نوع مقاله: پژوهشی
تاریخ دریافت: ۱۳۹۸/۰۷/۲۳ تاریخ پذیرش: ۱۳۹۸/۰۹/۱۷

بررسی روایی سازه مرکز ارزیابی و توسعه مدیران مورد مطالعه شرکت ملی نفت ایران

مجید سلیمی[۱] - علی دلاور[۲] - نورعلی فرخی[۳] - حسین شکرکن[۴] - عباس عباس‌پور[۵]

چکیده

هدف این مطالعه، بررسی روایی سازه مرکز ارزیابی و توسعه مدیران در شرکت ملی نفت ایران بود. روش پژوهش به لحاظ هدف، کاربردی و از نظر جمع‌آوری داده‌ها، توصیفی- تحلیلی به شمار می‌رود. نمونه آماری مشتمل بر داده‌های حاصل از ارزیابی ۱۰۲۲ مدیر بود که از بانک اطلاعات مدیران به روش تصادفی ساده انتخاب شدند. این داده‌ها بر اساس روش چندصفتی-چندروشی و تحلیل عاملی تأییدی مورد تحلیل قرار گرفت. نتایج نشان داد میانگین کل هم‌بستگی بین درجه‌بندی‌های مربوط به یک بُعدِ واحد (شایستگی) با روش‌های مختلف (۰،۳۲) در مقایسه با میانگین کل هم‌بستگی بین درجه‌بندی‌های ابعاد مختلف در همان روش (۰،۶۰)، کمتر است. بنابراین، هرچند درجه‌بندی ابعاد از اعتبار همگرایی نسبتاً متوسطی برخوردار بود، اعتبار تشخیصیِ رضایت بخشی نشان نداد. از سوی دیگر نتایج تحلیل عاملی نشان داد، سهم ابعاد (شایستگی‌ها) در توصیف تغییرات درجه‌بندی‌های صورت گرفته به مراتب بیش‌تر از سهم روش‌های ارزیابی است و بر این اساس می‌توان ادعا نمود که روایی تشخیصی بین ابعاد مورد مطالعه، تعیین‌کننده‌ترین عامل اصلی درجه‌بندی‌های مرکز ارزیابی می‌باشد.

واژگان کلیدی: مرکز ارزیابی، روایی سازه، روایی یابی چندصفتی-چندروشی، روایی همگرا/تشخیصی، تحلیل عاملی تأییدی.

[۱] دانشجوی دکتری سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران.نویسنده مسئول:(salimy28@yahoo.com)
[۲] استاد تمام گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران.Delavarali@yahoo.com
[۳] دانشیار گروه سنجش و اندازه‌گیری، دانشگاه علامه طباطبائی، تهران، ایران. farrokhinoorali@yahoo.com
[۴] استاد بازنشسته، دانشگاه شهید چمران اهواز، اهواز، ایران. shokrkon_h@yahoo.com
[۵] دانشیار گروه برنامه‌ریزی آموزشی، دانشگاه علامه طباطبائی، تهران، ایران. abbaspour@atu.ac.ir

[Downloaded from iieshrm.ir on 2026-11-05]

مقدمه

اساس دنیای جدیدِ کار، رقابت بی‌امان بین‌المللی و اقتصاد "دانش بنیان" نسبت به اقتصاد "فرآورده محور" است. فناوری‌هایی که با شتاب دگرگون می‌شوند، به افزایش تصاعدی اطلاعات موجود، می‌انجامد. برای برآوردن این نیازها، ساختارهای سازمانی، تخت شده، خط گزارش‌دهی مدیریتی کاهش یافته و سطح مسئولیت، استقلال و نیز گونه گونی مهارت کارکنان افزایش یافته است (Wood, Hoffman, &Twenge, 2012). بر این اساس، امروزه رقابت شدید و تغییرات فناوری، فشار روزافزونی بر سازمان‌ها و شیوه‌های متنوع آن‌ها در افزایش بهره‌وری نیروی انسانی‌شان وارد می‌سازد. در این میان عملکرد مدیران به طور فزاینده‌ای برای موفقیت سازمان‌ها در دنیای بدون مرز، حیاتی شناخته می‌شود (Jokinen, 2005). ازجمله رویکردهای سازمان‌ها جهت اطمینان از عملکرد مدیران، رویکرد شایستگی محور و به کارگیری آن در برنامه‌های انتخاب و توسعه مدیران است. اجزای اصلی یک سیستم مبتنی بر شایستگی عبارتند از: شناسایی معیارهای شایستگی، تدوین الگوهای شایستگی و ارزیابی شایستگی مدیران بر اساس الگوهای تدوین شده (Draganidis & Mentaz, 2006).

فناوری‌ای که به کمک آن رویکرد شایستگی محور اجرایی می‌گردد و الزام‌های مربوط به آن پاسخ داده می‌شود، فرایندی به نام مرکز ارزیابی است. فرایند مرکز ارزیابی با استفاده از شیوه‌های نو و هیجان انگیز، به سازمان‌ها کمک می‌کند تا چشم انداز دگرگون شونده کسب و کار را بپیمایند (Povah & Thornton, 2011).

درحال حاضر، مراکز ارزیابی جهت ارزیابی کارکنان و به طور منظم برای ارزیابی شایستگی مدیران در سازمان‌های مختلف خصوصی و عمومی هرساله مورداستفاده قرار می گیرد (Thornton, Byham, Arthur, & Day, 2003; McNelly, Edens, 1982). این مراکز در طول دهه‌ها به‌عنوان ابزار گزینش و آموزشِ عمومی آن گونه که فرصت‌های شغلی را در محیطی شبیه به محیط واقعیِ کار ارائه می‌دهند، مطرح بوده است (Gaugler, 1987 Rosenthal, Thornton, & Bentson 1982; Thornton, & Byham,).

[Downloaded from iieshrm.ir on 2026-11-05]

مراکز ارزیابی برای بررسی وضعیت فرد در تعدادی از سازه‌ها توسط مجموعه‌ای از تمرین‌ها در موقعیت‌های شبیه‌سازی شده شغلی، مفهوم‌پردازی شده‌اند (Cappelli & Keller, 2014). به بیان دیگر، مراکز ارزیابی از چندین ارزیابِ آموزش دیده که رفتار داوطلبان را در شبیه‌سازی فعالیت‌های مهم کاری، مشاهده می‌کنند، بهره می‌گیرند. ارزیابان، عملکرد شرکت کنندگان را در سازه‌های محوری از قبیل ویژگی‌ها و تکالیف ارزشیابی می‌کنند و ممکن است برای بهبود، توصیه‌هایی به عمل آورند. مرکز ارزیابی، ابزاری اثربخش برای کارمندیابی، انتخاب، برنامه‌ریزی جانشینی و بالندگی سرمایه‌های انسانی در سراسر لایه‌های سازمانی است (Thornton, Rupp, & Hoffman, 2015).

از شایستگی‌هایی که به طور معمول در این مراکز مورد ارزیابی قرار می‌گیرد، می‌توان به توانایی تصمیم‌گیری، حل مسئله، تفکر تحلیلی، ارتباطات و ... اشاره کرد که با استفاده از تکنیک‌های مختلف مانند: تمرین گروهی، تمرین نوشتاری، بحث گروهی، کارتابل، کازیه، مصاحبه و ... مورد ارزیابی قرار می‌گیرد (Greyling,Visser, & Fourie, 2003). غالباً هر بخش از فرایند ارزیابی و هم‌چنین ارزیابی داوطلبانِ مختلف توسط ارزیابان متفاوتی انجام می‌شود؛ از این رو فرایند انجام یافته در مراکز ارزیابی را می‌توان فرایندی چندروشی، چندصفتی و چندارزیاب نامید که به طور معمول در آن، ارزیابان برای ارزیابی و احصاء شایستگی‌ها و ابعاد هدف، تحت آموزش و کارورزی قرار می‌گیرند (Thornton & Rupp, 2006). وجود حداقل یک تمرین شبیه‌سازی مرتبط با شغل برای استخراج توانایی‌های عملکردی داوطلبان ضروری است. از سوی دیگر، مراکز ارزیابی می‌توانند براساس رفتارهای مورد هدف و میزان ارزش رفتارهای مشخص، کاملاً از هم متفاوت باشند (Joiner, 2000). درحالی که دیگر اَشکال ارزیابی در پی دستیابی به عملکرد حاضر فرد هستند، مزیت مراکز ارزیابی، توانایی شناساییِ پتانسیل و ظرفیت‌های داوطلبان به ویژه توانایی مدیریتی آنهاست(Charoux, 1991).

ارزیابی و سنجش شایستگی‌های مدیران، اقدامی مهم برای سازمان‌هایی است که در محیطی بین‌المللی فعالیت می‌کنند. باوجود آنکه مجموعه‌ای از ابزارها مانند: پرسش‌نامه‌ها، آزمون‌ها، مصاحبه و ... برای ارزیابی شایستگی‌های مدیریتی مؤثر هستند، اما روایی پیش

[Downloaded from iieshrm.ir on 2026-11-05]

بین و سازه آنها، مورد سؤال است (Bird & Steven, 2013). در مقابل، مراکز ارزیابی که کلیه مؤلفه‌های شایستگی، از جمله رفتارها و مهارت‌های خاص شغلی و هم‌چنین شخصیت و ابعاد نگرشی را اندازه‌گیری می‌کند، برای ارزیابی صلاحیت‌های مدیریتی مناسب هستند. (Povah & Thornton, 2011) اما سؤال اساسی با توجه به افزایش استفاده از مراکز ارزیابی، دانستن این مطلب است که مراکز ارزیابی تا چه حد دارای روایی است (Herd, Alagaraja, & Cumberland, 2016). این مسئله، موضوعی است که سال هاست ذهن محققان حوزه مرکز ارزیابی را به خود مشغول داشته است. مطابق این امر، در دهه‌های اخیر پژوهش‌های متعددی برای ارزیابی اثربخشی این فرایند در کشورهای مختلف انجام گرفته است (Hurst & Charoux, 1994; Nako, 2011). از پرسش‌های اساسی موجود در ادبیات پژوهشی مربوط به حوزه مرکز ارزیابی، می‌توان به روایی نتایج حاصل به ویژه روایی سازه آن‌ها اشاره کرد (Nako, 2011; Theron, 2009).

از آنجایی که استفاده از مراکز ارزیابی در ایران نیز دارای چندان قدمتی نبوده و افراد فعال در این مراکز نیز از بینشِ روان سنجی کافی برای استانداردسازی ابزارها برخوردار نبوده‌اند، تاکنون سؤال‌های جدّی در مورد روایی نتایج حاصل از مرکز ارزیابی از نظر فنی و سازمانی پرسیده نشده است. به تبع این سکوت می‌توان به جرأت گفت که تقریباً پژوهش مدوّن اندکی در این خصوص در ادبیات کشور وجود دارد. لذا پرداختن به این موضوع یک ضرورتِ اجتناب ناپذیر به نظر می‌رسد. مرکز ارزیابی و توسعه مدیرانِ شرکت ملی نفت نیز در راستای تصمیم‌سازی در حوزه شایسته گزینی و انتصاب مدیران و اجرای طرح‌های استعدادیابی، پس از طراحی مدل شایستگی‌های عمومی در چهارده بُعد (هم‌چون: نوآوری، رهبری، ارتباطات، پایش عملکرد، برنامه‌ریزی، تفکر استراتژیک و ...) این شایستگی‌ها را در فرایندی دو روزه در مرکز ارزیابی و توسعه مدیران و با ابزارهایی نظیر: آزمون‌های روان شناختی، مصاحبه، تمرین گروهی، بحث گروهی، موردکاوی و پرسش‌نامه مورد ارزیابی قرار می‌دهد و پس از جمع‌بندی و توافق در نمرات شایستگی‌ها، این امتیازات در بانک اطلاعات مدیرانِ شرکت ملی نفت ثبت می‌شود. بر این اساس هدف اصلی این پژوهش، ارزیابی و سنجش روایی سازه مرکز ارزیابی شرکت ملی نفت ایران است و اهداف فرعی این پژوهش

[Downloaded from iieshrm.ir on 2026-11-05]

عبارتند از: تعیین ابعادی که مرکز ارزیابی به منظور سنجیدن آن‌ها طراحی شده است.

بی شک استفاده از نتایج حاصل از درجه‌بندی و امتیازهای مراکز ارزیابی هنگامی قابل قبول است که داده‌های گردآوری شده پایایی و روایی مناسبی داشته باشند. بدون اطلاع از پایایی و روایی نمی‌توان در مورد نتایج حاصل از مراکز ارزیابی ادعایی کرد. هرچند مراکز ارزیابی به طورمعمول روایی محتوا و نیز روایی مرتبط با ملاک برای پیش‌بینی عملکرد را نشان می‌دهند (Meriac, Hoffman, Woehr, & Fleisher, 2008). روایی سازه این مراکز هم‌چنان یک مسئله حل نشده باقی مانده است و شواهد مرتبط با روایی سازه ابعاد مراکز ارزیابی کمتر امیدوارکننده است، به طوری که فقدان روایی همگرا و هم‌چنین نبود روایی تشخیصی این مراکز به طور گسترده‌ای در ادبیات پژوهشی گزارش شده است (Klimoski & Brickner, 1987; Michaels, Baker, & Brannick, 1989; Sackett & Harris, 1988). برای نمونه، بای سیو و همکاران (۱۹۸۷) در یک مطالعه بر روی ۱۱۷۰ آزمودنی در یک مرکز ارزیابی که آزمودنی‌ها بر اساس هشت بُعد (برنامه‌ریزی، تحلیل، تصمیم‌گیری، کنترل، ارتباطات شفاهی، روابط بین فردی، انعطاف‌پذیری و تأثیرگذاری) و در پنج تمرین (تمرین گروهی حل مسئله، کازیه، ایفای نقش، تمرین گروهیِ روابط انسانی و مصاحبه) درجه‌بندی شدند. با استفاده از تحلیل عاملی تأییدی بر روی ماتریس چندصفتی-چندروشی، پنج مدل مختلف -بر حسب تعداد عوامل- را مورد آزمون قرار دادند. بر اساس یافته‌های آنان: صرف نظر از مدل خاص، در درجه‌بندی ارزیابان، واریانس تمرین‌ها غالب بود. برای مثال در مدلی با پنج بُعد و پنج تمرین، متوسط واریانس توانایی ۰,۱۶ بود و برای مدلی با یک بُعد و پنج تمرین، برابر ۰,۰۴ بود. دی و همکاران (۲۰۰۳) در یک مطالعه فراتحلیل پیرامون روایی مرتبط با ملاک درجه‌بندی ابعاد مراکز ارزیابی، با بررسی ۳۴ مقاله مرتبط، دامنه‌ای بین ۰,۲۵ تا ۰,۳۹ برای روایی درجه‌بندی‌ها گزارش کردند.

مسئله روایی سازه مراکز ارزیابی از اوایل دهه ۱۹۸۰ شروع شد. در آن زمان، ساکِت و دری هر (۱۹۸۲) مطالعه پیش فرض‌های مبتنی برسازه‌های اندازه‌گیری شده در مراکز ارزیابی را مطرح کردند. تا آن زمان، این تصور که ابعاد فرضیِ اندازه‌گیری شده با روش‌های مختلفِ اندازه‌گیری دارای روایی‌اند، کم وبیش موردقبول بود. با این حال، ساکِت و دری هر، متوجه شدند که هم‌بستگی بین ابعاد مختلف در یک روش اندازه‌گیری بالاست، درحالی که

[Downloaded from iieshrm.ir on 2026-11-05]

هم‌بستگی‌های درونی یک بُعد در روش‌های مختلفِ اندازه‌گیری کم است. نتایج تحلیل مؤلفه اصلی نیز روش‌های اندازه‌گیری را به جای ابعاد نشان می‌داد. ازآنجایی که این نتایج در اغلب تحقیقات دیگر نیز تکرار شد (Brannick et al 1987 ;Bycio, Alvares, & ,1989; Hahn, 1985 Russell,)، روایی سازه مراکز ارزیابی با نگاهی انتقادآمیز روبه رو شد، چراکه ابعاد درجه‌بندی شده در تمرین‌ها، اغلب در نشان دادن روایی همگرا ناموفق بودند (2004 Cahoon, Bowler, & Bowle, 2012; Lance, Lambert, Gewin, Lievens, & Conway). باید توجه داشت که مراکز ارزیابی به منظور ارزشیابی افراد در ابعاد خاص و در موقعیت‌های (تمرین های) مختلف طراحی شده‌اند، بااین حال، پژوهش‌ها نشان می‌دهند که روش‌ها (تمرین‌ها) بهتر از ابعاد در ارزشیابی ارزیابی شوندگان، خود را نشان می‌دهد. بر این اساس، اغلب گفته می‌شود که مراکز ارزیابی فاقد روایی سازه‌اند و یا این که ارزیابان قادر به ارزیابی ابعاد موردنیاز نیستند. در این ارتباط، ادبیات تحقیق دلایل مختلفی را برای این کمبود، موردبحث قرار داده و تلاش می‌کند تا متغیرهایی را که روایی سازه را تعدیل می‌کنند، بیابد (Harris, Becker & Smith, 1996 ;Gaugler & Thornton, 1993 ,1989; Kleinmann, Kuptsch, & Köller ;1984 Neidig & Neidig 1986; Silverman, Dalessio, Woods, Johanson &). این مسئله که ابعاد مراکز ارزیابی دارای روایی مرتبط با محتوا و روایی مرتبط با ملاک هستند، اما روایی مرتبط با سازه را نشان نمی‌دهند، ازنظر تئوری تا حدودی متناقض به نظر می‌رسد (Guenole, Oleksandr, Chernyshenko, Stark, Cockerill, & Drasgow, 2013). به ویژه آنکه این دیدگاه با مفهوم‌سازی کنونی از روایی که آن را به‌عنوان یک مفهوم واحد می‌داند و شواهد مرتبط با محتوا و ملاک را استراتژی‌های مختلفی برای نشان دادن روایی سازه یک آزمون یا ابزار اندازه‌گیری مطرح می‌کند، ناسازگار است (Binning & Barret, 1989,). شاید بتوان دو دلیل برای حضور روایی مرتبط با ملاک در غیاب روایی همگرا و تشخیصی ارائه داد؛ اول آنکه ممکن است رویه‌ها، شیوه‌های اجرا و دیگر عوامل روش شناختیِ مراکز ارزیابی مقداری خطا که مانع روایی سنجی همگرا و تشخیصی است، به خطای اندازه‌گیری بیفزاید (Gaugler et ,1987; Lievens,al 1992 Jones, 1998;)، دوم اینکه، مراکز ارزیابی ممکن است سازه‌هایی غیر از آنچه را که موردنظر بوده، اندازه‌گیری کنند (Lance et al ; 1997 ;Raymark & ,1998

[Downloaded from iieshrm.ir on 2026-11-05]

Binning, 1995 Russell & Domm,). این توضیحِ دوم نشان می‌دهد که عدم وجود شواهد روایی همگرا و تشخیصی به خطای اندازه‌گیری مربوط نمی‌شود، بلکه به نامشخص بودن ساختار مکنون سازه موردنظر مرتبط است.

پیشنهادهای متعددی در مورد چگونگی حل معضل روایی سازه مراکز ارزیابی، ازجمله تصریح جهانی ابعاد و تمرین‌ها (Howard, 2008)، استفاده از مدل‌های مختلف ساختاری(Hoffman et al, 2001; Lievens & Conway, 2011) و هم‌چنین اندازه‌گیری کلی و طراحی ویژگی‌های مراکز ارزیابی (Arthur & Woehr, 2003; Arthur,Woehr, & Maldegen, 2000) مطرح شده است. بااین حال، به طورکلی، این تلاش‌ها در برطرف کردن مشکل روایی سازه مراکز ارزیابی ناکام مانده است، به طوری که برخی خواستار طراحی مجدد مراکز ارزیابی برای اندازه‌گیری تمرین‌ها به جای سازه‌ها بوده اند؛ به عبارت دیگر، آن‌ها خواستار حذف ابعاد از مطالعات مرکز ارزیابی شده و تنها تمرکز بر تمرین‌ها را پیشنهاد کرده‌اند (Jackson, Barney, Stillman, & Kirkley, 2007; Neidig & Neidig, 1984). چنین دیدگاهی احتمالاً به دلیل موضع متضاد با هدف اصلی مراکز ارزیابی، یعنی اندازه‌گیری موقعیت داوطلبان در ابعاد مختلف موردنیاز برای عملکرد در شغل شان، موردقبول نیست; (Bowler & Woehr, 2006, 2009; Howard, 2008). برخی محققان جهت رفع مشکلِ روایی سازه مراکز ارزیابی، بر روی انواع رویکردهای روش شناختی تمرکز کرده‌اند. برای نمونه، کلین من و کولر (۱۹۹۷) در تحلیل مجدد داده‌های بای سیو و همکاران، با استفاده از روش تحلیل عاملی، نشان دادند که نتایج محاسبه شده توسط بای سیو و همکاران احتمالاً زیر برآوردی از اثر ابعاد بوده که ناشی از رویکرد آماری مورداستفاده آن‌ها در مدل‌سازی روش‌های همبسته بوده است؛ و یا کلین من و همکاران (۱۹۹۶)، کاربرد روش‌های آماری مناسب، هم‌چون ماتریس چندصفتی-چندروشی را از شیوه‌های برآورد بهتر روایی سازه برای مراکز ارزیابی قلمداد می‌کنند. طی دهه‌های گذشته، پارادایم چندروشی-چندصفتی کمپل و فیسک (۱۹۵۹) به یکی از رایج‌ترین روش‌های کاربردی بررسی روایی سازه تبدیل شده است به طوری که ساختارهای مختلف ماتریس چندصفتی-چندروشی برای بررسی روایی سازه مراکز ارزیابی پیشنهاد شده

[Downloaded from iieshrm.ir on 2026-11-05]

است(Lance et al, 2004; Hoffman et al, 2011). هم‌چنین استفاده از مدل ساختاریِ ابعاد همبسته- تمرین‌ها همبسته نیز به شکل مفهومی مطرح شده است (Lance et al, 2012 ; Monahan, 2011). کاربرد مدل‌های ساختاری ابعاد همبسته- تمرین‌ها همبسته برای مراکز ارزیابی، علی رغم مقبولیت نظری آن، هنگام استفاده از تحلیل عاملی تأییدی اغلب با مشکلاتی همراه است. به کرّات مشاهده شده چنین مدل‌هایی اغلب به همگرایی نمی‌رسند و برازش مناسبی ندارند (Lance et al, 2007)، به همین دلیل، پیشنهادشده است که مشکلات مربوط به روایی سازه ممکن است به طور خاص به یک محصولِ مصنوعیِ اندازه‌گیری مربوط باشد (Hoffman et al, 2011).

محققان دریافته‌اند که با افزایش نسبت شاخص به عامل در یک ماتریس مشابه با ماتریس چندصفتی-چندروشی در تحلیل عامل تأییدی، احتمال بیش‌تری وجود دارد که یک راه‌حل همگرا و قابل قبول به دست آید (Tomas, Hontangas, & Oliver, 2000). یکی از ابزارهای افزایش نسبت شاخص به عامل، استفاده از ابعاد وسیع و یا در نظر گرفتن همه طبقاتی است که فهرستی از ابعاد قابل اندازه‌گیری را در چند دسته دیگر محدود می کند (Woehr & Arthur, 2003)، به‌عنوان مثال، ابعاد منحصربه فرد ارتباطات کلامی، ارتباطات نوشتاری و پاسخگویی می‌تواند در طبقه "ارتباطات" قرار گیرد و به این ترتیب به‌عنوان شاخص‌های این طبقه و به‌صورت گسترده‌تر عمل کند. با افزایش تعداد شاخص‌ها به نسبت تعداد عوامل با استفاده از ابعاد وسیع، ارزیابی روایی سازه یک فرایند ساده‌تر می شود که برخی از مسائل را معمولاً با مدل ابعاد همبسته- تمرین‌ها همبسته و در مطالعات مرکز ارزیابی حذف می‌کند.

به نظر می‌رسد چندین عامل روش شناختی دیگر و وضعیت طراحی بر ویژگی‌های روان سنجیِ درجه‌بندی مراکز ارزیابی تأثیر می‌گذارد؛ ازجمله تعداد ابعاد مشاهده شده، ثبت شده و رتبه‌بندی شده (Thornton& ,1987; Bycio et al Gaugler, 1989 ; Schmitt,1977)؛ آموزش ارزیابان / درجه‌بندی کنندگان و نسبت داوطلب به ارزیاب (Gaugler et al, 1987) و نوع رویکرد ارزیابی (به‌عنوان مثال، درون- ابعاد، در مقابل درون- تمرین ها) (Harris et al, 1993 1982,Silverman, ,;Dreher 1986; Sackett &

[Downloaded from iieshrm.ir on 2026-11-05]

Dalessio,& Woods). تعداد زیادی از پژوهش‌های روایی یابی همگرا/تشخیصیِ قبلی، به جای آنکه عملکرد آزمودنی‌ها را درون- ابعاد مطالعه کنند، درون- تمرین‌ها مورد بررسی قرار داده‌اند. به طور خاص، این دو رویکرد ارزیابی در تمام مراکز ارزیابی مطرح شده است (Sackett & Dreher,1982). در رویکرد درون–تمرینی، ارزیابی شوندگان پس از تکمیل هر تمرین، در هر بُعد درجه‌بندی می‌شوند. متقابلاً، ارزیابی ممکن است پس از آنکه تمام تمرین‌ها انجام شد، انجام گردد؛ یعنی زمانی که مجموعه‌ای از ابعاد بر اساس عملکرد همه تمرین‌ها درجه‌بندی شد، این رویکرد درون- ابعاد است. به طور مفهومی، استدلال می‌شود که درجه‌بندی ابعاد پس از هر تمرین موجب می‌شود، ارزیابان اطلاعات را در شرایط تمرین‌ها پردازش کنند. از طرف دیگر، اگر ارزیابان ابعاد را بعد از تکمیل تمام تمرین‌ها، درجه‌بندی کنند، احتمال زیادی وجود دارد که آن‌ها اطلاعات را برحسب ابعاد به جای تمرین‌ها پردازش نمایند.

سیلورمن و همکاران (۱۹۸۶) شواهدی را ارائه دادند که نشان می‌دهد انتخاب این رویکرد ممکن است، نتایج یافته‌های روایی همگرا/تشخیصی را در مراکز ارزیابی تعدیل نماید. اگرچه نتایج آن‌ها نشان می‌دهد که رویکرد درون- ابعاد به رویکرد درون-تمرین‌ها ترجیح داده می‌شود، هریس و همکاران (۱۹۹۳) موفق به تکرار یافته‌های آنان نشدند. به طور مشخص، آن‌ها "نشان دادند که هر دو رویکرد درجه‌بندی درون-ابعاد و درون-تمرین‌ها تقریباً هم‌بستگی‌های صفت واحد- روش‌های متعدد و هم‌بستگی‌های چندین صفت-روشِ واحد مشابهی به دست می‌دهند". ساکت و دری هر (۱۹۸۲) فرض کردند که محدودیت روایی سازه ممکن است به واسطه یک اثر هاله‌ای رخ دهد. نتایج پژوهش بایسو و همکاران (۱۹۸۷) نیز حاکی از آن است که ارزیابان در پردازشِ اطلاعات و پیچیدگی امر قضاوت، ظرفیت محدودی دارند که منجر به سوگیری شناختی آن‌ها خواهد شد. گاگلر و تورنتون (۱۹۸۹) در یک مطالعه نشان دادند که تغییر تعداد ابعاد (سه بُعد، شش بُعد، نُه بعد) بر روی پردازش اطلاعات ارزیابان مرکز ارزیابی و در نتیجه روایی سازه، به ویژه بر روی روایی همگرا تأثیر می‌گذارد. برای مراکز ارزیابی، ممکن است روایی سازه به این دلیل ضعیف باشد که تمرین‌های مختلف فرصت‌های مختلفی را برای نمایش رفتار مرتبط با

[Downloaded from iieshrm.ir on 2026-11-05]

ابعاد، ارائه می‌دهند. اما مشاهداتِ ساکت و دری هر (۱۹۸۲)، این ایده را رد نمود، چراکه هیچ هم‌بستگی صفت واحد-چند روش بالاتر از ۰,۲۹ نبود. آن‌ها استدلال کردند که حداقل برخی ضرایب تک صفتی– چندروشی باید به سطحی از معنی‌داری دست یابند، به طوری که بتوان آن‌ها را به خوبی مشاهده کرد. دیگر عوامل، شامل نوع ارزیابان (روان‌شناسان در مقابل مدیران و سرپرستان؛ دادن بازخورد؛ مقدار آموزش ارزیابان و تعداد روزهای مشاهده نیز توسط گاگلر و همکاران (۷۹۸۷) مورد بررسی قرارگرفته است.

با توجه به اهمیت روایی درجه‌بندی‌های مراکز ارزیابی و نبود اطلاعات روشنی از روایی مرکز ارزیابی شرکت ملی نفت ایران، مطالعه حاضر بر آن بود تا ضمن گسترش تحقیقات قبلی به بررسی روایی سازه مرکز ارزیابی شرکت ملی نفت ایران و تعیین روایی همگرا و تشخیصی آن با استفاده از رویکرد هم‌بستگی چندروشی-چندصفتی و نیز روش تحلیل عاملی تأییدی بپردازد. لازم به ذکر است که قصد این مطالعه، بررسی اثرات یک طراحی خاص و پیاده‌سازی آن نبود. چنین کاری مستلزم اجرای یک طرح تجربی با دست‌کاری برخی ویژگی‌ها و ثابت نگه داشتن برخی دیگر از ویژگی‌ها در یک گروه کنترل است. البته اجرای چنین طرحی در مرکز ارزیابی فعلی امکان‌پذیر نیست؛ بنابراین هدف مطالعه، صرفاً توصیف و تحلیل یک پدیده در یک موقعیت خاص بود.

روش‌شناسی

طرح پژوهش حاضر، طرح هم‌بستگی بر مبنای ماتریس چندصفتی- چندروشی و هم‌چنین تحلیل عامل تأییدی است. در پژوهش حاضر جامعه آماری، تمامی کارکنان شرکت ملی نفت ایران که از ابتدای سال ۱۳۹۱ تا ۱۳۹۷ در مرکز ارزیابی و توسعه مدیران، مورد سنجش قرار گرفته‌اند، بودند. به منظور انتخاب نمونه از جامعه مدیران ارزیابی شده از روش نمونه‌گیری تصادفی ساده استفاده گردید. تعداد مدیرانی که در مرکز ارزیابی بین سال‌های ۱۳۹۱تا ۱۳۹۷ ارزیابی شدند ۶۰۰۰ نفر بودند که با توجه به فرمول نمونه‌گیری کوکران برای جامعه مشخص (با مقدار اشتباه مجاز ۳٪ و z برابر ۱,۹۶ (، حجم نمونه مورد بررسی ۹۱۰ نفر محاسبه گردید، اما با توجه به ماهیت پژوهش حاضر، این حجم به ۱۰۲۲

[Downloaded from iieshrm.ir on 2026-11-05]

نفر افزایش داده شد و به این ترتیب، جمعیت نمونه این تحقیق را تشکیل دادند.

داده‌های این پژوهش، بر اساس چندین بُعد شایستگی (رهبری، برنامه‌ریزی، توان نقد، پایش عملکرد، ارتباطات، شناخت مسئله، نوآوری، تصمیم‌گیری، تفکر استراتژیک، تفکر تحلیلی و کلان نگری) و با چهار روش (تمرین گروهی، بحث گروهی، پرسش‌نامه و مصاحبه) در یک مقیاس ۹ درجه‌ای (یک تا نُه) درجه‌بندی شدند که بر اساس روش ماتریس چندصفتی-چندروشی و تحلیل عاملی تأییدی مورد مطالعه قرار گرفت. بررسی‌های قبلی روایی همگرا/تشخیصی مرکز ارزیابی اغلب داده‌های خود را با استفاده از تجزیه وتحلیل هم‌بستگی صفر مرتبه یا یک روش تحلیل عاملی (به‌عنوان مثال، تجزیه وتحلیل چندصفتی-چندروشی)، موردمطالعه قرار داده‌اند. درنتیجه، برای مقاصد تطبیقی، داده‌های این مطالعه نیز با استفاده از ماتریس چندصفتی- چندروشی و تحلیل عامل تأییدی مورد بررسی قرار گرفت.

کمپل و فیسک (۱۹۵۹) طرحی را ارائه دادند که در آن، هر مجموعه از صفات چندگانه توسط چندین روش (تمرین) ارزیابی شود. افزون بر آن، راهنمایی‌هایی را برای ارزیابی ماتریس چندصفتی-چندروشی ارائه دادند. وقتی تفاوت‌های افراد را با روش اندازه‌گیری خاصی نشان می‌دهیم، بخشی از این واریانس ناشی از جنبه‌های اختصاصی روش اندازه‌گیری و بخشی ناشی از تفاوت جایگاه افراد روی پیوستار صفتِ مورد اندازه‌گیری است. لذا واریانس منظم ناشی از دو عنصر ۱) روش اندازه‌گیری و ۲) خصوصیات افراد است. در روش بررسیِ روایی چندصفتی-چندروشی، ماتریس هم‌بستگی بین چند متغیر که با روش‌های مختلف اندازه‌گیری شده‌اند، تشکیل می‌شود. قطر اصلی این ماتریس، شاخص پایایی اندازه‌ها هستند و عناصر غیرقطری، هم‌بستگی بین متغیرهایی که با روش واحدی اندازه‌گیری شده‌اند یا هم‌بستگی بین متغیرهایی که با روش‌های مختلفی اندازه‌گیری شده‌اند را نشان می‌دهد. کمپبل و فیسک، در بررسی کلاسیک و بنیادین خود (۱۹۵۹) برای بررسی روایی کامل، مطابق شکل (۱)، شرایط زیر را پیشنهاد می‌کنند:

شرط یکم- ضرایب هم‌بستگی بین اندازه‌های متغیر واحدی که با روش‌های مختلف اندازه‌گیری شده، [۱]، باید مثبت و به طور قابل توجهی از صفر بیش‌تر باشد. این معیار

[Downloaded from iieshrm.ir on 2026-11-05]

معمولاً شرط کافی شناخته می‌شود (روایی همگرا).

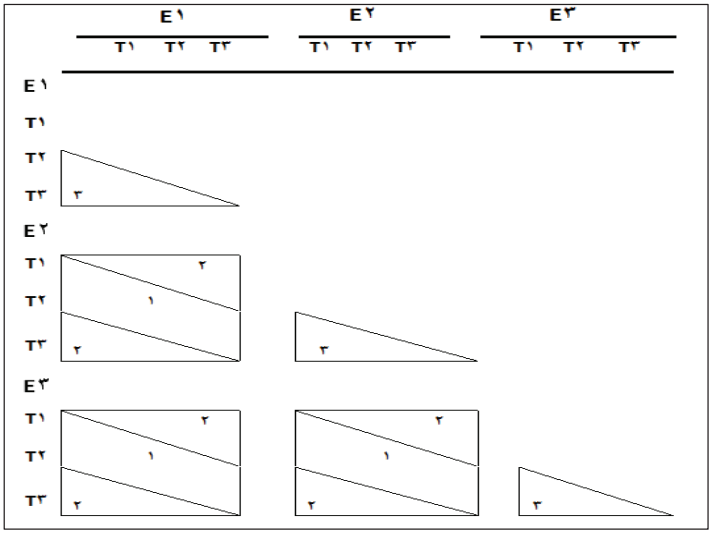
شرط دوم- ضرایب هم‌بستگی بین چندین صفت که با روش واحدی اندازه‌گیری شده، [۳]، باید به طور معنی‌دار از هم‌بستگی بین اندازه‌های متغیر واحدی که با روش‌های مختلف اندازه‌گیری شده [۱]، کوچک‌تر باشد (روایی تشخیصی).

شرط سوم-هم‌بستگی بین اندازه‌های چندین صفت که با روش‌های مختلف اندازه‌گیری شده [۲]، نیز باید به طور معنی‌دار از هم‌بستگی بین اندازه‌های متغیر واحدی که با روش‌های مختلف اندازه‌گیری شده [۱]، کوچک‌تر باشد (روایی تشخیصی).

شرط چهارم- هم‌بستگی بین چندین صفت که با روش واحدی اندازه‌گیری شده [۳] باید همان الگوی هم‌بستگی بین اندازه‌های چندین صفت که با روش‌های مختلف اندازه‌گیری شده‌اند [۲] را نشان دهد.

توجه: قطر ماتریس با شماره ۱ اشاره به هم‌بستگی‌های یک صفت واحد- با چند روش، درایه‌های غیر قطری با شماره ۲ اشاره به هم‌بستگی بین چند صفت- با چند روش و درایه‌های غیر قطری با شماره ۳، نشانگر هم‌بستگی چند صفت- با روش واحد است.

[Downloaded from iieshrm.ir on 2026-11-05]



شکل ۱- مثالی از ماتریس چندصفتی-چندروشی با سه صفت و سه روش. (Salkind, 2010)

اگر شرط اول برقرار باشد، می‌گوییم که روش‌های اندازه‌گیری دارای روایی همگرا هستند؛ اما نباید به این شرط تنها اکتفا کرد. باید نشان داد که شرط دوم و سوم نیز صادق است. در این صورت می‌گوییم روش‌های اندازه‌گیری دارای روایی تشخیصی هستند. برخی محققان شرط چهارم را واقع بینانه نمی‌دانند و حصول دقیق آن را محال می‌دانند (مگنوسون، ۱۹۶۷، ترجمه براهنی، ۱۳۷۵). برخی از محققان معتقدند که شرایط کمپبل و فیسک را نمی‌توان جدا از درجه پایایی روش‌ها ارزیابی کرد. اگر نتایج اندازه‌گیری فاقد پایایی کافی باشد، در این صورت باید عدول مختصر از شرایط روایی تشخیصی را تحمل کرد. هرچند تجزیه وتحلیل هم‌بستگی، شواهدی در رابطه با اثرات ابعاد یا روش‌ها فراهم می‌کند؛ اما امکان آزمون کلی این اثرات را به دست نمی‌دهد؛ بنابراین، در این مطالعه، از تحلیل عامل تأییدی برای ارزیابی مدلی که هر دو اثر ابعاد و روش‌ها را ارائه دهد، استفاده شد. به نظر می‌رسد، روش‌های تحلیل عاملی تأییدی برای ارزیابی داده‌های ماتریس چندصفتی-چندروشی بسیار مناسب و محبوب باشند (Marsh & Grayson, 1995). در

[Downloaded from iieshrm.ir on 2026-11-05]

رویکرد تحلیل عاملی تأییدی، عواملی که به کمک چندین اندازه مربوط به همان صفت تعریف می‌شوند، منعکس‌کننده روایی سازه اندازه هاست، درحالی که عوامل مبتنی بر اندازه‌های صفات مختلف که با همان ابزار است، نشان‌دهنده اثر روش هاست. مزیت بزرگ تحلیل عاملیِ تأییدی، امکان تثبیت بارهای عاملی و هم‌بستگی بین عامل هاست. تجزیه وتحلیل این نوع هم‌بستگی‌ها دربرگیرنده مقایسه میانگین هم‌بستگی‌های درون-ابعاد بین-روش‌ها، با میانگین هم‌بستگی‌های درون-روش‌ها بین-ابعاد است. ارزش‌های بالاتر برای هم‌بستگی‌های درون-ابعاد، بین-روش‌ها نسبت به هم‌بستگی‌های درون-روش‌ها، بین-ابعاد نشانگر روایی همگرا/تشخیصی است (Salkind, 2010).

یافته‌ها

ماتریس هم‌بستگی متغیر ابعاد (شایستگی‌ها) (بُعد رهبری، برنامه‌ریزی، توان نقد، پایش عملکرد، ارتباطات، شناخت مسئله، نوآوری، تصمیم‌گیری) با استفاده از روش‌های مختلف (تمرین گروهی، مصاحبه، بحث گروهی و پرسش‌نامه) تشکیل شد. با فرض این که روش‌های مختلف اندازه‌گیری از لحاظ خصوصیات مختلف هرچه بیش‌تر ناهمانند باشد، می توان ضرایب هم‌بستگی حاصل را شاخص روایی سازه دانست (تحلیل ماتریس چندصفتی – چندروشی).

روابط درونی بین درجه‌بندی‌ها برای هر بُعد حاصل از هر روش (ابزار) در جداول (۱) و (۳) نشان داده شده است. در این دو جدول، ضرایب هم‌بستگی بین ابعاد که با روش واحدی اندازه‌گیری شده‌اند، در داخل مثلث‌هایی با خطوط پیوسته، کنار قطر اصلیِ ماتریس نشان داده شده است. چون در این مورد، هر بُعد با روش واحدی اندازه‌گیری شده است، خصوصیات روش واحد سبب پیدایش نوعی واریانس مشترک در همه ابعاد می‌شود. از طرفی ضرایب هم‌بستگی بین ابعاد مختلف که با روش‌های مختلف اندازه‌گیری شده‌اند، در داخل دو مثلث با خطوط منقطع (گوشه سمت راست پایین) نشان داده شده‌اند. مقادیر قطری بین این دو مثلث شاخص ضرایب هم‌بستگی بین اندازه‌های بعدی واحد است که با روش‌های مختلف اندازه‌گیری شده‌اند. این ضرایب اخیر، ضرایب روایی همگرا هستند.

[Downloaded from iieshrm.ir on 2026-11-05]

- برای بُعد رهبری با استفاده از تحلیل جدول (۱) سه نتیجه حاصل گردید؛

مطابق جدول (۱)، ضریب هم‌بستگی برای بُعد رهبری که توسط ابزارهای اندازه‌گیری مختلف (تمرین گروهی و مصاحبه) به دست آمد، مثبت و در سطح ۰/۰۱ آماری معنی‌دار بود، هرچند مقدار آن چندان قابل توجه نبود (۰/۳۶۳). بنابراین بُعد رهبری اولین شرط کمپبل و فیسک را نسبتاً فراهم می‌کند.

ضرایب هم‌بستگی بین بُعد رهبری با سایر ابعاد که توسط ابزار اندازه‌گیری یکسانی به دست آمد، همگی مثبت و بالاتر از ۰/۴۶ (میانگین= ۰,۶۲) و نیز بالاتر از مقدار محاسبه شده در بند ۱ بود. به عبارت دیگر، شرط دوم کمپبل و فیسک در مورد بُعد رهبری برقرار نیست.

ضرایب هم‌بستگی بین بُعد رهبری و دیگر ابعادی که با روش‌های متفاوتی اندازه‌گیری شده است، همگی مثبت و کمتر از ۰/۲۵ بود (میانگین= ۰,۱۸). همان‌طور که مشاهده می شود، در رابطه با بُعد رهبری، شرط سوم کمپبلو فیسک تنها برای برخی خانه‌ها برقرار است جدول (۱).

- برای بُعد برنامه‌ریزی با استفاده از تحلیل جدول (۱) نیز سه نتیجه زیر به دست آمد؛

ضریب هم‌بستگی برای بُعد برنامه‌ریزی که توسط ابزارهای اندازه‌گیری مختلف (تمرین گروهی و مصاحبه) به دست آمد، مثبت و در سطح ۰/۰۱ آماری معنی‌دار بود، مقدار این ضریب ۰/۳۱۷ برآورد شد که نسبتا رضایت بخش است.

ضرایب هم‌بستگی بین بُعد برنامه‌ریزی با سایر ابعاد که توسط ابزار اندازه‌گیری یکسانی به دست آمد، همگی مثبت و بالاتر از ۰/۴۷ بود (میانگین= ۰,۶۲). این مقادیر در ضمن بالاتر از مقدار محاسبه شده در بند۱ بود؛ بنابراین، شرط دوم کمپبل و فیسک در مورد بُعد برنامه‌ریزی برقرار نیست.

ضرایب هم‌بستگی بین بُعد برنامه‌ریزی و دیگر ابعادی که با روش‌های متفاوتی اندازه‌گیری شده‌اند، همگی مثبت و کمتر از ۰/۲۳ بود (میانگین= ۰,۱۹) همان‌طور که مشاهده می‌شود، در رابطه با متغیر برنامه‌ریزی، شرط سوم کمپبل و فیسک برای برخی خانه‌ها برقرار نیست.

[Downloaded from iieshrm.ir on 2026-11-05]

جدول ۱- ماتریس درونی ابعاد و روش‌ها (MTMM) برای پنج بُعد و دو روش (یافته‌های پژوهشگران)

		تمرین گروهی					مصاحبه				
		رهبری	برنامه‌ریزی	شناخت مساله	نوآوری	تصمیم	رهبری	برنامه‌ریزی	شناخت مساله	نوآوری	تصمیم
تمرین گروهی	رهبری	–									
	برنامه‌ریزی	۰,۷۵۲	–								
	شناخت‌مساله	۰,۶۱۸	۰,۶۵۷	–							
	نوآوری	۰,۶۷۶	۰,۶۶۴	۰,۵۱۵	–						
	تصمیم‌گیری	۰,۷۱۱	۰,۷۱۶	۰,۵۸۹	۰,۶۰۶	–					
مصاحبه	رهبری	۰,۳۶۳	۰,۱۶۷	۰,۱۲۳	۰,۱۲۳	۰,۱۶۲	–				
	برنامه‌ریزی	۰,۲۲۶	۰,۳۱۷	۰,۱۸۷	۰,۱۲۱	۰,۲۱۰	۰,۶۲۵	–			
	شناخت مساله	۰,۲۲۳	۰,۲۱۳	۰,۲۸۵	۰,۱۴۵	۰,۲۲۰	۰,۵۴۳	۰,۵۲۵	–		
	نوآوری	۰,۲۵۰	۰,۲۰۷	۰,۱۹۸	۰,۳۶۵	۰,۲۳۸	۰,۴۶۷	۰,۴۷۴	۰,۵۹۶	–	
	تصمیم‌گیری	۰,۱۹۹	۰,۱۹۴	۰,۱۹۵	۰,۱۴۱	۰,۳۸۷	۰,۶۱۱	۰,۵۸۶	۰,۶۳۲	۰,۴۳۱	–

تمامی هم‌بستگی‌ها در سطح ۰,۰۱ آماری معنی‌دار هستند.

- برای بُعد شناخت مسئله نیز با استفاده از تحلیل جدول (۱) نتایج زیر به دست آمد؛

ضریب هم‌بستگی برای بُعد شناخت مسئله که توسط ابزارهای اندازه‌گیری مختلف (تمرین گروهی و مصاحبه) به دست آمد مثبت و در سطح ۰/۰۱ آماری معنی‌دار بود، مقدار این ضریب ۰/۲۸۵ برآورد شد.

ضرایب هم‌بستگی بین بُعد شناخت مسئله با سایر ابعاد که توسط ابزار اندازه‌گیری یکسانی به دست آمد، همگی مثبت و بالاتر از ۰/۵۰ بودند (میانگین=۰,۵۸). این مقادیر در ضمن بالاتر از مقدار محاسبه شده در بند۱ بود؛ بنابراین، شرط دوم کمپبل و فیسک در

[Downloaded from iieshrm.ir on 2026-11-05]

مورد بُعد برنامه‌ریزی برقرار نیست.

ضرایب هم‌بستگی بین بُعد شناخت مسئله و دیگر ابعادی که با روش‌های متفاوتی اندازه‌گیری شده‌اند، همگی مثبت و کمتر از ۰/۲۳ بود (میانگین=۰,۱۹). همان‌طور که مشاهده می‌شود، در رابطه با متغیر برنامه‌ریزی، شرط سوم کمپبل و فیسک برای برخی خانه‌ها برقرار نیست.

- برای بُعد نوآوری با استفاده از تحلیل جدول (۱) نیز نتایج زیر به دست آمد؛

ضریب هم‌بستگی برای بُعد نوآوری که توسط ابزارهای اندازه‌گیری مختلف (تمرین گروهی و مصاحبه) به دست آمد مثبت و در سطح ۰/۰۱ آماری معنی‌دار بود. مقدار این ضریب ۰/۳۶۵ برآورد شد که نسبتاً رضایت بخش است.

ضرایب هم‌بستگی بین بُعد نوآوری با سایر ابعاد که توسط ابزار اندازه‌گیری یکسانی به دست آمد، همگی مثبت و بالاتر از ۰/۴۳ بودند (میانگین= ۰,۵۵). این مقادیر در ضمن به میزان اندکی بالاتر از مقدار محاسبه شده در بند۱ بود. به عبارت دیگر شرط دوم کمپبل و فیسک در مورد بُعد نوآوری چندان برقرار نیست.

ضرایب هم‌بستگی بین بُعد نوآوری و دیگر ابعادی که با روش‌های متفاوتی اندازه‌گیری شده است، همگی مثبت و کمتر از ۰/۲۴ بود (میانگین= ۰,۱۸). همان‌طور که مشاهده می شود، در رابطه با متغیر نوآوری، شرط سوم کمپبل و فیسک برای برخی خانه‌ها برقرار نیست.

- برای بُعد تصمیم‌گیری با استفاده از تحلیل جدول (۱) نیز نتایج زیر به دست آمد؛

ضریب هم‌بستگی برای بُعد تصمیم‌گیری که توسط ابزارهای اندازه‌گیری مختلف (تمرین گروهی و مصاحبه) به دست آمد مثبت و در سطح ۰/۰۱ آماری معنی‌دار بود. مقدار این ضریب ۰/۳۸۷ به دست آمد.

ضرایب هم‌بستگی بین بُعد تصمیم‌گیری با سایر ابعاد شایستگی که توسط ابزار اندازه‌گیری یکسانی به دست آمد، همگی مثبت و بالاتر از ۰/۴۳ بودند (میانگین= ۰,۶۱). این مقادیر در ضمن بالاتر از مقدار محاسبه شده در بند۱ بود. به عبارت دیگر شرط دوم کمپبل و فیسک در مورد بُعد تصمیم‌گیری برقرار نیست.

[Downloaded from iieshrm.ir on 2026-11-05]

ضرایب هم‌بستگی بین بُعد تصمیم‌گیری و دیگر ابعادی که با روش‌های متفاوتی اندازه‌گیری شده است، همگی مثبت و کمتر از ۰/۲۴ بود (میانگین= ۰,۱۹). همان‌طور که مشاهده می‌شود، در رابطه با متغیر تصمیم‌گیری، شرط سوم کمپبل و فیسک برای اغلب خانه‌ها برقرار نیست.

جدول (۲) میانگین هم‌بستگی‌های درون ابعاد- بین روش‌ها و درون روش ها- بین ابعاد برای جدول (۱) را نشان می‌دهد.

جدول ۲- میانگین درون ابعاد-بین روش‌ها و درون روش‌ها – بین ابعاد (یافته‌های پژوهشگران)

میانگین هم‌بستگی‌های درون-ابعاد، بین-روش‌ها	میانگین هم‌بستگی‌های درون-روش‌ها، بین-ابعاد
رهبری = ۰,۲۶۳ برنامه‌ریزی = ۰,۲۸۵ شناخت مساله= ۰,۳۱۷ نوآوری = ۰,۳۶۵ تصمیم‌گیری = ۰,۳۸۷	تمرین گروهی = ۰,۶۵۰ مصاحبه = ۰,۵۴۹
میانگین کل = ۰,۳۲۳	میانگین کل = ۰,۵۹۹

همان‌طور که مشاهده می‌شود، میانگین کل هم‌بستگی بین درجه‌بندی‌های مربوط به یک بُعد واحد که با دو روش مختلف (تمرین گروهی و مصاحبه) اندازه‌گیری شده بود، ۰/۳۲ به دست آمد که در مقایسه با میانگین کل هم‌بستگی بین درجه‌بندی‌های ابعاد مختلف در همان روش (یعنی ۰/۵۹)، کمتر می‌باشد. به عبارت دیگر، ضرایب هم‌بستگی بین چندین صفت که با روش واحدی اندازه‌گیری شده به طور معنی‌دار بزرگ‌تر از هم‌بستگی بین اندازه‌های متغیر واحدی است که با روش‌های مختلف اندازه‌گیری شده است. بنابراین هرچند اندازه‌های مربوط به ابعاد (رهبری، برنامه‌ریزی، شناخت مسئله، نوآوری و تصمیم‌گیری) از اعتبار همگرایی نسبتاً خوبی برخوردارند، اما فاقد اعتبار تشخیصی هستند.

در جدول (۳) روابط درونی بین درجه‌بندی‌ها برای دیگر ابعاد شایستگی (توان نقد،

[Downloaded from iieshrm.ir on 2026-11-05]

پایش عملکرد و ارتباطات) حاصل از هر روش (ابزار) نشان داده است. برای بُعد توان نقد با استفاده از تحلیل جدول (۳) نیز نتایج زیر به دست آمد؛

مطابق جدول (۳)، ضریب هم‌بستگی برای بُعد توان نقد که توسط ابزارهای اندازه‌گیری مختلف (تمرین گروهی و بحث گروهی) به دست آمد (۰,۲۸۵)، مثبت و در سطح ۰/۰۱ آماری معنی‌دار بود.

ضرایب هم‌بستگی بین بُعد توان نقد با دو بُعد دیگر که توسط ابزار اندازه‌گیری یکسانی به دست آمد، مثبت و بالاتر از ۰/۵۷ (میانگین = ۰,۵۹) و نیز بالاتر از مقدار محاسبه شده در بند۱ بود. به عبارت دیگر، شرط دوم کمپبل و فیسک در مورد بُعد توانِ نقد برقرار نیست.

ضرایب هم‌بستگی بین بُعد توان نقد و دو بُعد دیگر که با روش‌های متفاوتی اندازه‌گیری شده بود، مثبت و کمتر از ۰/۲۵ بود (میانگین = ۰,۲۸). همان‌طور که مشاهده می‌شود، در رابطه با بُعد توانایی نقد، شرط سوم کمپبل و فیسک تنها برای برخی خانه‌ها برقرار است.

- برای بُعد پایش عملکرد با استفاده از تحلیل جدول (۳) نیز نتایج زیر به دست آمد؛

ضریب هم‌بستگی برای بُعد پایش عملکرد که توسط دو ابزار اندازه‌گیری مختلف (تمرین گروهی و بحث گروهی) به دست آمد، مثبت و در سطح ۰/۰۱ آماری معنی‌دار بود، مقدار آن (۰/۲۵۱) به دست آمد.

ضرایب هم‌بستگی بین بُعد پایش عملکرد با سایر ابعاد که توسط ابزار اندازه‌گیری یکسانی به دست آمد، همگی مثبت و بالاتر از ۰/۵۹ (میانگین = ۰,۶۱) و نیز بالاتر از مقادیر محاسبه شده دربند۱ بود. به عبارت دیگر، شرط دوم کمپبل و فیسک در مورد بُعد پایش عملکرد برقرار نیست.

ضرایب هم‌بستگی بین بُعد پایش عملکرد و دیگر ابعادی که با روش‌های متفاوتی اندازه‌گیری شده است، همگی مثبت و کمتر از ۰/۲۵ بود (میانگین = ۰,۲۸). همان‌طور که مشاهده می‌شود، در رابطه با بُعد پایش عملکرد، شرط سوم کمپبل و فیسک تنها برای برخی خانه‌ها برقرار است.

- برای بُعد ارتباطات با استفاده از تحلیل جدول (۳) نیز نتایج زیر به دست آمد؛

[Downloaded from iieshrm.ir on 2026-11-05]

ضریب هم‌بستگی بین بُعد ارتباطات که توسط ابزارهای اندازه‌گیری مختلف (تمرین گروهی و بحث گروهی) به دست آمد، مثبت و در سطح ۰/۰۱ آماری معنی‌دار بود. مقدار ضریب هم‌بستگی (۰/۳۹۷) به دست آمد.

ضرایب هم‌بستگی بین بُعد ارتباطات با سایر ابعاد که توسط ابزار اندازه‌گیری یکسانی به دست آمد، همگی مثبت و بالاتر از ۰/۵۷ (میانگین = ۰,۵۹) و نیز بالاتر از مقادیر محاسبه شده در بند ۱ بود. به عبارت دیگر، شرط دوم کمپبل و فیسک در مورد بُعد ارتباطات برقرار نیست.

ضرایب هم‌بستگی بین بُعد ارتباطات و دیگر ابعادی که با روش‌های متفاوتی اندازه‌گیری شده است، همگی مثبت و کمتر از ۰/۳۰ بود (میانگین= ۰,۳۱). لذا، در رابطه با بُعد ارتباطات، شرط سوم کمپبل و فیسک برقرار است.

در قسمت انتهایی جدول (۳)، میانگین هم‌بستگی‌های درون ابعاد-بین روش‌ها و درون روش ها–بین ابعاد داده شده است. همان‌طور که مشاهده می‌شود، میانگین کل هم‌بستگی بین درجه‌بندی‌های مربوط به یک بُعد واحد که با دو روش مختلف (تمرین گروهی و بحث گروهی) اندازه‌گیری شده بود، ۰/۳۱ به دست آمد که در مقایسه با میانگین کل هم‌بستگی بین درجه‌بندی‌های ابعاد مختلف در همان روش که ۰/۶۰ است، بسیار کمتر است. بنابراین، نتایج نشان می‌دهد که اندازه‌های مربوط به ابعاد (توانایی نقد، پایش عملکرد و ارتباطات) از اعتبار همگرایی ضعیفی برخوردارند، ضمن آنکه فاقد اعتبار تشخیصی نیز هستند.

جدول ۳- ماتریس درونی ابعاد و روش‌ها (MTMM) برای سه بعد و دو روش (یافته‌های پژوهشگران)

		تمرین گروهی			بحث		
		توانایی نقد	پایش	ارتباطات	توانایی نقد	پایش	ارتباطات
تمرین گروهی	توانایی نقد						
	پایش	۰,۶۲۶					

[Downloaded from iieshrm.ir on 2026-11-05]

بحث	ارتباطات	۰,۵۷۲	۰,۵۹۴			
	توانایی نقد	۰,۲۸۵	۰,۲۵۱	۰,۳۰۴		
	پایش	۰,۲۵۳	۰,۲۵۱	۰,۳۳۸	۰,۶۰۴	
	ارتباطات	۰,۳۰۶	۰,۲۹۹	۰,۳۹۷	۰,۵۷۴	۰,۶۰۶
		میانگین هم‌بستگی های درون-ابعاد، بین-روش ها			میانگین هم‌بستگی های درون-روش‌ها، بین-ابعاد	
		توانایی نقد = ۰,۲۸۵ پایش عملکرد = ۰,۲۵۱ ارتباطات = ۰,۳۹۷			تمرین گروهی= ۰,۵۹۷ بحث = ۰,۵۹۵	
		میانگین کل = ۰,۳۱۱			میانگین کل = ۰,۵۹۶	

تمامی هم‌بستگی‌ها در سطح ۰,۰۱ آماری معنی‌دار هستند.

جدول (۴) هم‌بستگی‌های درون ابعاد-بین روش‌ها برای سه بُعد دیگر (تفکر استراتژیک، تفکر تحلیلی و کلان نگری) که با روش‌های مختلف به دست آمد را نشان می‌دهد. همان‌طور که مشاهده می‌شود، میانگین این هم‌بستگی‌ها ۰,۲۱ بود.

جدول ۴- هم‌بستگی‌های درون ابعاد-بین روش‌ها برای سه بُعد تفکر استراتژیک، تفکر تحلیلی، کلان نگری (یافته‌های پژوهشگران)

ابعاد	روش		هم‌بستگی
تفکر استراتژیک	بحث	پرسش‌نامه	۰,۱۸۱
تفکر تحلیلی	بحث	مورد کاوی	۰,۲۲۰
کلان نگری	بحث	مصاحبه	۰,۲۳۹
		میانگین	۰,۲۱

[Downloaded from iieshrm.ir on 2026-11-05]

تمامی هم‌بستگی‌ها در سطح ۰,۰۱ آماری معنی‌دار هستند.

جدول (۵) هم‌بستگی‌های درون روش ها-بین ابعاد برای سه بُعد تفکر استراتژیک، تفکر تحلیلی و کلان نگری که با روش‌های مختلف به دست آمد را نشان می‌دهد. همان‌طور که مشاهده می‌شود، میانگین این هم‌بستگی‌ها ۰,۶۷ بود که در مقایسه با هم‌بستگی‌های درون ابعاد-بین روش‌ها جدول (۴) به مراتب بیش‌تر است و نشان از نبود اعتبار تشخیصی اندازه‌ها دارد.

جدول ۵- هم‌بستگی‌های درون ابعاد-بین روش‌ها برای سه بُعد تفکر استراتژیک، تفکر تحلیلی، کلان نگری (یافته‌های پژوهشگران)

روش	ابعاد		هم‌بستگی
بحث	تفکر استراتژیک	کلان نگری	۰,۷۰۶
بحث	تفکر استراتژیک	تفکر تحلیلی	۰,۷۵۶
بحث	کلان نگری	تفکر تحلیلی	۰,۷۹۳
	میانگین		۰,۷۴

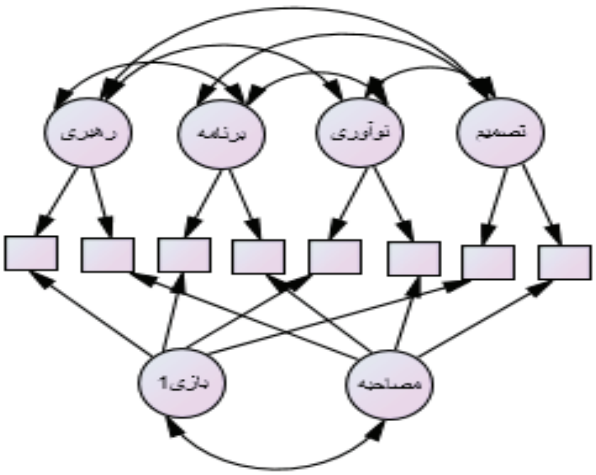
تمامی هم‌بستگی‌ها در سطح ۰,۰۱ آماری معنی‌دار هستند.

در ادامه این مطالعه، از تحلیل عامل تأییدی (CFA) برای ارزیابی مدلی که هر دو اثر ابعاد و روش‌ها را ارائه دهد، استفاده شد. جهت ارزیابی برازش مدلِ عاملیِ تأییدی تدوین شده، از برنامه آموس استفاده شد و کوواریانس‌های بین هشت درجه‌بندی (یک درجه‌بندی برای هریک از چهار بُعد بر اساس هریک از دو روش) موردبررسی قرار گرفت.

جهت جمع‌آوری داده‌های موردنیاز CFA، دو ارزیاب به شکلی انتخاب شد که بر روی گروه یکسانی از افراد و با توجه به ابعادی یکسان ولی با روش‌های ارزیابی متفاوت، عمل درجه‌بندی را انجام داده باشند؛ برای مثال اگر ارزیاب۱، افراد را در بُعد رهبری، مورد ارزیابی قرار داده است، ارزیاب ۲ نیز همان افراد را در همان بُعد مورد ارزیابی قرار داده

[Downloaded from iieshrm.ir on 2026-11-05]

باشد. تنها تمایز در این بود که هر ارزیاب، افراد را با یک روش ارزیابی جداگانه مورد درجه‌بندی قرار داده بود. مثلا اگر ارزیاب ۱ افراد را با روش تمرین گروهی مورد درجه‌بندی قرار داده بود، ارزیاب ۲ نیز همان افراد را با تمرین مصاحبه، درجه‌بندی نموده باشد. بنابراین از یک طرحِ کاملاً متقاطع جهت جمع‌آوری داده‌ها استفاده شد. استفاده از طرح متقاطع سبب شد تا تعداد ابعاد و روش‌های مورد مطالعه به چهار بُعد و دو روش و حجم نمونه به ۴۲۱ نفر تقلیل یابد.



شکل ۲- مدل تدوین شده تحلیل عاملی ابعاد-روش‌ها (یافته‌های پژوهشگران)

مدل تدوین شده مورد ارزیابی در این مطالعه، مطابق شکل (۲) شامل شش متغیر نهفته بود. چهار متغیر نهفته، عوامل مربوط به چهار بُعد شایستگی (رهبری، برنامه‌ریزی، نوآوری و تصمیم‌گیری) و دو متغیر نهفته نشان‌دهنده عامل‌های مربوط به روش اندازه‌گیری (تمرین گروهی و مصاحبه) است. اندازه‌های کلی برازش برای این مدل نشان می‌دهد که به خوبی یک مدلِ مشخص شده با چهار بُعد و دو روش به داده‌ها مربوط می شود. علاوه براین، مقایسه مقادیر پارامترهای برآورد شده عاملِ ابعاد روی درجه‌بندی‌ها در

[Downloaded from iieshrm.ir on 2026-11-05]

مقابلِ برآورد پارامترهای عاملِ روش‌ها روی درجه‌بندی‌ها، شاخصی از اندازه نسبی اثرات ابعاد و روش‌ها را نشان می‌دهد. به طور خاص، بارهای عاملی بزرگ برای ابعاد، نشان‌دهنده وجود اعتبار همگرا، بارهای عاملی بزرگ برای روش‌ها نشان‌دهنده اثرات روش اندازه‌گیری و هم‌بستگی‌های بالا (به‌عنوان مثال، نزدیک به ۱) بین ابعاد نشان‌دهنده فقدان اعتبار تشخیصی است (Marsh & Grayson, 1995).

جدول (۶)، بارهای عاملی متعلق به هر شاخص اندازه‌گیری را بر روی متغیرهای نهفته ابعاد و روش‌ها نشان می‌دهد. در این جدول ستون اول از راست، نشان‌دهنده هشت متغیر مشاهده شده مدل تحلیل عاملی است. چهارستون بعدی، تحت عنوان ابعاد، چهار متغیر نهفته مربوطه را به نمایش می‌گذارد و دو ستون آخر تحت عنوان روش‌ها دو متغیر نهفته مربوط به روش‌های ارزیابی را نشان داده، مقادیر درون خانه‌های جدول، بارهای عاملی برآورد شده را مشخص می‌کند. برای مثال، شاخص "B، رهبری" مربوط به اندازه‌گیری شایستگی رهبری از طریق تمرین گروهی است. این شاخص روی متغیر نهفته رهبری دارای بار عاملی به میزان ۰,۶۷۷ و بر روی متغیر نهفته تمرین گروهی، دارای بار عاملی به میزان ۰,۵۸۱ است. به همین ترتیب، شاخص "M تصمیم‌گیری" مربوط به اندازه‌گیری شایستگی تصمیم‌گیری از طریق تمرین مصاحبه است. همان‌طور که مشاهده می‌شود، این شاخص روی متغیر نهفته تصمیم‌گیری دارای بار عاملی به میزان ۰,۷۱۳ و بر روی متغیر نهفته مصاحبه، دارای بار عاملی به میزان ۰,۲۳۴ است.

جدول ۶- برآورد پارامترهای استاندارد مدل تحلیل عاملی تأییدی چند روشی-چند صفتی (یافته‌های پژوهشگران)

شاخص اندازه‌گیری	ابعاد				روش‌ها (تمرین)	
	رهبری	برنامه‌ریزی	نوآوری	تصمیم‌گیری	تمرین گروهی (B)	مصاحبه (M)
B، رهبری	۰,۶۷۷				۰,۵۸۱	
B، برنامه‌ریزی		۰,۷۶۸			۰,۴۱۵	
B، نوآوری			۰,۶۴۵		۰,۳۹۰	

[Downloaded from iieshrm.ir on 2026-11-05]

B، تصمیم‌گیری			۰,۶۰۶	۰,۶۷۷
M، رهبری	۰,۷۹۴			۰,۱۹۰
M، برنامه‌ریزی		۰,۷۲۵		۰,۲۷۳
M، نوآوری			۰,۵۰۶	۰,۳۴۲
M، تصمیم‌گیری			۰,۷۱۳	۰,۲۳۴
میانگین بار عاملی ابعاد = ۰,۶۷۹				
B= تمرین گروهی، M= مصاحبه				

به طورکلی این جدول نشان می‌دهد که متوسط هم‌بستگی بین شاخص‌های اندازه‌گیری با متغیر مکنون روش (۰,۳۸۸) به مراتب کمتر از هم‌بستگی بین شاخص‌های اندازه‌گیری با متغیر مکنون ابعاد (۰,۶۷۹) است جدول (۶).

جداول (۷) و (۸)، بارهای عاملی جدول (۶) را به تفکیک برای ابعاد و روش‌ها نشان داده است. برای مثال جدول (۷) نشان می‌دهد که ارزش بارهای عاملی "درونِ روش ها-بینِ ابعاد"، در اغلب موارد کوچک‌تر از ۰/۵۰ برآورد شده است. درحالی که ارزش بارهای عاملی "درونِ ابعاد-بینِ روش ها" در همه موارد بزرگ‌تر از ۰/۵۰ به دست آمده است جدول (۸).

جدول ۷- بارهای عاملی درون-روش‌ها، بین-ابعاد (یافته‌های پژوهشگران)

روش اندازه‌گیری	شاخص اندازه‌گیری	بار عاملی	R^2
تمرین گروهی	رهبری	۰/۵۸۱	۰/۳۳۷
	برنامه‌ریزی	۰/۱۹۰	۰/۰۳۶
	نوآوری	۰/۴۱۵	۰/۱۷۲
	تصمیم‌گیری	۰/۲۷۳	۰/۰۷۴
مصاحبه	رهبری	۰/۳۹۰	۰/۱۵۲
	برنامه‌ریزی	۰/۳۴۲	۰/۱۱۷

[Downloaded from iieshrm.ir on 2026-11-05]

	نوآوری	۰/۶۷۷	۰/۴۵۸
	تصمیم‌گیری	۰/۲۳۴	۰/۰۵۵
متوسط واریانس تبیین شده توسط عامل روش ها			۰/۱۷۵

هم‌چنین مقایسه نسبت واریانس تبیین شده (R^2) توسط روش‌ها و ابعاد نشان می‌دهد که سهم ابعاد در توصیف تغییرات درجه‌بندی‌های صورت گرفته به مراتب بیش‌تر از سهم روش هاست. به طوری که متوسط واریانس تبیین شده برای ابعاد ۴۶/۹ درصد جدول (۸) به دست آمد، درحالی که برای روش‌ها ۱۷/۵ درصد حاصل گردید جدول (۷). با توجه به ارزیابی اعتبار سازه، این نتایج می‌تواند به‌عنوان تأثیر نسبتاً اندک روش‌ها در واریانس رفتار درجه‌بندی‌ها و متقابلاً تأثیر زیاد ابعاد در واریانس درجه‌بندی‌ها قلمداد شود.

جدول ۸- بارهای عاملی درون-ابعاد، بین-روش‌ها (یافته‌های پژوهشگران)

ابعاد شایستگی	شاخص اندازه‌گیری	بار عاملی	R^2
رهبری	تمرین گروهی ۱	۰/۶۷۷	۰/۴۵۸
	مصاحبه	۰/۷۶۸	۰/۵۹۰
برنامه‌ریزی	تمرین گروهی ۱	۰/۶۴۵	۰/۴۱۶
	مصاحبه	۰/۶۰۶	۰/۳۶۷
نوآوری	تمرین گروهی ۱	۰/۷۹۴	۰/۶۳۰
	مصاحبه	۰/۷۲۵	۰/۵۲۶
تصمیم‌گیری	تمرین گروهی ۱	۰/۵۰۶	۰/۲۵۶
	مصاحبه	۰/۷۱۳	۰/۵۰۸
متوسط واریانس تبیین شده توسط عامل ابعاد			۰/۴۶۹

جدول (۹) آماره‌های خوبی برازش تحلیل عاملی را نشان می‌دهد. شاخص‌های کلی برازش نشان داد که مدل، یک برازش رضایت بخش را فراهم نموده است؛ مقدار کای اسکور

[Downloaded from iieshrm.ir on 2026-11-05]

با هفت درجه آزادی برابر با ۹/۵۴ به دست آمد که معنی‌دار نیست و موجب عدم رد فرض صفر برازندگی می‌شود، اما از طرف دیگر مقدار نسبت مجذور خی به درجه آزادی (χ^2/df)، ۱/۳۶ برآورد شد که کمتر از ارزش پیشنهادی ۳ است. مقدار جذر برآورد واریانس خطای تقریب (RMSEA) نیز ۰/۰۱۹ به دست آمد که کمتر از ارزش بحرانی ۰/۰۵ است. دیگر شاخص‌ها نیز (CFI = ۰/۹۹۹ و NFI = ۰/۹۹۸) مقادیر قابل قبولی را نشان می‌دهد.

جدول ۹- شاخص‌های برازش و اعتبار (یافته‌های پژوهشگران)

مشخصه	برآورد
کای اسکور (df =7)	۹/۵۴۴
نسبت مجذور خی به درجه آزادی (χ^2/df)	۱/۳۶۳
جذر برآورد واریانس خطای تقریب (RMSEA)	۰/۰۱۹
شاخص برازندگی تطبیقی (CFI)	۰/۹۹۹
شاخص نرم شده برازندگی (NFI)	۰/۹۹۸

بحث و نتیجه‌گیری

هدف از این مطالعه، بررسی اعتبار سازه مرکز ارزیابی شرکت ملی نفت ایران با استفاده از روش‌های تحلیل چند صفتی- چند روشی و تحلیل عاملی تأییدی بود. داده‌های مربوط به ۱۰۲۲ مدیر که بر اساس چندین بعد شایستگی (رهبری، برنامه‌ریزی، توان نقد، پایش عملکرد، ارتباطات، شناخت مساله، نوآوری، تصمیم‌گیری، تفکر استراتژیک، تفکر تحلیلی و کلان نگری) و با چهار روش (تمرین گروهی، بحث گروهی، پرسش‌نامه و مصاحبه) در یک مقیاس ۹ درجه‌ای (۱ تا ۹) درجه‌بندی شده بودند از بانک اطلاعات شایستگی مدیران سازمان به‌صورت تصادفی استخراج شد و بر اساس روش چند صفتی-چند روشی و تحلیل عاملی تأییدی موردمطالعه قرار گرفت. تاکنون مطالعات متعددی درباره اعتبار سازه مراکز ارزیابی صورت گرفته که اغلب آن‌ها نشان داده‌اند که سازه‌های موردنظر، اندازه‌گیری نشده است (Brannick et al, 1989 ;Bycio et al, 1987 ;Russell & Domm, 1995 Sackett;

[Downloaded from iieshrm.ir on 2026-11-05]

& Harris, 1988). نتایج مطالعه حاضر مبتنی بر روش چندروشی-چندصفتی نیز هم راستا با نتایج مطالعات قبل، نشان داد میانگین کل هم‌بستگی بین درجه‌بندی‌های مربوط به یک بُعدِ واحد با روش‌های ارزیابی مختلف (۰,۳۲) در مقایسه با میانگین کل هم‌بستگی بین درجه‌بندی‌های ابعاد مختلف در همان روش (۰,۶۰)، کمتر است؛ بنابراین، به نظر می‌رسد، هرچند اندازه‌ها از اعتبار همگرایی متوسطی برخوردارند، اما اعتبار تشخیصیِ رضایت بخشی ندارند. نتایج این بخش از مطالعه حاضر همسو با نتایج برخی تحقیقات قبلی است. شاید بتوان دو دلیل برای نبود اعتبار تشخیصی ارائه داد؛ اول آنکه ممکن است رویه‌ها، شیوه‌های اجرا و دیگر عواملِ روش شناختیِ مراکز ارزیابی مقداری خطا که مانع اعتبارسنجی تشخیصی است، به خطای اندازه‌گیری بیفزاید (Gaugler et al, 1987; Jones, 1992 ; Lievens, 1998). برای نمونه، برخی از محققین معتقدند که شرایط کمپبل و فیسک را نمی‌توان جدا از درجه پایایی روش‌ها ارزیابی کرد. اگر نتایج اندازه‌گیری فاقد پایایی کافی باشد، در این صورت باید عدول مختصر از شرایط اعتبار تشخیصی را تحمل کرد. دوم اینکه، مراکز ارزیابی ممکن است سازه‌هایی غیر ازآنچه را که موردنظر بوده، اندازه‌گیری می کند(Lance, 2008; Russell & Domm, 1995 ; Raymark & Binning, 1997).

از سوی دیگر، نتایج تحلیل عاملی چهار بُعد (رهبری، برنامه‌ریزی، نوآوری، تصمیم‌گیری) و دو روش (تمرین گروهی و مصاحبه) مورد مطالعه با استفاده از یک طرح متقاطع، نشان داد که سهم ابعاد در توصیف تغییرات درجه‌بندی‌های صورت گرفته به مراتب بیش‌تر از سهم روش هاست، به طوری که متوسط این سهم برای دو روش مورد بررسی ۱۷/۵ درصد درحالی که برای چهار بُعد موردمطالعه ۴۶/۹ درصد به دست آمد. هم‌چنین، برآوردهای پارامتر ابعاد، تقریباً در همه موارد به میزان قابل توجهی بزرگ‌تر از برآوردهای پارامترهای روش‌ها بود، به طوری که میانگین بارهای عاملی برآورد شده برای دو روش و چهار بُعدِ مورد بررسی به ترتیب ۰/۳۹ و ۰/۶۸ بود. علاوه بر آن، هم‌بستگی درونی بین عامل‌های مکنون ابعاد، بین ۰/۳۵ تا ۰/۶۷ (میانگین ۰/۵۳) بود که شاهدی مبنی بر اعتبار تشخیصی ابعاد مورد اندازه‌گیری است. همان‌طور که مارش و گریسون (۱۹۹۵) ذکر کرده‌اند، بارهای عاملی بزرگ برای ابعاد نشان‌دهنده وجود اعتبار همگرا،

[Downloaded from iieshrm.ir on 2026-11-05]

بارهای عاملی بزرگ برای روش‌ها نشان‌دهنده اثرات روش اندازه‌گیری و هم‌بستگی‌های بالا بین ابعاد نشان‌دهنده فقدان اعتبار تشخیصی است؛ بنابراین چنین نتیجه گرفته می‌شود که ابعاد مورد مطالعه، عوامل تعیین‌کننده اصلیِ درجه‌بندی‌های مرکز ارزیابی هستند و اثر روش‌ها (تمرین‌ها) کمتر از آن هاست. چنین نتایجی در مطالعات دیگر نیز مشاهده شده است (Arthur & et al, 2000). به هرحال شاید بتوان یک دلیل تفاوت بین نتایج تحلیل هم‌بستگی و تحلیل عاملی را استفاده از طرح کاملاً متقاطع و تقلیل ابعاد و روش‌های مورد مطالعه در تحلیل CFA دانست. این امکان وجود دارد که محدود کردن ابعاد و تمرینات در تجزیه و تحلیل عاملی، تنوع در درجه‌بندی ارزیابان مرکز ارزیابی را کاهش داده است و از آنجایی که هر ارزیاب با استفاده از یک روش تمرین به درجه‌بندی افراد می‌پردازد، سبب کاهش واریانس روش‌ها نسبت به ابعاد شده است. بر پایـه شواهد مربوط به ابعاد و تمرین‌ها در ساختار درونی، پژوهش‌ها کم کم روی تفسیرهای چندوجهی‌تری تمرکز می‌کنند که در طراحی، نمره‌گذاری و تفسیرِ مراکز ارزیابی، هم به ابعاد توجه دارند و هم به تمرین‌ها (Hoffman, 2012 ;Putka & Hoffman, 2013).

البته در مطالعات دیگری تأثیر دیگر خصوصیات طراحی نیز بر روایی سازه مورد توجه قرارگرفته است. خصوصیات مشخصِ طراحی عبارتند از آموزش ارزیاب (Schleicher, Day, Mayes, & Riggio, 2002)، بهره‌مندی از روان‌شناسان در برابر مدیران به‌عنوان ارزیاب (Woehr, & Arthur, 2003)، شمار ویژگی‌های ارزیابی شده (Woehr & Arthur, 2003)، خصوصیات تمرین‌ها (Lievens, Chasteen, Day, & Christiansen, 2006) و هدف مرکز ارزیابی (Woehr & Arthur, 2003). بااینکه نتایج طراحی مرکز ارزیابی در بررسی‌های علمی تا حدی نامنسجم و اثرهای هر خصوصیت منفردِ طراحی نسبتاً ضعیف بوده است، چند الگوی منسجم و یکپارچه هم به چشم می‌خورد. مثلاً نشان داده شده است، مراکز ارزیابی‌ای که برای بالندگی به کاررفته‌اند (نه برای گزینش)، و در آن از روان‌شناسان و ارزیابانِ آموزش دیده استفاده شده (نه از مدیران)، قدرت تشخیص یا افتراق بیش‌تری میان ویژگی‌های مختلفی دارند که در یک تمرین درجه‌بندی می‌شوند (Lievens & Conway, 2001). افزون براین، (Lievens & et al, 2006) یک مطالعه علمی گسترده

[Downloaded from iieshrm.ir on 2026-11-05]

انجام دادند که نشان داد تمرین‌هایی که فعال‌کننده صفت‌های یکسانی بودند، هم‌بستگی بیش‌تری داشتند تا تمرین‌هایی که انتظار نمی‌رفت صفت‌های یکسانی را فعال سازند.

بر اساس این مطالعات و هم‌چنین یافته‌های این پژوهش، پیشنهاد می‌شود در مطالعات بعدی هردو تحلیل هم‌بستگی و تحلیل عاملی بر اساس طرحِ کاملاً متقاطع مورد مقایسه قرار گیرد. یک توضیح احتمالی برای تفاوت بین نتایج حاصل از مطالعه تحلیل عاملی این پژوهش با سایر مطالعات قبلی، مربوط به این واقعیت است که مراکز ارزیابی مستعد متفاوت از نظر عوامل توسعه‌ای و اجرایی هستند که ممکن است بر احتمال یافتن اعتبار همگرا و تشخیصی تأثیر بگذارد (Lievens, 1998). با توجه به این مسئله، استفاده از شیوه‌های حرفه‌ای برای افزایش توانایی ارزیابان در اندازه‌گیری سازه‌ها / ابعاد مورد علاقه در مرکز ارزیابی توصیه می‌شود.

منابع

مگنوسون، دیوید (۱۳۷۵). *مبانی نظری آزمون‌های روانی*، چاپ سوم، ترجمه محمدتقی براهنی، تهران: دانشگاه تهران.

Arthur, J., Woehr, D. J., & Maldegen, R. (2000). Convergent and discriminant validity of assessment center dimensions: A conceptual and empirical re-examination of the assessment center construct-related validity paradox. *Journal of Management*, 26, 813-835.

Arthur, W. R., Day, E., McNelly, T. L., & Edens, P. S. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. *Personnel Psychology*, 56, 125-154.

Binning, J. F. & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal of Applied Psychology*, 74: 478–494.

Bird, A. & Stevens, M. J. (2013). Assessing global leadership competencies. *In Global Leadership 2e* (pp. 125-152). Routledge.

Bowler, M. C. & Woehr, D. J. (2006). A meta-analytic evaluation of the impact of dimension and exercise factors on assessment center ratings. *Journal of Applied Psychology*, 91, 1114-1124.

[Downloaded from iieshrm.ir on 2026-11-05]

Bowler, M. C. & Woehr, D. J. (2009). Assessment center construct-related validity: Stepping beyond the MTMM matrix. *Journal of Vocational Behavior*, 75, 173-182.

Brannick, M. T., Michaels, C. E., & Baker, D. P. (1989). Construct validity of in-basket scores. *Journal of Applied Psychology*, 74, 957-963.

Bycio, P., Alvares, K. M., & Hahn, J. (1987). Situational specificity in assessment center ratings: A confirmatory factor analysis. *Journal of Applied Psychology*, 72, 463-474.

Cahoon, M. V., Bowler, M. C., & Bowler, J. L. (2012). A reevaluation of assessment center construct-related validity. *International Journal of Business and Management*, 7, 3-19.

Campbell, D. T. & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105.

Cappelli, P. & Keller, J. R. (2014). Talent management: Conceptual approaches and practical challenges. *Annual Review of Organizational Psychology and Organizational Behavior*, 1, 305–331.

Charoux, E. (1991). Assessment centers: A route to black advancement. *HRM*, 15 (3), 20-28.

Draganidis F. & Mentzas, G. (2006). Competency based management: A review of systems and approaches. *Information Management & Computer Security*. 14 (1), 51-64.

Gaugler, B. B. & Thornton, G. C. (1989). Number of assessment center dimensions as a determinant of assessor accuracy. *Journal of Applied Psychology*, 74, 611-618.

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., & Bentson, B. (1987). Meta-analysis of assessment center validity. *Journal of Applied Psychology,* 72, 493–511.

Greyling, L., Visser, D., & Fourie, L. (2003). Construct validity of competency dimensions in a team leader assessment centre. *SA Journal of Industrial Psychology*, 29 (2), 10-19.

Guenole, N., Chernyshenko, O. S., Stark, S., Cockerill, T., & Drasgow, F. (2013). More than a mirage: A large scale assessment centre with more dimension variance than exercise variance. *Journal of Occupational and Organizational Psychology*, 86 (1), 5-21.

[Downloaded from iieshrm.ir on 2026-11-05]

Harris, M. M., Becker, A. S., & Smith, D. E. (1993). Does the assessment center scoring method affect the cross-situational consistency of ratings? *Journal of Applied Psychology, 78,* 675-678.

Herd, A. M., Alagaraja, M., & Cumberland, D. M. (2016). Assessing global leadership competencies: The critical role of assessment centre methodology. Human Resource Development International, 19 (1), 27-43

Hoffman, B. J. (2012). Exercises, dimensions, and the great battle of Lilliput: A call for multifaceted interpretation of assessment center performance. In D. Jackson, C. Lance, & B. Hoffman (Eds.), *The psychology of assessment centers* (pp. 281–308). New York, NY:Routledge.

Hoffman, B. J., Melchers, K. G., Blair, C. A., Kleinmann, M., & Ladd, R. T. (2011). Exercises and dimensions are the currency of assessment centers. *Personnel Psychology*, 64, 351-395.

Hurst, D. & Charoux, J. (1994). The assessment centre: Testing the fairness hypothesis. *Journal of Industrial Psychology*, 20 (2), 21-25..

Jackson, D. R., Barney, A. R., Stillman, J. A., & Kirkley, W. (2007). When traits are behaviors: The relationship between behavioral responses and trait-based overall assessment center ratings. *Human Performance*, 20, 415-432.

Jones, R. G. (1992). Construct validation of assessment center final dimension ratings: Definition and measurement issues. *Human Resources Management Review*, 2: 195–220.

Joiner, D. A. (2000). Guidelines for ethical consideration for assessment center operations: International Task Force on Assessment Centre Guidelines. *Public Personnel Management*, 29 (3), 315-331.

Jokinen, T. (2005). Global leadership competencies: A review and discussion. *Journal of European Industrial Training*, 29 (3), 199-216.

Kleinmann, M., Kuptsch, C., & Köller, O. (1996). Transparency: A necessary requirement for the construct validity of assessment centres. *Applied Psychology: An International Review*, 45, 67-8.

Kleinmann, M. & Koller O. (1997). Construct Validity of Assessment Centers: Appropriate Use of Confirmatory Factor Analysis and Suitable Construction Principles. *Journal of social behavior and personality,* 12 (5), 65-84.

Klimoski, R. & Brickner, M. (1987). Why do assessment centers work? The puzzle of assessment center validity. *Personnel Psychology*, 40, 243-260.

[Downloaded from iieshrm.ir on 2026-11-05]

Lance, C. E. (2008). Why assessment centers do not work the way they are supposed to. *Industrial and Organizational Psychology*: Perspectives on Science and Practice, 1, 84-97.

Lance, C. E., Gatewood, R. D., Newbolt, W. H., Foster, M. S., French, N. R., & Smith, D. E. (1998). Assessment center exercise factors represent cross-situational specificity, not method bias. *Unpublished manuscript. University of Georgia*, Athens, GA.

Lance, C. E., Lambert, T. A., Gewin, A. G., Lievens, F., & Conway, J. M. (2004). Revised estimates of dimension and exercise variance components in assessment center postexercise dimension ratings. *Journal of Applied Psychology*, 89, 377-385.

Lance, C., Foster, R., Nemeth, Y., Gentry, W., & Drollinger, S. (2007). Extending the nomological network of assessment center construct validity: prediction of cross-situationally consistent and specific aspects of assessment center performance. *Human Performance*, 20, 345-362.

Lance, C. E. (2012). Research into task-based assessment centers. In D. R. Jackson, C. E. Lance, B. J. Hoffman (Eds.), *The Psychology of Assessment Centers* (pp. 218-233). New York: Routledge.

Lievens, F. (1998). Factors which improve the construct validity of assessment centers: A review. *International Journal of Selection And Assessment*, 6 (3), 141-152.

Lievens, F. & Conway, J. M. (2001). Dimension and exercise variance in assessment center scores: A large-scale evaluation of multitrait-multimethod studies. *Journal of Applied Psychology*, 86, 1202-1222.

Lievens, F., Chasteen, C. S., Day, E. A., & Christiansen, N. D. (2006). Large-scale investigation of the role of trait activation theory for understanding assessment center convergent and discriminant validity. *Journal of Applied Psychology*, 91, 247–258.

Marsh, H. W. & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications.* Thousand Oaks, CA: Sage Publications

Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology*, 93, 1042-1052.

[Downloaded from iieshrm.ir on 2026-11-05]

Monahan, E. L. (2011). *The search for the mythical assessment center dimension: Measurement artifact vs. substantive conclusion* (Unpublished master's thesis). University of Georgia, Athens, GA.

Nako, Z. C. (2011). *Construct validity of managerial assessment center*. Submitted in accordance with the requirements for the degree of M. A., University of South Africa.

Neidig, R. D. & Neidig, P. J. (1984). Multiple assessment center exercises and job relatedness. *Journal of Applied Psychology, 69,* 182-186.

Putka, D. & Hoffman, B. J. (2013). A closer look at reliable and unreliable sources of variance in assessment center ratings. *Journal of Applied Psychology*, 98, 114–133.

Povah, N. & Thornton, G., C. III. (Eds.). (2011). *Assessment centers and global talent management*. Farnham, England: Gower.

Raymark, P. H. & Binning, J. F. (1997). *Explaining assessment center validity: A test of the criterion contamination hypothesis.* Paper presented at the 1997 Academy of Management meeting, Boston, MA.

Russell, C. J. (1985). Individual decision processes in an assessment center. *Journal of Applied Psychology*, 70, 737-746.

Russell, C. J. & Domm, D. R. (1995). Two field test of an explanation of assessment center validity. *Journal of Occupational and Organizational Psychology,* 68: 25–47.

Sackett, P. R., & Dreher, G. F.) 1982 (. Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology,* 67: 401–410.

Sackett, P. R. & Harris, M. M. (1988). A further examination of the constructs underlying assessment center ratings. Recent developments in nontraditional employment testing |Special issue]. *Journal of Business Psychology, 3{2),* 214-229.

Salkind, N. J. (2010). Encyclopedia of research design Thousand Oaks, CA: *SAGE Publications*, Inc. doi: 10.4135/9781412961288

Schleicher, D. J., Day, D. V., Mayes, B. T., & Riggio, R. E. (2002). A new frame for frame of reference training: Enhancing the construct validity of assessment centers. *Journal of Applied Psychology*, 87, 735–746.

Schmitt, N. (1977). Interrater agreement in dimensionality and combination of assessment center judgments. *Journal of Applied Psychology*, 62: 171–176.

[Downloaded from iieshrm.ir on 2026-11-05]

Silverman, W. H., Dalessio, A., Woods, S. B., & Johanson, R. L. (1986) Influence of assessment center methods on assessor's ratings. *Personnel Psychology*, 39, 565-578.

Thornton, G. C. & Byham, W. C. (1982). *Assessment centers and managerial performance.* NY: Academic Press.

Thornton, G. C. & Rupp, D. E. (2006). *Assessment Centers in Human Resource Management. Strategies for prediction, diagnosis and development*. New Jersey: Lawrence Erlbaum Associates, Publishers.

Thornton G. C. III., Rupp, D. E., & Hoffman, B. J. (2015). *Assessment center perspectives for talent management strategies (Second edition).* New York: Routledge.

Theron, C. (2009). The diversity-validity dilemma: In search of minimum adverse impact and maximum utility. *South African Journal of Industrial Psychology*. 35 (1), 1-13.

Tomas, J. M., Hontangas, P. M., & Oliver, A. (2000). Linear confirmatory factor models to evaluate multitrait-multimethod matrices: The effects of number of indicators and correlation among methods. *Multivariate Behavioral Research*, 35, 469-499

Woehr, D. J. & Arthur, W. R. (2003). The construct-related validity of assessment center ratings: A review and meta-analysis of the role of methodological factors. *Journal of Management*, 29, 231-258.

Wood, L. A., Hoffman, B. J., & Twenge, J. M. (2012). *Job characteristics: The times they are a-changing*. Poster presented as an All-Conference. Top Poster at the 27th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.

[Downloaded from iieshrm.ir on 2026-11-05]

[Downloaded from iieshrm.ir on 2026-11-05]