادفی ارائه می‌شود.

جدول (۵) اندازه اثر ترکیبی مدل تصادفی روایی نقشه‌های مفهومی بر اساس روش‌های برآورد

#### روایی

مقدار P	مقدار Z	فاصله اطمینان ۹۵٪		اندازه اثر ترکیبی	تعداد اندازه‌های اثر	روش برآورد روایی
		حد بالا	حد پایین			
۰/۰۰۰۱	۱۶/۱۱۷	۰/۵۲۵	۰/۴۲۸	۰/۴۷۸	۱۳۱	ملاکی
۰/۰۰۰۱	۶/۴۷۵	۰/۶۸۵	۰/۴۲۱	۰/۵۶۷	۱۲	افتراقی

نتایج فراتحلیل در زمینه نقش متغیر تعدیل‌کننده روش برآورد روایی (جدول ۵) نشان داد که اندازه اثر ترکیبی با مدل تصادفی در پژوهش‌هایی که شیوه برآورد روایی نقشه‌های مفهومی از نوع محتوایی بود برابر با ۰/۷۷۵؛ در پژوهش‌هایی که شیوه برآورد روایی از نوع ملاکی بود برابر با ۰/۴۷۸؛ و در پژوهش‌هایی که شیوه برآورد روایی از نوع افتراقی بود برابر با ۰/۵۶۷ بود. بنابراین اندازه اثر ترکیبی در پژوهش‌هایی که شیوه برآورد روایی از نوع محتوایی بود بالاتر از دو شیوه دیگر برآورد روایی بود.

جدول (۶) اندازه اثر ترکیبی مدل تصادفی روایی نقشه‌های مفهومی بر اساس شیوه انجام تکالیف نقشه مفهومی

مقدار P	مقدار Z	فاصله اطمینان ۹۵٪		اندازه اثر ترکیبی	تعداد اندازه‌های اثر	شیوه انجام تکالیف
		حد بالا	حد پایین			
.۰۰۰۱	۱۵/۷۶۵	.۴۹۵	.۳۹۹	.۴۴۸	۱۱۵	ساخت نقشه مفهومی
.۰۰۰۱	۱۲/۱۶۷	.۶۷۹	.۵۳۶	.۶۱۳	۲۸	تکمیل نقشه مفهومی

نتایج فراتحلیل در زمینه نقش متغیر تعدیل‌کننده شیوه انجام تکالیف نقشه مفهومی (جدول ۶) نشان داد که اندازه اثر ترکیبی با مدل تصادفی در پژوهش‌هایی که شیوه انجام تکالیف شامل ساخت نقشه مفهومی توسط یادگیرندگان بود برابر با  $۰/۴۵۱$  و در پژوهش‌هایی که شیوه انجام تکالیف شامل کامل کردن نقشه مفهومی بود برابر با  $۰/۶۲۱$  بود. بنابراین اندازه اثر ترکیبی در پژوهش‌هایی که یادگیرندگان خود اقدام به ساخت نقشه‌های مفهومی می‌کردند پایین‌تر از اندازه اثر ترکیبی در پژوهش‌هایی بود که یادگیرندگان به تکمیل نقشه‌های مفهومی می‌پرداختند.

جدول (۷) اندازه اثر ترکیبی مدل تصادفی روایی نقشه‌های مفهومی بر اساس شیوه تهیه نقشه‌های مفهومی توسط آزمودنی‌ها

مقدار P	مقدار Z	فاصله اطمینان ۹۵٪		اندازه اثر ترکیبی	تعداد اندازه‌های اثر	شیوه تهیه نقشه‌ها
		حد بالا	حد پایین			
.۰۰۰۱	۱۴/۸	.۵۵۷	.۴۴۸	.۵۰۵	۱۰۳	مداد - کاغذی
.۰۰۰۱	۹/۵۱۴	.۵۰۵	.۳۵	.۴۳۱	۳۹	کامپیوتری

نتایج فراتحلیل در زمینه نقشه متغیر تعدیل‌کننده شیوه تهیه نقشه‌های مفهومی (جدول ۷) نشان داد که اندازه اثر ترکیبی با مدل تصادفی در پژوهش‌هایی که شیوه تهیه نقشه‌های مفهومی مدادکاغذی بود برابر با  $۰/۵۰۵$  و در پژوهش‌هایی که شیوه تهیه نقشه‌های مفهومی کامپیوتری بود برابر با  $۰/۴۳۱$  بود. بنابراین اندازه اثر ترکیبی در

پژوهش‌هایی که شیوه تهیه نقشه‌های مفهومی مداد - کاغذی بود بالاتر از پژوهش‌هایی بود که شیوه تهیه نقشه‌های مفهومی کامپیوتری بود.

جدول (۸) اندازه اثر ترکیبی مدل تصادفی روایی نقشه‌های مفهومی بر اساس نوع آزمودنی‌های مورد استفاده در پژوهش‌ها

مقدار P	مقدار Z	فاصله اطمینان ۹۵٪		اندازه اثر ترکیبی	تعداد اندازه‌های اثر	نوع آزمودنی
		حد بالا	حد پایین			
۰/۰۰۱	۱۳/۳۲۲	۰/۵۵۵	۰/۴۳۴	۰/۴۹۷	۸۲	دانش‌آموز
۰/۰۰۱	۱۱/۲۲۱	۰/۵۳۶	۰/۳۹۷	۰/۴۶۹	۶۱	دانشجو

نتایج فراتحلیل در زمینه نقشه متغیر تعدیل‌کننده نوع آزمودنی‌های مورد استفاده در پژوهش‌ها (جدول ۸) نشان داد که اندازه اثر ترکیبی با مدل تصادفی در پژوهش‌هایی که آزمودنی‌ها دانش‌آموز بودند برابر با ۰/۴۹۷، و در پژوهش‌هایی که آزمودنی‌ها دانشجو بودند برابر با ۰/۴۶۹، بود. بنابراین اندازه اثر ترکیبی در پژوهش‌هایی که آزمودنی‌های مورد استفاده دانش‌آموز بودند بالاتر از اندازه اثر ترکیبی در پژوهش‌هایی بود که آزمودنی‌های مورد استفاده دانشجویان بودند.

### بحث و نتیجه‌گیری

هدف از اجرای پژوهش حاضر، فراتحلیل پژوهش‌هایی بود که روایی نقشه‌های مفهومی را بررسی کرده بودند. نتایج پژوهش نشان داد که به‌طور کلی روایی نقشه‌های مفهومی دارای اندازه اثر ترکیبی ۰/۴۸۶، است که بر اساس ملاک کوهن (۱۹۸۸) اندازه اثر بزرگی محسوب می‌شود. بر طبق ملاک کوهن در پژوهش‌های همبستگی، مقدار اندازه‌های اثر ۰/۱ و کمتر از آن کوچک، اندازه‌های اثر ۰/۳ و کمتر از آن متوسط، و اندازه‌های اثر ۰/۵ و بیشتر از آن بزرگ به حساب می‌آیند. همچنین در مورد نقشه متغیرهای تعدیل‌کننده نیز نتایج نشان داد که اندازه اثر ترکیبی در پژوهش‌هایی که روش برآورد روایی نقشه‌های مفهومی از نوع افتراقی بود برابر با ۰/۵۶۷؛ و در پژوهش‌هایی که روش برآورد روایی از نوع ملاکی بود برابر با ۰/۴۷۸، بود. میزان اندازه اثر ترکیبی در پژوهش‌هایی که یادگیرندگان، نقشه‌های مفهومی را تکمیل می‌کردند

برابر با ۶۱۳. و در پژوهش‌هایی که یادگیرندگان به ساخت نقشه‌های مفهومی اقدام می‌کردند برابر با ۴۴۸. بود. میزان اندازه اثر ترکیبی در پژوهش‌هایی که شیوه تهیه نقشه‌های مفهومی مدادکاغذی بود برابر با ۵۰۵. و در پژوهش‌هایی که شیوه تهیه نقشه‌های مفهومی کامپیوتری بود برابر با ۴۳۱. بود. همچنین اندازه اثر ترکیبی در پژوهش‌هایی که آزمودنی‌ها دانش‌آموز بودند برابر با ۴۹۷. و در پژوهش‌هایی که آزمودنی‌ها دانشجو بودند برابر با ۴۶۹. بود. بنابراین تمام نتایج فراتحلیل حاضر نشان دادند که نقشه‌های مفهومی از روایی مناسبی برخوردار بوده و توانایی استفاده شدن به‌عنوان ابزار ارزیابی را دارا می‌باشند.

در مورد روایی افتراقی، نتایج پژوهش‌ها نشان می‌دهد که نقشه‌های مفهومی ابزارهای قابل اعتمادی هستند. برای مثال وست، پارک، پومروی و ساندوال<sup>۱</sup> (۲۰۰۲) در پژوهش خود نشان دادند که آزمون‌های مبتنی بر نقشه مفهومی می‌توانند تفاوت‌های مورد انتظار بین آزمودنی‌ها از لحاظ سطوح متفاوت آموزش را به‌درستی پس از یک مداخله آموزشی تشخیص دهند. برای این منظور، آنها امتیازهای نقشه مفهومی یادگیرندگان باتجربه‌تر (دانشجویان سال‌های دوم و سوم) را با امتیازهای نقشه مفهومی یادگیرندگان کم‌تجربه‌تر (دانشجویان سال اول) با استفاده از یک آزمون ویلکاکسون مقایسه کردند. نتایج آنها نشان داد که امتیازهای نقشه‌های مفهومی یادگیرندگان باتجربه‌تر به‌صورت معنی‌داری بالاتر از امتیازهای نقشه‌های مفهومی یادگیرندگان کم‌تجربه‌تر بود. البته روش نمره‌دهی نقشه‌های مفهومی در این رابطه تأثیرگذار بود؛ چراکه این معنی‌داری در روش نمره‌دهی ساختاری وجود داشت در حالی که در روش نمره‌دهی رابطه‌ای چنین تفاوت معنی‌داری وجود نداشت.

در رابطه با روایی ملاکی نقشه‌های مفهومی نیز پژوهشگران اظهار داشته‌اند که امتیازهای ارزیابی‌های مبتنی بر نقشه مفهومی هرگز به‌صورت کامل با ارزیابی‌های سنتی هم‌بسته نیست زیرا آزمون‌های مبتنی بر نقشه مفهومی به‌گونه‌ای مستقیم‌تر درک ساختاری و سازمان‌یافته شاگردان را اندازه می‌گیرند (نواک و گوین، ۱۹۸۴). رایس و همکاران (۱۹۹۸) نیز اظهار داشتند که روش امتیازدهی نقشه‌های مفهومی - به‌ویژه معیارهای مختص به امتیازدهی - تعیین‌کننده مهم نیرومندی رابطه بین امتیازهای نقشه مفهومی و امتیازهای دیگر ابزارهای ارزیابی است. برای مثال، زمانی که سطوح سلسله

1. West, Park, Pomeroy, & Sandoval

مراتب از معیارهای امتیازدهی نقشه‌های مفهومی حذف می‌شود همبستگی بین امتیازهای نقشه‌های مفهومی و آزمون‌های سنتی به شکل دراماتیکی افزایش می‌یابد. در تبیین روایی مناسب نقشه‌های مفهومی می‌توان از پژوهش وست و همکاران (۲۰۰۲) استفاده کرد. این پژوهشگران به‌منظور برآورد روایی پیش‌بینی (یکی از انواع روایی ملاکی) امتیازات نقشه‌های مفهومی را پیش و پس از یک مداخله آموزشی مقایسه کردند. مقایسه کیفی نقشه‌های مفهومی پیش و پس از مداخله به‌خوبی تلفیق مفاهیم جدید به داخل نقشه‌ها، تمایز بیشتر مفاهیم موجود و روابط عرضی (همبرها) جدید را نشان داد. نتایج کمی پژوهش نیز نشان داد که امتیازهای نقشه‌های مفهومی پس از آموزش به‌صورت معنی‌داری افزایش یافت هر چند که باز هم این معنی‌داری در روش نمره‌دهی ساختاری دیده شد. این نتایج نشان‌دهنده روایی ملاکی مناسب نقشه‌های مفهومی است.

مسئله‌ای که در رابطه با روایی ملاکی مطرح است پیدا کردن مقیاس دیگری (مقیاس ملاک یا معیار) است که بتواند به‌صورتی قابل اطمینان نمره‌هایی را تولید کند و سازه یا صفت یکسانی را اندازه بگیرد که هدف ارزیابی است. اگر نمره‌های مقیاس معیار، روا و معتبر باشد در این صورت این پرسش مطرح می‌شود که چرا خود مقیاس ملاک به جای مقیاسی که در حال آزمون شدن است (مثلاً ارزیابی مبتنی بر نقشه مفهومی) مورد استفاده قرار نمی‌گیرد؟ در رابطه با امتیازهای ارزیابی‌های مبتنی بر نقشه مفهومی باید اظهار داشت که مقیاس‌های ملاکی که برای برآورد روایی مورد استفاده قرار می‌گیرند یا نمره‌دهی آنها خیلی وقت گیر است (مانند مصاحبه‌ها) یا زیر سایه دیگر سازه‌ها قرار می‌گیرند (مانند توانایی نوشتن در استفاده از آزمون‌های تشریحی) یا اینکه چارچوب‌های مفهومی دانش‌آموزان را به‌خوبی نشان نمی‌دهند (مانند آزمون‌های چندگزینه‌ای). بنابراین ارزیابی‌های مبتنی بر نقشه مفهومی نسبت به دیگر ارزیابی‌ها برتری محسوسی دارند.

در مجموع، همان‌گونه که مک‌کلور و همکاران (۱۹۹۹) اظهار داشتند به‌منظور بهینه ساختن کاربرد نقشه‌های مفهومی به‌عنوان ابزارهای ارزیابی، معلمان باید تکالیف نقشه مفهومی را یک ارزیابی اولیه در نظر گیرند که می‌تواند حوزه‌هایی که به تحلیل عمیق تری نیاز دارند را نشان دهد. برای نمونه، نمره کم در یک نقشه مفهومی می‌تواند به ما هشدار دهد تا به‌منظور تشخیص کج‌فهمی‌های شاگردان نگاه دقیق‌تری به موضوع داشته باشیم. به‌علاوه، زمانی که مجموعه‌ای از نقشه‌ها نمره‌گذاری می‌شوند، معلم باید



مراقب بازپیدایی الگوهای در نقشه‌های مفهومی شاگردان باشد زیرا ممکن است نشان‌دهنده مسائلی در زمینه نحوه آموزش باشند. با این شیوه، ارزیابی نقشه‌های مفهومی شاگردان ممکن است به معلمان اجازه دهد تا تعادل بین هزینه‌ها و فایده‌های تکالیف ارزیابی نقشه‌های مفهومی را بهینه سازند.

بر اساس یافته‌های پژوهش حاضر می‌توان به برنامه‌ریزان و مسئولان امور آموزشی در کشور پیشنهاد کرد که:

با توجه به روایی مطلوب نقشه‌های مفهومی معلمان، مربیان، استادان و مسئولان امور ارزشیابی کشور می‌توانند از نقشه‌های مفهومی به‌عنوان یکی از روش‌های ارزیابی جانشین برای ارزیابی هر چه بهتر فرایند یاددهی - یادگیری استفاده کنند. همچنین می‌توانند در سطوح مختلف تحصیلی از نقشه‌های مفهومی به‌عنوان ابزارهای ارزشیابی آغازین، تکوینی (جهت پی بردن به کج‌فهمی‌ها و نقاط قوت و ضعف یادگیری شاگردان) و تراکمی بهره ببرند.

از جمله محدودیت‌های پژوهش حاضر نیز می‌توان به موارد زیر اشاره کرد:

این احتمال وجود داشت که پژوهش‌های غیر معنی‌دار در ارتباط با روایی نقشه‌های مفهومی کمتر گزارش شده باشند. پژوهش‌های زیادی از نقشه‌های مفهومی به‌عنوان ابزارهای ارزیابی استفاده کرده بودند اما از ارائه شواهد روایی نقشه‌ها خودداری کرده بودند. دسترسی به تعدادی از پژوهش‌ها با توجه به چارچوب نمونه‌گیری میسر نبود. در بعضی موارد اطلاعات مربوط به نوع تکالیف نقشه مفهومی، سال و مکان اجرای پژوهش‌ها ناقص بود. همچنین در این فرا تحلیل صرفاً به داده‌های کمی پژوهش‌ها برای انجام فرا تحلیل بسنده شد بنابراین، دقت و صحت این داده‌ها بر عهده مؤلفان پژوهش‌های اولیه است.

## منابع

- ایزائلو، 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 - ۸۲.
- رعایی، فرزانه؛ دلاور، علی و فرخی، نورعلی (۱۳۸۹). فراتحلیل تحقیقات انجام شده در حوزه اضطراب و افسردگی. *فصلنامه اندازه‌گیری تربیتی*، ۱ (۴)، ۱ - ۲۵.
- سیف، علی‌اکبر (۱۳۸۹). *روان‌شناسی پرورشی نوین: روان‌شناسی یادگیری و آموزش*، ویراست ششم. تهران: نشر دوران.
- عبادالله‌وند، کوثر (۱۳۹۵). *بررسی روایی و پایایی آزمون‌های تکمیل مبتنی بر نقشه مفهومی در ماده درسی علوم تجربی پایه ششم*. پایان‌نامه کارشناسی ارشد، دانشگاه شهید مدنی آذربایجان.
- مصراآبادی، جواد (۱۳۹۰). *بررسی روایی و پایایی شیوه‌های نمره‌گذاری آزمون‌های مبتنی بر نقشه مفهومی*. *فصلنامه اندازه‌گیری تربیتی*، ۷ (۲)، ۹ - ۳۲.
- مصراآبادی، جواد (۱۳۸۸). *امکان‌سنجی به کارگیری روش‌های ارزشیابی مبتنی بر نقشه مفهومی در زیست‌شناسی دوره متوسطه*. وزارت آموزش و پرورش، سازمان پژوهش و برنامه‌ریزی درسی، مؤسسه پژوهشی برنامه‌ریزی درسی و نوآوری‌های آموزشی.
- هومن، حیدرعلی (۱۳۹۰). *شناخت روش علمی در علوم رفتاری*. تهران: سازمان مطالعه و تدوین کتب علوم انسانی دانشگاه‌ها (سمت)، مرکز تحقیق و توسعه علوم انسانی.

- Arneson, B. T. (2005). *On the role of concept mapping assessment in today's constructivist classroom*. Doctoral dissertation, The University of Texas at Austin.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd, UK.
- Chang, K.-E.; Sung, Y.-T.; Chang, R.-B. & Lin, S.-C. (2005). A New Assessment for Computer-based Concept Mapping. *Educational Technology & Society*, 8 (3), 138-148.
- Ciprian Sas, I. (1999). *The multiple-choice concept map (MCCM): an interactive computer-based assessment method*. Doctoral dissertation, University of Nevada, Las Vegas.

- Clariana, R.B. (2009). A comparison of pair-wise, list-wise, and clustering approaches for eliciting structural knowledge. *International Journal of Instructional Media*, 36 (3), 287-302.
- Cline, B. E.; Brewster, C. C. & Fell, R.D. (2010). A rule-based system for automatically evaluating student concept maps. *Expert Systems with Applications*. 37, 2282–2291.
- Conradty, C. & Bogner, F. X. (2012) Knowledge presented in concept maps: correlations with conventional cognitive knowledge tests. *Educational Studies*, 38 (3), 341-354
- Coul, R.; Clariana, R. B. & Salehi, R. (2005). Comparing several human and computer-based methods for scoring concept maps and essays. *Journal of Educational Computing Research*, 32 (3), 227-239.
- Cravalho, P. F. (2010). *Learning Statistics using Concept Maps: Effects on Anxiety and Performance*. Master dissertation, San Jose State University.
- Ekin, B.; Ulcinar Sagir, S. & Saltan, F. (2016). The Comparison on Evaluation of Concept Map and Structured Grid with Multiple-Choice Test. *Participatory Educational Research (PER)*, Special Issue 2016-II; 100-111.
- Eroglu, M. G. & Kelecioglu, H. K. (2011). An analysis on the validity and reliability of concept map and structural communication grid scores. *Journal of Education*, 40, 210 - 220.
- Hung, C. H. & Lin, C. U. (2015). Using concept mapping to evaluate knowledge structure in problem-based learning. *BMC Medical Education*, 15, 212. DOI 10.1186/s12909-015-0496-x.
- İngeç, S. K. (2009). Analysing Concept Maps as an Assessment Tool in Teaching Physics and Comparison with the Achievement Tests. *International Journal of Science Education*, 31 (14), 1897-1915.
- Jonsson, A. & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2, 130–144.
- Kassab, S. E. & Hussain, S. (2010). Concept mapping assessment in a problem-based medical curriculum. *Journal of Medical Teacher*. 32, 926 – 931.
- Kaya, O. N. & Kilic, Z. (2004). Student-Centered Reliability, Concurrent Validity and Instructional Sensitivity in Scoring of Students' Concept Maps in a University Science Laboratory. Poster presented at 18<sup>th</sup> International Conference on Chemical, İstanbul, Turkey.

- Klein, D. C. D.; Chung, K. W. K.; Osmundson, E. & Herl, H. E. (2002). *The validity of knowledge mapping as a measure of elementary student's scientific understanding*. CRESST/University of California, Los Angeles.
- Ley, S. L.; Krabbe, H. & Fischer, H. E. (2012). Convergent validity: concept maps and test for student's diagnosis in physics. *Fifth International Conference on Concept Mapping Valletta, Malta*
- Liping, Z.; Yan, W.; Boqing, D. & Zengyao, Z. (2009). The comparison study of Chinese and American secondary school students' knowledge structure - an experimental research based on concept map assessment technique. *Front. Educ China*, 4 (2), 286–297. DOI 10.1007/s11516-009-0015-0.
- Lim, Y. K. (2008). *The effect of concept mapping with different levels of generativity and learners self-regulated learning skills on knowledge acquisition and representation*. Doctoral dissertation, The Pennsylvania State University.
- Liu, J. (2013). The Assessment Agent System: design, development, and evaluation. *Education Tech Research Dev.* 61,197–215. DOI 10.1007/s11423-013-9286-5.
- Liu, X. & Hinchey, M. (1996). The internal consistency of a concept mapping scoring scheme and its effect on prediction validity. *International Journal of Science Education*, 18 (8), 921 - 937.
- Liu, T. C., Lin, Y. C., & Tsai, C. C. (2009). Identifying senior high school student's misconceptions about statistical correlation, and their possible causes: an exploratory study using concept mapping with interviews. *International Journal of Science & Mathematics Education*. 7, 791 – 820.
- McClure, J. R., Sonak, B., & Suen, H. K. (1999). Concept map assessment of classroom learning: Reliability, validity, and logistical practicality. *Journal of Research in Science Teaching*, 36 (4), 475–492.
- Nicoll, G.; Francisco, J. & Nakhleh, M. (2001). A three-tier system for assessing concept map links: a methodological study. *International Journal of Science Education*, 8, 863 – 875.
- Novak, J. D. & Gowin, B. (1984). *Learning how to learn*. New York: Cambridge University Press.
- Plummer, K. J. (2008). *Analysis of the psychometric properties of two different concept-map assessment tasks*. Doctoral dissertation, Brigham Young University.

- Reiska, P.; Soika, K.; Mollits, A.; Rannikmae, M. & Soobard, R. (2015). Using Concept Mapping Method for Assessing Students' Scientific Literacy. *Procedia - Social & Behavioral Sciences*, 177, 352 – 357.
- Rice, D. C.; Ryan, J. M. & Samson, S. M. (1998). Using concept maps to assess student learning in the sciences classroom: Must different methods compete? *Journal of Research in Science Teaching*, 35 (10), 1103-1127.
- Ruiz-Primo, M. A. (2000). On the Use of Concept Maps as an Assessment Tool in Science: What We Have Learned so Far. *Revista Electrónica de Investigación Educativa*, 2 (1), 2000-2030.
- Ruiz-Primo, M. A.; Schultz, S. E.; Li, M. & Shavelson, R. J. (2001). Comparison of the reliability and validity of scores from two concept-mapping techniques. *Journal of Research in Science Education*, 38 (2), 260–278.
- Ruiz-Primo, M. A. & Shavelson, R. J. (1996). Problems and issues in use of concept maps in science assessment. *Journal of Research in Science Teaching*, 33, 569-600.
- Rye, J. & Rubba, P. A. (2002). Scoring Concept Maps: An Expert Map-Based Scheme Weighted for Relationships. *Journal of School Science & Mathematics*, 102 (1).
- Sasson, E.; Ravid, G. & Pliskin, N. (2015). Improving similarity measures of relatedness proximity: Toward augmented concept maps. *Journal of Informetrics*, 9, 618 – 628.
- Schau, C.; Mattern, N.; Zeilik, M.; Teague, K. W. & Weber, R. J. (2001). Select-and-Fill-in Concept Map Scores as a Measure of Students' Connected Understanding of Science. *Educational & Psychological Measurement*, 61, 136.
- Schwendimann, B. A. (2015). Concept maps as versatile tools to integrate complex ideas: From kindergarten to higher and professional education. *Knowledge Management & E-Learning: Special Issue on Novakian Concept Mapping in University & Professional Education*, 7 (1), 73-99.
- Stoddart, T, Abrams, R., Gasper, E. & Canaday, D. (2000). Concept maps as assessment in science inquiry learning - a report of methodology. *International Journal of Science Education*, 22, 1221 – 1246.
- Strautmane, M. (2012). Concept map-based knowledge assessment tasks and their scoring criteria: an overview. *Concept Maps:*

- Theory, Methodology, Technology Proc. of the Fifth Int. Conference on Concept Mapping Valletta, Malta 2012.
- Tsai, C. C.; Lin, S. S. J. & Yuan, S. M. (2001). Students' use of web-based concept map testing and strategies for learning. *Journal of Computer Assisted Learning*, 17, 72 – 84.
- Tzeng, J. Y. (2014). Mapping for depth and variety: using a “Six W’s” scaffold to facilitate concept mapping for different history concepts with different degrees of freedom. *Educational Studies*, 40 (3), 253-276, DOI: 10.1080/03055698.2014.889595.
- Vodovozov, V. & Raud, Z. (2015). Concept Maps for Teaching, Learning, and Assessment in Electronics. *Hindawi Publishing Corporation Education Research International*, Volume 2015, Article ID 849678, 9 pages. <http://dx.doi.org/10.1155/2015/849678>.
- Watson, M. K.; Pelkey, J.; Noyes, C. & Rodgers, M. (2016). Assessing impacts of a learning-cycle-based module on students' conceptual sustainability knowledge using concept maps and surveys. *Journal of Cleaner Production*, 133, 544 – 556.
- West, D. C.; Park, J. K.; Pomeroy, J. R. & Sandoval, J. (2002). Concept mapping assessment in medical education: A comparison of two scoring systems. *Medical Education*, 36, 820-826.
- Won, M.; Krabbe, H.; Siv Ling Ley, S. L.; Treagust, D. F. & Fisher, H. E. (2017). Science Teachers' Use of a Concept Map Marking Guide as a Formative Assessment Tool for the Concept of Energy. *Educational Assessment*, 22 (2), 95 - 110.
- Wu, P. H.; Hwang, G. J.; Milrad, M.; Ke, H. R. & Huang, Y.M. (2012). An innovative concept map approach for improving students' learning performance with an instant feedback mechanism. *British Journal of Educational Technology*, 2, 217 – 232.
- Yin, Y. & Shavelson, R. J. (2008). Application of generalizability theory to concept map assessment research. *Applied Measurement in Education*, 21, 273-291.
- Yin, Y.; Vanides, J.; Ruiz-Primo, M. A.; Ayala, C. C. & Shavelson, R. J. (2005). Comparison of Two Concept-Mapping Techniques: Implications for Scoring, Interpretation, and Use. *Journal of Research in Science Teaching*, 42 (2), 166–184.