

## مقایسه کانون ارزیابی با ارزیابی در الگوی کامل والگوی رایانه‌ای

حمیدرضا عریضی<sup>۱\*</sup>، هاجر براتی<sup>۲</sup>

۱- استاد گروه روانشناسی، دانشکده علوم تربیتی و روانشناسی، دانشگاه اصفهان، اصفهان، ایران.

۲- استادیار گروه روانشناسی، موسسه آموزش عالی المهدی مهر اصفهان، اصفهان، ایران.

پذیرش: ۱۳۹۵/۵/۱۶

دریافت: ۱۳۹۴/۱۰/۲

### چکیده

بسیاری از افراد در سازمان‌ها، باورهای مثبتی در مورد نقش رایانه در سیستم منابع انسانی مانند رعایت انصاف در گزینش و مصاحبه‌ها دارند. چنین باورهایی به زمینه‌هایی مانند ارزیابی در کانون ارزیابی نیز گسترش یافته است. با این وجود ارزیابی در کانون ارزیابی از رویه‌ای چند مرحله‌ای ایجاد می‌شود که با محو شدن این رویه در ارزیابی با رایانه، ممکن است ضعیف‌تر شود. هدف پژوهش حاضر مقایسه ارزیابی رایانه‌ای با الگوی سنتی شامل شش مرحله ارزیابی گروهی است. دو مطالعه مجزا از دو نمونه از مدیران ( $N_1=200$  و  $N_2=147$ ) سازمان‌های (خصوصی و دولتی) اصفهان گزارش شده است که به روش نمونه‌گیری در دسترس، انتخاب شده بودند. ابزارهای پژوهش عبارت بودند از پرسشنامه ارزیابی عملکرد ۳۶۰ درجه، ابزار رایانه‌ای و ابزار انتخاب اجباری و ارزیابان به دو دسته خبره و تازه‌کار تقسیم شده بودند. نتایج نشان داد که الگوی کامل ارزیابی انسانی بر الگوی رایانه‌ای برتری دارد و شاخص‌های پایایی و اعتبار آن نیز بهتر است. با توجه به نتایج و نیز اهمیت و کاربرد بالای کانون‌های ارزیابی، استفاده از ارزیاب‌های انسانی به جای ارزیابی رایانه‌ای توصیه می‌شود. ممکن است بهتر باشد نقش رایانه را به جای ارزیابی به استفاده در تمرین‌ها عوض کنیم.



واژه‌های کلیدی: کانون ارزیابی، الگوی کامل، الگوی رایانه‌ای.

## ۱- مقدمه

در سال‌های اخیر سازمان‌های کشور به اهمیت استفاده از کانون ارزیابی بیش از پیش پی برده‌اند، بنابراین استفاده از این ابزار در بسیاری از سازمان‌ها گسترش پیدا کرده است. ارزیابی در الگوی اصلی کانون ارزیابی از طریق ارزیاب‌های انسانی طی مراحل صورت می‌گیرد و در نهایت مجموع نظر ارزیابان به صورت نتیجه نهایی استنباط می‌شود. با این وجود برخی از کانون‌های ارزیابی داخلی، ارزیاب انسانی را حذف کرده‌اند و از رایانه به عنوان ارزیاب استفاده می‌کنند. سؤالی که اینجا مطرح است این است که آیا رایانه می‌تواند جایگزین آدمی شود؟ چرا که اگر رایانه بتواند در کانون ارزیابی جایگزین ارزیاب انسانی شود، احتمالاً کانون ارزیابی با سهولت بیشتر و شاید با هزینه کمتر اجرایی شود. در واقع مسئله پژوهش این است که در یک تمرین متجانس، ارزیابی از ابعاد آن که به وسیله الگوی کانون ارزیابی صورت می‌گیرد با ارزیابی رایانه‌ای چه تفاوتی از نظر ویژگی‌های روان‌سنجی دارد؟ هرچند در جامعه مدرن، ماشین بری از خطا دانسته شده و بهتر از خردجمعی ارزیابان در کانون ارزیابی تصور می‌شود [۱]؛ اما سؤال این است که آیا ارزیابی رایانه نسبت به جمع نظرات ارزیابان صحت و دقت یا به زبان روان‌سنجی پایایی و اعتبار بالاتری دارد؟

در طول سه دهه گذشته دو دیدگاه مقابل با یکدیگر در کانون ارزیابی (AC)<sup>۱</sup> رشد کرده است که یکی بر اهمیت بعد یا شایستگی‌ها در این ابزار همه جانبه منابع انسانی تأکید کرده است [۲] و دیگری بر تمرین‌ها و اهمیت آنها در پژوهش‌های مربوط به کانون ارزیابی اصرار ورزیده است [۳؛ ۴]. پژوهش حاضر متعلق به دیدگاه اول بوده و برای آنکه نقش عامل بعد و ارزیابی آن مشخص شود، تمرین ثابت نگاه داشته شده و فقط از یک تمرین (کشف حقیقت) استفاده شده است تا مقایسه‌ای بین ارزیابی از طریق مجموعه‌ای از ارزیابان در جریان گفتگوی تعاملی چند مرحله‌ای برای رسیدن به نتایج نهایی با نتیجه کانون ارزیابی رایانه‌ای مقایسه شود.



## ۱-۱- پیشینه تحقیق

رایانه در بسیاری از نظام‌های مربوط به مدیریت منابع انسانی از جمله نظام جانشینی [۵] یا گزینش کارکنان [۶؛ ۷] نقش مهمی یافته است. در برخی مواقع الگوهای روان‌سنجی با استفاده از رایانه تحول یافته است، برای مثال الگوی فنون نظریه سؤال - پاسخ (IRT)<sup>۱</sup> که در آن آزمون‌های هوش و استعداد رایانه‌ای کاملاً نسبت به آزمون‌های مداد - کاغذی برتری دارند (مانند تای، مید و سائو [۸] و بایلاری و منتظر [۹]). اما در مورد کانون ارزیابی مسئله کاملاً فرق می‌کند. کانون یا مرکز ارزیابی، نه یک مکان فیزیکی بلکه یک فرایند ارزیابی است [۱۰]. این رویکرد بر ارزیابی گروهی از شرکت‌کنندگان به وسیله تیم ارزیابان با استفاده از تمرینات و آزمون‌های متنوع دلالت دارد که برای انتخاب افراد واجد شرایط برای تصدی مناصب استفاده می‌شود [۱۱].

در کانون دو سطح تمرین‌ها و ارزیابی وجود دارد [۱۱] که ارزیاب باید از سطح دانش بالایی برخوردار باشد، برای مثال دانش روان‌شناختی و قوانین مربوط به درون‌گنجی طبقه‌ای یا شناخت اجتماعی و انواع سوگیری‌های مندرج در آن را بداند تا بتواند به درستی تصمیم بگیرد و با توجه به تنوع پاسخ‌ها در هر موقعیت بتواند رفتار را درون طبقات قرار دهد [۱۲]. در واقع رفتار آدمی به طبقات طبیعی باز می‌گردد که برخلاف طبقات مصنوعی به خوبی تعریف نشده است [۱۳] و بنابراین رایانه نمی‌تواند درون‌گنجی در طبقات را به خوبی انجام دهد. یک علت این است که دستورالعمل‌های داده شده به رایانه بر اساس منطق ارسطویی است. رایانه برای قضاوت برای طبقات مصنوعی ساخته شده است. طبقاتی که بشر تعریف کرده و دارای ویژگی‌های شناخته شده‌ای هستند. این طبقات در قالب الگوریتم‌ها تعریف می‌شوند مانند طبقه چهارضلعی‌ها که برای مثال مربع را می‌توان به سادگی درون آن قرار داد. ولی طبقات طبیعی دارای ویژگی‌هایی نیستند که

---

1. Item-Response Theory  
2. Tay, Meade & Cao  
3. Baylari & Montazer



تعریف شده باشند مانند طبقه میوه‌ها که با پیش‌نمونه<sup>۱</sup> تعریف می‌شود [۱۴] و بنابراین ارزیابی آن نمی‌تواند به وسیله رایانه انجام شود.

از طرف دیگر ارزیابی در کانون ارزیابی شامل مراحل ۱- مشاهده و ثبت رفتار؛ ۲- طبقه‌بندی رفتار؛ ۳- ارائه گزارش؛ ۴- تعیین امتیاز اولیه عملکرد کلی هر فرد در یک بعد؛ ۵- استخراج امتیاز نهایی افراد در هر بعد؛ ۶- امتیاز نهایی ارزیابی [۱۵] است که با گام‌های متعدد صورت می‌گیرد و کمک می‌کند رفتار از زوایای متعدد بررسی و ارزیابی شود. ارزیابی یکباره شرکت‌کننده به وسیله رایانه دارای این نقص عمده است که این نگرش همه‌جانبه وجود ندارد. در واقع رایانه بهتر از آدمی در نمره‌دهی یک مقیاس شخصیت عمل می‌کند [۱۶] زیرا می‌تواند همزمان با ورود نمره‌ها، محاسبات روان‌سنجی اعم از فنون کلاسیک یا نظریه سؤال- پاسخ را انجام دهد و پایایی را محاسبه کند. همچنین اثر هاله‌ای را با مقایسه پایایی درون و بین درجه‌بندی‌کنندگان مشخص کند، زیرا این دانش‌ها از نوع محاسباتی و سازگار با رایانه است. اما دانش شناخت و ادراک اجتماعی را ندارد و احتمالاً نتواند به خوبی ارزیاب انسانی عمل کند که دانش جامعی در این زمینه‌ها دارد [۱۷؛ ۱۸]. این انتقاد را می‌توان به بسیاری از ACهای امروز ایران وارد کرد که از این نرم‌افزارها سود می‌برند. هر ادعایی در مورد توانایی رایانه در انجام ارزیابی باید از این طریق صورت گیرد که نخست تمرین‌ها اجرا و سپس ارزیابی هم به وسیله مراحل کانون ارزیابی و هم به وسیله رایانه انجام شود، سپس تحلیل پایایی و اعتبار صورت گیرد تا مشخص شود کدام موفق‌تر عمل کرده‌اند. احتمالاً رایانه را بهتر بتوان در شبیه‌سازی سازمانی و اجرای تمرین‌ها به کار گرفت تا در ارزیابی. یک دلیل عمده آن این است که به طور مثال در ایفای نقش، هر تغییر کوچکی در بلندی صدا یا توانایی ایفاگر نقش یا یک لبخند غیرمحسوس می‌تواند بر نتیجه تأثیر بگذارد و رایانه کمک می‌کند که شرایط اجرا مشابه یا به اصطلاح روان‌سنجی تراز شده باشد. در این صورت می‌توان هزینه کانون ارزیابی را نیز کاهش داد، به عبارت دیگر استفاده از رایانه در کانون ارزیابی در جهت متغیر مستقل و نه متغیر وابسته باید صورت گیرد.



## ۲- روش

### ۲-۱- جامعه، نمونه و روش اجرای پژوهش

جامعه آماری پژوهش شامل تمام مدیران شرکت‌های صنعتی (شامل نزدیک به چهار هزار واحد صنعتی در شهرک‌های محمدآباد، جی، سجزی، اشترجان، مبارکه، منظریه، نجف‌آباد، محمودآباد، موجه‌خورت، شهرک بزرگ صنعتی اصفهان، سروش بادران و خمینی شهر)، مدیران مجموعه نفت (شامل پخش نفت اصفهان، گازرسانی، خطوط لوله گاز، پالایشگاه اصفهان و شرکت پتروشیمی اصفهان) و مجموعه نیرو (شامل مدیران شرکت توزیع برق اصفهان، شرکت آب اصفهان، تصفیه فاضلاب اصفهان، برق استان، برق منطقه‌ای اصفهان، نیروگاه برق شهید منتظری، نیروگاه اصفهان، و نیروگاه‌های هسا و زواره) و اداره‌های تعاون و کمرگ استان اصفهان بودند. دو نمونه مجزا از این افراد انتخاب و دو مطالعه بر روی آنها انجام شد. نمونه اول شامل تعداد ۲۰۰ نفر و نمونه دوم شامل ۱۴۷ نفر از مدیران مذکور می‌شد که به وسیله شرکت خدمات مهندسی و از طریق مدیر کل ارتباط با صنعت دانشگاه به مرکز تحقیقات دانشگاه اصفهان معرفی شده بودند. به منظور انجام این پژوهش از روش نمونه‌گیری در دسترس<sup>۱</sup> استفاده شد. نمونه‌گیری در دسترس را می‌توان به دو دسته نمونه‌گیری آسان و افراد داوطلب تقسیم کرد [۱۹] که نمونه پژوهش حاضر هر دو ویژگی را دارد. به طور کلی در نمونه پژوهش، ۲۲ درصد از مدیران شهرک‌های صنعتی، ۴۲ درصد از مدیران مجموعه نفت، ۲۱ درصد از مدیران مجموعه نیرو، ۸ درصد از مدیران اداره تعاون و ۷ درصد از مدیران گمرک انتخاب شده بودند. میانگین سنی آنان به طور کلی و به ترتیب ۳۱/۴۲، ۳۷/۶۸، ۳۵/۴۹، ۳۲/۲۶، و ۳۱/۸۷ با انحراف معیار ۷/۵۳، ۱۵/۴۱، ۱۲/۲۴، ۹/۸۳، و ۸/۲۹ بوده است. میانگین سنوات تحصیلی آنها نیز به ترتیب برابر ۱۳/۵۸، ۱۵/۲۹، ۱۷/۲۶، ۱۳/۶۹، و ۱۱/۶۳ با انحراف معیار ۴/۲۲، ۳/۱۱، ۳/۲۹، ۴/۳۶، و ۵/۱۱ بوده است. عضویت افراد در دو گروه به صورت تصادفی انجام شده است و تفاوت معناداری بین آنها از نظر سن و تحصیل وجود نداشته است.

---

1. Convenience Sampling



ارزیابان نیز گروهی از روان‌شناسان صنعتی و سازمانی بودند که در یک دوره دو ماهه مورد آموزش قرار گرفتند. در این دوره در مورد ارزیابی و تمرینات و شش مرحله اجرای کانون ارزیابی به آنها آموزش داده شد. در پایان برای نزدیک‌تر کردن دیدگاه آنان روش‌های چارچوب مرجع<sup>۱</sup> به آنها آموزش داده شد تا بتوانند به درستی ابعاد را شناسایی کنند. چون مبنای نظام ارزشیابی ابعاد تحلیل شغل بود؛ در زمینه ابعاد نیز افراد آموزش دیدند و سپس در مورد مشاهده‌ها و انواع مشاهده‌ها به‌خصوص مشاهده استنباطی و سوگیری‌ها و راه‌های غلبه بر آن، آموزش‌های لازم داده شد.

لازم به ذکر است با توجه به اینکه تعداد ۲۰۰ جلسه برگزار و در هر جلسه حداقل ۷ ارزیاب حضور داشته‌اند، در مجموع، ۱۴۰۰ نفر (که بیشتر آنها تکراری بوده‌اند) در مجموع جلسه‌ها حضور پیدا کردند. در اینجا سعی شد به عنوان لنگرگاه جلسه، افرادی که در همه دوره‌های آموزشی شرکت کرده بودند، همواره در جلسه حضور داشته باشند. دو گروه چهار و پنج نفری در بسیاری از جلسه‌ها حضور پیدا کردند، بنابراین داده‌های مربوط به آنان در تحلیل عاملی به صورت گروه‌های همگن تحلیل شده است. در جلساتی که بقیه افراد (۸ نفر باقیمانده) حضور داشته‌اند و معمولاً بین ۱ تا ۳ نفر ارزیابان مشترک بیشتر نبوده است به دلیل کم اثر شدن چارچوب مرجع از اصل تشکیل جلسه‌های آموزشی کانون ارزیابی دور شده و بنابراین فقط برای مدیریت داده‌ها آنها نیز تحلیل عاملی شده‌اند که در بخش یافته‌ها به آن اشاره خواهد شد. این بخش داده‌ها که یافته‌های آن در پژوهش حاضر و در جدول‌های ۵ و ۵ گزارش شده است، از دو مقیاس سود جسته است: ۱- در مقابل ابزار هنجاری از ابزاری با گزینه اجباری<sup>۲</sup> استفاده شده است. ابزار هنجاری یا همان طیف لیکرتی یا هنجار مرجع، ابزاری است که در آن همبستگی بین نمره‌ها (با اصلاح برای ضریب پایایی) واقعی است در صورتی که در گزینه باید این همبستگی‌ها کمتر از مقدار واقعی برآورد می‌شد [۲۰]. اما حسن مهم آن این است که در این حالت وانمودسازی<sup>۳</sup> و مطلوبیت اجتماعی<sup>۴</sup> کنترل می‌شود [۲۱]، چون

- 
1. Frame of Reference
  2. Ipsative Instruments
  3. Faking
  4. Social Desirability



شرکت‌کننده آزاد نیست تا به دلخواه خود (مانند طیف لیکرتی) در جهتی که می‌خواهد همه نمرات را بالا ببرد. ضریب اعتبار ملاکی در سنجش با گزینه اجباری پایین‌تر به دست می‌آید و اگر از روش‌های سنجشی یا آزمایشی ضعیف استفاده شود، افت در پایایی و اعتبار بیشتر است [۲۲] که می‌توان از آن به عنوان روشی برای مقایسه دقت‌های ارزیابی در دو الگوی کامل و ارزیابی در الگوی رایانه‌ای استفاده کرد.

## ۲-۲- ابزار: برای دستیابی به اهداف پژوهش از دو دسته ابزار استفاده شده است:

ابزار رایانه‌ای: برای مقایسه با رایانه از تمرینی مشابه در کانون ارزیابی در الگوی کامل استفاده شد. این تمرین کشف حقیقت است که می‌توان برای آن یک ابزار رایانه‌ای ساخت. برای محتوای تمرین، معادل‌های آن از پرسشنامه تحلیل سمت حرفه‌ای و مدیریتی (PMPQ) ساخته شد. هر آزمودنی در مقابل یک تمرین کشف حقیقت قرار می‌گرفت (برای مثال تمرین حل مسئله) و به این ترتیب ۲۲ تمرین کازیه برای ابعاد پردازش، برنامه‌ریزی، قضاوت، مهارت بین‌فردی و ویژگی شخصیتی ساخته شد که همگی از متن یک تمرین کشف حقیقت است و نزدیک‌ترین تمرین به بازی‌های تجاری است که امروزه در برخی از کانون‌های ارزیابی عرضه می‌شود. محاسبه و نمره‌دهی برای ثبت رفتار از طریق عامل انسانی به عنوان ورودی و همزمان در طبقه‌بندی مربوط قرار می‌گرفت. اما در اینجا دیگر از تعامل ارزیابان برای رسیدن به یک نتیجه نهایی خبری نبود و کامپیوتر محاسبه‌های خود را انجام می‌داد.

متغیر ملاک: برای سنجش متغیر ملاک، یعنی عملکرد از پرسشنامه ارزیابی عملکرد ۳۶۰ درجه استفاده شد. این پرسشنامه را نوری، عریضی، مولوی، کلانتری، نشاط‌دوست و همکاران در سال ۱۳۸۵ ساختند [۲۳]. پرسشنامه ۸۲ سؤال و ۹ زیرمقیاس شامل برنامه‌ریزی، تصمیم‌گیری، قضاوت، ارتباطات، پردازش اطلاعات، مهارت‌های فنی، روابط بین فردی، رهبری و کنترل دارد. پرسش‌ها با مقیاس ۵ درجه‌ای لیکرت (از همیشه تا به‌ندرت) به اضافه یک مقوله «نامرتب» برای هر پرسش تنظیم شده است. در پژوهش



نوری و همکاران [۲۳]، اعتبار محتوایی<sup>۱</sup> پرسشنامه تأیید شد. همچنین ضرایب پایایی (آلفای کرونباخ) برای ۹ قسمت این پرسشنامه به ترتیب برنامه‌ریزی ۰/۹۳۹، قضاوت ۰/۸۷۱، تصمیم‌گیری ۰/۸۹۳، ارتباطات ۰/۸۹۱، پردازش اطلاعات ۰/۹۳۳، مهارت‌های فنی ۰/۹۰، رهبری ۰/۸۸۲، روابط بین فردی ۰/۹۴۹ و کنترل ۰/۸۹۵ به دست آمد که در حد قابل قبولی است. دو نمونه از سؤال‌های این پرسشنامه از زیرمقیاس برنامه‌ریزی و تصمیم‌گیری به ترتیب عبارت است از «فعالیت‌های واحد خود را اولویت‌بندی می‌کند»، «در تصمیم‌گیری‌های مهم با افراد ذیصلاح مشورت می‌کند» [۲۴]. در پژوهش حاضر از سؤال‌های مربوط قسمت‌های تصمیم‌گیری و مهارت‌های فنی استفاده نشده است. علت استفاده نکردن از این دو تمرین، این مسئله بوده است که مهارت‌های فنی در مهارت‌های مدیریتی بالا ارزش کمی داشته و فقط در سطح پایین مدیریت (مانند مشاغل سرپرستی) اهمیت دارد و برای مشاغل سرپرستی به تحلیل شغل PMPQ افزوده شده است. حذف تصمیم‌گیری هم به دلیل آن بوده است که در پژوهش قدیمی [۲۵] تمرین کشف حقیقت برای مهارت تصمیم‌گیری به وسیله ۲۰ ارزیاب امتیاز کمی داشته و در دیگر تمرین‌های AC ترجیح داده شده بود. از آنجا که در مطالعه دوم، متغیر ملاک عملکرد سازمانی بوده است، از مقیاس‌بندی عملکرد در سازمان‌های مبدأ استفاده شده است. در اینجا اشکالی اساسی وجود دارد که سبک نمره‌دهی، مجموع نمرات و نمرات گرایش مرکزی و پراکندگی نمرات سازمانی در بین سازمان‌ها متفاوت بوده است. برای حل مشکل، تمام نمرات به نمرات تراز شده ( $z = \frac{x}{s} = \frac{x-\bar{x}}{s}$ ) تبدیل شد و نمرات استاندارد  $\sigma_x = 0/92$  و  $\mu = 0$  (میانگین z) ملاک عمل قرار گرفت.

ابزار ارزیابی در مطالعه دوم: از دو فرم گزینه پاسخ، یکی با طیف لیکرتی و دیگری با طیف گزینه اجباری هم در الگوی کامل و هم در مطالعه رایانه‌ای استفاده شده است. تفاوت این دو گزینه پاسخ آن است که در طیف لیکرتی آزمودنی در پاسخ به سؤال‌های آزاد است و نمره او در دامنه بی‌پایانی تغییر می‌کند. در حالی که در طیف گزینه اجباری، جمع نمرات آزمودنی‌ها مقدار ثابتی خواهد شد.

#### 1. Content Validity





افراد خیره و تازه کار: افراد خیره، افرادی هستند که در سه ملاک با افراد تازه کار متفاوت هستند. نخست آنها در سطح بالاتر (ارشد یا دکتری) بودند همزمان که دوره آموزش ارزیابی و از جمله بحث مفصل روان‌شناختی و سوگیری‌های شناخت اجتماعی و فنون ارزیابی و طبقه‌بندی رفتار را گذرانیده بودند. ضریب تاو کندال بین افراد خیره  $w=0/71$  و بین افراد تازه کار  $w=0/58$  بود. ضریب همبستگی کندال که با نماد  $w$  نشان داده می‌شود، یک آزمون ناپارامتریک است و برای تعیین میزان هماهنگی میان نظرات استفاده می‌شود. ضریب کندال بین ۰ و ۱ متغیر است. صفر، یعنی عدم توافق کامل و یک، یعنی توافق کامل وجود دارد.

### ۳- یافته‌ها

در جدول ۱ زیرمقیاس‌های ابعاد AC و مقایسه الگوی کامل با ارزیابی رایانه‌ای در این زیرمقیاس‌ها آورده شده است. اعتبار سازه در این پژوهش با دو روش در تحلیل عاملی (اکتشافی و تأییدی) و در دو الگو سنجش شده است. در جدول ۱ نتایج حاصل از تحلیل عاملی، عامل‌های ۲۲ گانه کانون ارزیابی در ارزیابی با الگوی کامل و رایانه نشان می‌دهد که عامل‌ها در ارزیابی با الگوی کامل دارای بار عاملی مناسب، یعنی بالاتر از ۰/۳ و دارای مقادیر ویژه مناسب، یعنی بالاتر از ۱ هستند [۲۶]. با این وجود بارهای عاملی برای همان عامل‌ها در ارزیابی رایانه‌ای کمتر شده است؛ هر چند همچنان مقدار مطلوب بالاتر از ۰/۳ را داراست. در سنجش دوم مقادیر ویژه نیز کمتر شده‌اند و حتی در یک مورد هم مقدار ویژه به پایین تر از ۱ رسیده است، به عبارتی اعتبار سازه روش ارزیابی رایانه‌ای کمتر از ارزیابی با الگوی کامل است. اعتبار سازه کیفیت کلی عملیات اندازه‌گیری است که در تبدیل سازه‌ها در فرضیه پژوهش به متغیرهای مورد بررسی به کار می‌رود، به عبارت دیگر این اعتبار به مواردی مربوط می‌شود که هدف دانستن این نکته است که تا چه اندازه می‌توان نمره‌های آزمون را به عنوان صفات یا خصوصیات نظری مشخص مورد تفسیر قرار داد [۱۹].



جدول ۱ زیر مقایسه‌های ابعاد AC، مقایسه مدل کامل با ارزیابی رایانه‌ای

الگوی رایانه‌ای ارزیابی			الگوی کامل ارزیابی				
مقدار ویژه	بار عاملی	عامل	مقدار ویژه	بار عاملی	عامل	ابعاد	ردیف
۱/۷۶-I	۰/۵۷	I	۲/۸۵	۰/۷۸	II	تسلط بر خود	۱
	۰/۵۱	I		۰/۷۴	II	سازگاری	۲
۱/۴۸-II	۰/۵۴	I		۰/۶۸	II	خطرپذیری	۳
	۰/۵۵	II		۰/۶۱	II	داشتن انرژی	۴
	۰/۴۷	II	۰/۸۴	I	مربی‌گری	۵	
۱/۴۱-III	۰/۵۲	II	۳/۸۲	۰/۷۸	I	مصاحبه‌گری	۶
	۰/۵۶	III		۰/۷۵	I	متقاعدسازی	۷
۱/۳۸-IV	۰/۴۹	II		۰/۶۹	I	آگاهی از محیط اجتماعی	۸
۱/۳۲-V	۰/۴۴	V		۰/۶۵	I	انعطاف رفتاری	۹
	۰/۴۳	IV		۰/۶۳	I	رهبری	۱۰
۱/۲۴-VI	۰/۵۲	III	۲/۶۱	۰/۷۵	III	حل مسئله	۱۱
	۰/۵۶	IV		۰/۷۱	III	سازماندهی	۱۲
۱/۲۱-VII	۰/۶۱	V		۰/۶۹	III	تفسیر	۱۳
	۰/۵۷	VII		۰/۵۹	III	تصمیم‌گیری	۱۴
۱/۱۸-III	۰/۶۲	VI	۲/۲۸	۰/۸۴	IV	بیان کلامی	۱۵
۰/۷۲-IX	۰/۵۲	IV		۰/۷۹	IV	سیالی نوشتار	۱۶
	۰/۴۷	III		۰/۷۶	IV	تراز سطح بیان	۱۷
	۰/۵۲	V	۱/۸۳	۰/۷۲	V	استدلال عددی	۱۸
	۰/۶۲	II		۰/۶۵	V	میان‌بر شناختی	۱۹
	۰/۴۰	II		۰/۶۸	V	پیش‌بینی	۲۰
	۰/۴۱	VI	۱/۴۳	۰/۸۱	VI	قضاوت و برنامه انسانی	۲۱
	۰/۵۹	VII		۰/۷۲	VI	قضاوت برنامه‌ها	۲۲



برای مقایسه الگوها شاخص‌های برازش در جدول ۲ آورده شده است. با توجه به اینکه مقدار  $\chi^2/df$  و دو پارامتری هستند که در بخش آماری مربوط به تحلیل عاملی تأییدی با هم یک الگو را تأیید می‌کنند، توصیه شده است تا از شاخص  $\chi^2/df$  استفاده شود. هرچقدر این شاخص کوچک‌تر باشد، به معنای تأیید بیشتر الگو است. قاسمی [۲۷] مقدار مطلوب برای آن را زیر ۵ دانسته است که تنها الگوی کامل عاملی دارای این ویژگی است. همچنین این شاخص در تمام ارزیابی‌های الگوی کامل وضعیت بهتری نسبت به ارزیابی رایانه‌ای دارد، همچنین برازش الگو در ارزیابی با الگوی کامل بهبودی یافته و شاخص‌های AGFI، NFI، GFI و CFI به سطح قابل قبول ۰/۹۰ رسیده‌اند [۲۷]، در حالی که در ارزیابی رایانه‌ای چنین وضعیتی ندارند و نیز میزان RMSEA و RMR نیز در ارزیابی با الگوی کامل در میزان مطلوب زیر ۰/۰۹ [۲۷] قرار می‌گیرد در صورتی که این شاخص‌ها در الگوی رایانه‌ای از حد مطلوب بالاتر هستند. بنابراین الگوی کامل صحیح است و اعتبار سازه آن تأیید می‌شود.

جدول ۲ مقایسه الگوی کامل برحسب تعداد ارزیاب‌ها با الگوی رایانه‌ای

شماره ارزیاب	$\chi^2/df$	GFI	CFI	NFI	AGFI	RMSEA	RMR	
۱	۴/۴۷	۰/۹۵	۰/۹۳	۰/۹۲	۰/۹۴	۰/۰۲	۰/۰۳	الگوی کامل
۲	۴/۱۹	۰/۹۳	۰/۹۵	۰/۹۰	۰/۹۱	۰/۰۴	۰/۰۵	
۳	۵/۲۷	۰/۹۲	۰/۹۴	۰/۹۰	۰/۹۰	۰/۰۳	۰/۰۴	
۴	۶/۴۲	۰/۹۱	۰/۹۳	۰/۹۰	۰/۹۰	۰/۰۴	۰/۰۳	
۵	۶/۷۱	۰/۸۹	۰/۹۱	۰/۹۰	۰/۸۸	۰/۰۶	۰/۰۴	
۶	۳/۹۶	۰/۹۲	۰/۹۰	۰/۹۲	۰/۹۴	۰/۰۳	۰/۰۲	
۷	۴/۰۸	۰/۹۴	۰/۹۱	۰/۹۱	۰/۹۲	۰/۰۴	۰/۰۳	
ارزیابی کلی	۸/۴۲	۰/۷۷	۰/۷۵	۰/۷۳	۰/۷۹	۰/۱۸	۰/۱۶	الگوی رایانه‌ای



در جدول ۳ ضریب اعتبار ملاکی مربوط به عوامل کانون ارزیابی در الگوی کامل آورده شده است. متغیر ملاک، عملکرد سازمانی بوده است که شامل ۷ بعد است. در اینجا مشاهده می‌شود که بیشترین همبستگی مربوط مهارت بین فردی و ارتباط بین فردی و نیز پردازش اطلاعات در کانون ارزیابی و پردازش اطلاعات در شغل است. همچنین رابطه ابعاد مختلف کانون ارزیابی با متغیر ملاک و ابعاد آن معنادار است و اعتبار ملاکی تأیید می‌شود. همبستگی گزارش شده با تصحیح همبستگی برای عدم پایایی ابزارها ( $\frac{r_{obs}}{\sqrt{r_{11}\sqrt{r_{11}}}}$ ) به دست آمد. بنابراین همبستگی‌های گزارش شده در جدول ۳ از همبستگی‌های مشاهده شده ( $r_{obs}$ ) بزرگ‌تر است با این حال، همبستگی بین عوامل کانون ارزیابی و برخی از متغیرهای ملاک بسیار بالاتر از حد انتظار به دست آمده است (برای مثال ضریب همبستگی بین مهارت‌های بین فردی و رهبری) که احتمالاً تا حدودی در آن سوگیری روش وجود دارد، زیرا هم متغیرهای پیش‌بین و هم متغیرهای ملاک بر مبنای ابعاد مقیاس تحلیل شغل PMPQ ساخته شده‌اند و واریانس روش آن را متورم ساخته باشد. به همین دلیل در نمونه دوم از ارزیابی عملکرد سازمانی در سازمان مبدأ استفاده شده است.

جدول ۳ ضریب اعتبار ملاکی مربوط به عوامل کانون ارزیابی (ملاک = عملکرد سازمانی) در

#### الگوی کامل

ردیف	عامل	برنامه‌ریزی	پردازش اطلاعات	قضاوت	روابط بین فردی	ارتباطات	رهبری	کنترل
I	شخصیت	۰/۲۸*	۰/۲۵*	۰/۲۷*	۰/۲۹*	۰/۲۶*	۰/۵۳**	۰/۴۱**
II	مهارت بین فردی	۰/۲۴*	۰/۲۸*	۰/۴۷*	۰/۵۹**	۰/۴۹**	۰/۵۶**	۰/۳۷**
III	برنامه‌ریزی	۰/۲۹*	۰/۴۴**	۰/۳۱**	۰/۱۱	۰/۱۷	۰/۳۲**	۰/۳۵**
IV	مهارت ارتباطی	۰/۱۸	۰/۱۵	۰/۲۲*	۰/۵۴**	۰/۴۱**	۰/۳۵**	۰/۳۵**
V	پردازش اطلاعات	۰/۳۸**	۰/۵۹**	۰/۳۷**	۰/۱۱	۰/۰۹	۰/۲۲*	۰/۱۷
VI	قضاوت	۰/۳۴*	۰/۳۹**	۰/۵۴**	۰/۳۶**	۰/۲۹*	۰/۳۷**	۰/۳۵**

\*\* p<0.01 \* <0.05

در جدول ۴ پایایی زیرمقیاس پژوهش در الگوی کامل (افراد خبره و تازه‌کار) و رایانه آورده شده است. در اینجا ملاحظه می‌شود که پایایی در الگوی کامل افراد خبره با هر دو روش طیف

لیکرت و گزینه اجباری در حد مطلوب است. اما این پایایی در ارزیابی الگوی کامل با افراد تازه کار و ارزیابی رایانه‌ای به‌ویژه در گزینه اجبار بسیار افت داشته و از حد مطلوب خارج شده است؛ یعنی پایایی رایانه مشابه ضریب پایایی ارزیابان تازه کار و بسیار کمتر از ارزیابان خبره است و این نشان می‌دهد که رایانه نتوانسته است با شایستگی‌ای در حد ارزیابان خبره که مسلح به دانش روان‌شناسی شناختی و طبقه‌بندی بر طبق پیش نمونه‌ها و شناسایی سوگیری‌ها در شناخت اجتماعی هستند، عمل کند. بنابراین ارزیابی رایانه‌ای به لحاظ پایایی نیز نسبت به الگوی کامل با افراد خبره بسیار ضعف دارد.

جدول ۴ پایایی زیر مقیاس پژوهش در الگوی کامل (افراد خبره و تازه کار) و رایانه در نمونه دوم پژوهش

ضریب پایایی رایانه		ضریب پایایی افراد تازه کار		ضریب پایایی افراد خبره		بعد
گزینه اجباری	طیف لیکرتی	گزینه اجباری	طیف لیکرتی	گزینه اجباری	طیف لیکرتی	
۰/۳۱	۰/۶۹	۰/۲۹	۰/۶۱	۰/۶۷	۰/۷۵	شخصیت
۰/۳۲	۰/۶۱	۰/۳۵	۰/۶۵	۰/۵۹	۰/۷۲	مهارت بین فردی
۰/۳۱	۰/۶۵	۰/۲۸	۰/۶۲	۰/۵۴	۰/۷۷	برنامه‌ریزی
۰/۲۸	۰/۵۴	۰/۳۳	۰/۶۵	۰/۶۵	۰/۷۳	مهارت ارتباطی
۰/۳۴	۰/۷۱	۰/۲۸	۰/۶۷	۰/۵۷	۰/۷۱	پردازش اطلاعات
۰/۲۷	۰/۵۲	۰/۳۱	۰/۵۹	۰/۶۳	۰/۷۰	قضات

در جدول ۵ ضریب اعتبار ملاکی مربوط به مطالعه دوم آورده شده است. متغیر ملاک، عملکرد سازمانی بوده است که شامل شش بعد می‌شود.



جدول ۵ اعتبار ملاکی (ملاک = عملکرد سنجش شده در سازمان اصلی) در نمونه دوم پژوهش

کضاوت (VI)	پردازش اطلاعات (VII)	مهارت ارتباطی (IV)	برنامه‌ریزی (III)	مهارت بین‌فردی (II)	شخصیت (I)		
۰/۲۷*	۰/۲۶*	۰/۲۵*	۰/۲۷*	۰/۲۱*	۰/۲۴*	طیف لیکرت	الگوی کامل
۰/۳۱*	۰/۳۸**	۰/۳۴**	۰/۳۷**	۰/۳۳*	۰/۳۵**	گزینه اجباری	ارزیابی
۰/۰۷	۰/۱۱	۰/۰۸	۰/۱۳	۰/۰۹	۰/۱۱	طیف لیکرت	الگوی رایانه‌ای
۰/۱۳	۰/۱۵	۰/۱۱	۰/۲۱*	۰/۱۷	۰/۱۹*	گزینه اجباری	

\*\*= $p < 0.01$  , \*= $p < 0.05$

همان‌طور که در جدول ۵ مشاهده می‌شود، بیشترین همبستگی‌ها مربوط به ابعاد عملکرد در الگوی کامل ارزیابی در گزینه اجباری است. همبستگی‌ها در ارزیابی الگوی کامل طیف لیکرت هم معنادار هستند و این نشان می‌دهد که اعتبار ملاکی الگوی کامل تأیید شده است. اما همین روابط، یعنی رابطه ارزیابی با الگوی رایانه‌ای کانون ارزیابی با متغیر ملاک و ابعاد آن جز در دو مورد شخصیت و برنامه‌ریزی در گزینه اجباری، معنادار نیست و اعتبار ملاکی تأیید نمی‌شود.

#### ۴- نتیجه‌گیری

از زمان داستان کودکانه پینوکیو آفریده کالوکلودی، این که انسان گاهی دلبسته مصنوعات خویش شده و به آن اعتماد می‌کند، موضوع تفکر بشر قرار گرفته است. این موضوع در مورد رایانه نیز صادق است و در بیشتر مواقع «رایانه» به‌خوبی توانسته است به این اعتماد پاسخ دهد، برای مثال زمانی که رایانه به جای توصیه‌ها و روابط، ملاک‌های تصمیم‌گیری گزینش را اعمال کند؛ محیط عادلانه‌تر خواهد بود. رایانه تحت تأثیر اثر هاله‌ای، خطاها و سوگیری‌هایی مانند آن قرار نمی‌گیرد و می‌تواند پردازش اطلاعات وسیع و پیچیده را با دقت زیاد انجام دهد [۲۸]. اما زمانی که بررسی‌هایی از قبیل کانون ارزیابی رخ می‌دهد، این تصور خاتمه پیدا می‌کند و توانمندی رایانه رخ می‌بازد. نتایج پژوهش حاضر نشان داد که رایانه در قضاوت و در بهترین حالت، تنها شبیه یک ارزیاب تازه‌کار عمل می‌کند و شاخص‌های پایایی و اعتبار ارزیابی‌های رایانه‌ای و ارزیاب‌های تازه‌کار بسیار کمتر از ارزیاب‌های حرفه‌ای بوده است. در تبیین یافته‌ها باید گفت که اعتماد به رایانه، از عملکرد آن در محیط‌هایی کاملاً متفاوت پدید



آمده است. تمامی فلوجارت‌ها و مبنای اولیه هر نرم‌افزار رایانه‌ای بر مبنای اصل طبقه‌بندی عمل می‌کنند که در هر مرحله باید مشخص کند که آیا عنصر خاصی مربوط به آن طبقه وجود دارد یا خیر؟ اما این درون‌گنجی طبقه‌ای مربوط به طبقات مصنوعی است و طبقات طبیعی را که به خوبی تعریف نشده‌اند، شامل نمی‌شود. تورنتون [۱۵] نیز با بیان یکی از مهم‌ترین خطاها که در مرحله دوم فرایند ارزیابی رخ می‌دهد، به این مسئله اشاره کرده است. رفتار انسانی که مبنای کانون ارزیابی است، نیز جزء طبقات طبیعی است. اما در مرحله دوم کانون ارزیابی فرایند تعامل ارزیابان خبره رخ می‌دهد و بحث‌هایی که انجام می‌گیرد کمک می‌کند تا به این «درون‌گنجی طبقه‌ای» از زوایای مختلف نگاه شود، بنابراین صحت و دقت ارزیابی‌های ارزیاب‌های انسانی خبره بالا می‌رود. مسئله دیگر «سطح خبرگی» در مفاهیم روانشناسی شناختی است که سبب می‌شود ارزیابان خبره در دام یا تله «سوگیری‌های شناختی» گرفتار نشوند. تورنتون [۱۵] در مرحله چهارم کانون ارزیابی به خبرگی اشاره مهمی کرده و تسلط بر مفاهیم روان‌شناسی اجتماعی و شناختی در مورد «شناخت اجتماعی» و به‌خصوص دانش مربوط به سوگیری‌ها را در قضاوت صحیح ارزیابان مهم دیده است. ارزیابان خبره به دلیل دانش و آگاهی نسبت به این سوگیری‌ها و شناخت اجتماعی، ارزیابی بهتری انجام می‌دهند. قدیمی [۲۵] در کانون ارزیابی تیم‌سازی مدیران متوجه شد که این سوگیری‌ها فقط در یک بحث مکفی و بسنده بین ارزیابان که به آن آگاهند، مشخص شده و افراد ناشی و تازه‌کار که در مورد این مفاهیم روان‌شناسی شناختی آموزش ندارند، در دام آن گرفتار می‌شوند. این دانش در رایانه‌هایی که به قضاوت رفتار آدمی می‌پردازند؛ وجود ندارد.

سبک‌های پاسخدهی شرکت‌کننده در کانون ارزیابی منعکس می‌شود. این سبک‌ها از جمله سبک وانمودسازی، سبک مطلوبیت اجتماعی و سبک تصدیق را می‌توان تا حدی از طریق گزینه اجباری کنترل کرد [۴]. در این پژوهش مشاهده شد که نمرات شرکت‌کنندگان در این مقیاس برای افراد ناشی و نیز رایانه دارای پایایی بسیار اندکی است. با این حال نباید به طور کلی کانون ارزیابی رایانه‌ای را زیر سؤال برد و آن را مورد تردید کلی قرار داد؛ به این معنا که ماشین و در اینجا رایانه نمی‌تواند جایگزین انسان به عنوان ارزیاب شود؛ اما ارزیاب‌ها هم باید برای ارزیابی‌های دقیق‌تر نقاط ضعف خود را برطرف کنند و بر نقاط قوت خود بیفزایند. رایانه را می‌توان در کانون ارزیابی و برای سنجش استعداد، هوش و شخصیت مدیران مورد استفاده



قرار داد [۳۰؛ ۲۹] ولی ارزیابی رفتار بهتر است با همان الگوی شش مرحله‌ای صورت گیرد. در این الگو، تعامل بین ارزیابی‌کنندگان اهمیت اساسی در ارزیابی دارد. در واقع ریشه‌های آن را می‌توان به یادگیری از خطا در جمع نسبت داد. پیازه دریافته بود که نمی‌توان با تصحیح خطا در دوران کودکی (برای مثال از ناحیه بزرگسال) رشد را تسریع کرد و به همین دلیل با برونر مخالف بود که خواهان شتاب‌بخشی رشد از طریق تصحیح خطا بود. پیازه فقط یک مورد را که یادگیری از همسالان در تصحیح خطا بود، مستثنا کرد [۳۱]. توافق روی نمره یک شرکت‌کننده در بین ارزیابان، در واقع همین جنبه یادگیری از سایر شرکت‌کنندگان را دارد.

یک سؤال اساسی این است که چرا برگزارکنندگان کانون‌های ارزیابی که در ایران از طریق نرم‌افزاری، فرایند ارزیابی را انجام می‌دهند به این خطاها واقف نمی‌شوند؟ در پاسخ باید گفت که یک پزشک هرگاه عمل جراحی خود را نادرست انجام دهد؛ بیمار می‌میرد یا بهبود پیدا نمی‌کند و اگر مهندس سازه‌ای پلی را ناستوار بنا کند، آن پل فرو می‌ریزد؛ به عبارت دیگر اعمال آنها در خودشان قضاوت شده و ملاک قضاوت به‌خودی خود (le soi) وجود دارد، در صورتی که این ملاک در علوم انسانی هر چند به‌خودی خود است؛ اما آشکاری آن نیاز به ملاک‌هایی بیرون از خود دارد. این ملاک‌ها برای مثال عملکرد واقعی افراد در محیط کار، در کانون ارزیابی، به عنوان متغیر ملاک مورد سنجش قرار نمی‌گیرد، تا نداشتن اعتبار ملاکی ضعیف آنها را آشکار سازد و این کار در مقاله حاضر انجام شده است.

محدودیت اصلی این پژوهش، برنامه نرم‌افزاری کشف حقیقت است که محدودیت یافته‌ها به تمرین کشف حقیقت و شایستگی‌های این پژوهش را در بردارد.

## ۵- منابع

[۱] عریضی حمیدرضا، خشوعی مهدیه‌السادات، نوری ابوالقاسم (۱۳۹۱) «کاربرد کانون ارزیابی و تحلیل شغل در تعیین شایستگی‌های مدیریتی»، *روانشناسی معاصر*، ۷(۱): ۸۵-۹۹.

[2] Ozpinar A., Civikgedey B. (2010) *Blue-collar employee selection with computer aided assessment center in manufacturing*; Information Technology





Based Higher Education and Training (ITHET), 9th International Conference on, April 29, pp.276 – 279.

- [3] Jackson D. J. R. (2012) "Task-based assessment centers: Theoretical perspectives"; In D. J. R. Jackson, C. L. Lance, & B. J. Hoffman (Eds.), *The psychology of Assessment Centers* (pp. 173–189), New York, NY: Routledge.
- [4] Jackson D. J. R., Stillman J. A., Atkins S. G. (2005) "Rating tasks versus dimensions in assessment centers: A psychometric comparison"; *Human Performance*, 18: 213–241.
- [5] راث ول، ویلیام. جی. (۱۳۸۴) *برنامه ریزی جانشینی مؤثر: حصول اطمینان از تداوم رهبری و پرورش استعدادها از درون سازمان؛ ترجمه احمدرضا حراف، چاپ اول، تهران: انتشارات شرکت ملی پتروشیمی ایران.*
- [6] Hawkes B. (2013) *Simulation technologies*; In M. S. Fetzter & K. A. Tuzinski (Eds.), *Simulations for Personnel Selection*, New York: Springer
- [7] Bruk-Lee V., Drew E., Hawkes B. (2013) *Applicant reactions*, In M. S. Fetzter & K. A. Tuzinski (Eds.), *Simulations for Personnel Selection*, New York: Springer.
- [8] Rupp D. E., Thornton G. C., Gibbons A. M. (2008) "The construct validity of the assessment center method and usefulness of dimensions as focus constructs"; *Industrial and Organizational Psychology*, 1: 116-120.
- [9] Baylari A., Montazer Gh. A. (2009) "Design a personalized e-learning system based on item response theory and artificial neural network approach"; *Expert Systems with Applications*, 36: 8013–8021.
- [۱۰] عریضی حمیدرضا، براتی هاجر (۱۳۹۱) «نقش واسطه‌ای پهنای نقش در رابطه بین عملکرد مدیران در کانون ارزیابی، توانایی های شناختی و خودپیروی در شغل با عملکرد شغلی»، *روانشناسی معاصر*، ۹ (۲) (پیاپی ۱۸): ۱۰۱ – ۱۱۲.
- [۱۱] بالانتاین ی. و پوان. (۲۰۰۴) *کانون‌های ارزیابی و توسعه مدیران؛ ترجمه مسعود سلطانی (۱۳۸۵)*، تهران: انتشارات آسیا.
- [12] Kramer S. (2014) "Computer-aided discovery and categorization of personality axioms"; *Journal of Logics and Their Applications*, 1(2): 107-133.



- [13] Hulton A. (2014) "Natural categories and human kinds: Classification in the natural and social sciences"; *International Studies in the Philosophy of Science*, 28(1): 102-105.
- [14] Jackson D. N., Wroblewski V. R., Ashton M. C. (2000) "The impact of faking on employment tests: Does forced-choice offer a solution?"; *Human Performance*, 13: 371-388.
- [۱۵] تورنتن، ج. سی (۱۹۹۲) *کانون‌های ارزیابی در مدیریت منابع انسانی*؛ ترجمه سعید جعفری مقدم، مجید سلیمی و سعید شهباز مرادی (۱۳۸۶)، تهران: مؤسسه انتشارات و چاپ دانشگاه تهران.
- [16] Tay L., Meade A.W., Cao M. (2015) "An overview and practical guide to IRT Measurement Equivalence Analysis"; *Organizational Research Methods*, 18(1): 3-46.
- [17] Meade A. (2004) "Psychometric problems and issues involved with creating and using measures for selection"; *Journal of Occupational and Organizational Psychology*, 77: 531-552.
- [۱۸] بارون رابرت، بیرن دان، برنسکامپ نایلا (۱۳۸۸) *روان‌شناسی اجتماعی*؛ ترجمه یوسف کریمی. تهران: نشر روان.
- [۱۹] بورگ، و.، گال ج.، گال م. (۱۹۹۶) *روش‌های تحقیق کمی و کیفی در علوم تربیتی و روان‌شناسی*؛ ترجمه احمدرضا نصر، حمیدرضا عریضی، محمود ابوالقاسمی، محمدجعفر پاکسرشت، علیرضا کیامنش، خسرو باقری و همکاران (۱۳۹۰)، ج ۱، تهران: انتشارات سمت.
- [20] Bartram D. (2005) "The Great Eight Competencies: A criterion-centric approach to validation"; *Journal of Applied Psychology*, 90: 1185-1203.
- [21] Christiansen N., Burns G.N., Montgomery G.E. (2005) "Reconsidering Forced-Choice Item Formats for Applicant Personality Assessment"; *Human Factors*, 18: 267-307.



- [22] Krause D. E., Anderson N., Rossberger R. J., Parastuty Z. (2014) "Assessment center practices in indonesia: An exploratory study"; *International Journal of Selection and Assessment*, 22(4): 384-398.
- [23] نوری، ا. عریضی، ح. مولوی، ح. کلانتری، م. نشاطدوست، ح. شفتی، ع. موسویان، ع. گرگانی، ح. و جنتیان، س. (۱۳۸۵) مهارت های روانشناختی مورد نیاز در مشاغل مدیریتی: حوزه مدیریت شرکت گاز استان اصفهان. گزارش نهایی پروژه تحقیقاتی معاونت پژوهشی دانشگاه اصفهان و امور پژوهش و توسعه شرکت ملی گاز ایران.
- [24] حقیقی عالیه، سماواتیان حسین (۱۳۹۰) «رابطه ویژگی های شخصیتی همکاران و زیردستان با نحوه ارزیابی آنان از عملکرد سرپرستان و مدیران، دانش و پژوهش در روانشناسی کاربردی»، ۱۲ (۴ (پیاپی ۴۶)): ۸۲-۹۰.
- [25] قدیمی فهیمه (۱۳۹۴) انتخاب و اجرای تمرین های کانون ارزیابی برای شایستگی تیم سازی مدیران شرکت پالایش و پخش فرآورده های نفتی ایران و تدوین الگویی بریا مجموعه شایستگی ها، پایان نامه کارشناسی ارشد روانشناسی صنعتی و سازمانی، دانشگاه اصفهان.
- [26] مولوی، حسین (۱۳۸۶) راهنمای عملی SPSS ۱۰-۱۳-۱۴ در علوم رفتاری، اصفهان: انتشارات پویا اندیشه.
- [27] قاسمی وحید (۱۳۹۲) مدل سازی معادله ساختاری در پژوهش های اجتماعی با کاربرد Amos Graphics، تهران: انتشارات جامعه شناسان.
- [28] Morgeson F. P., Campion M. A. (2010). *A framework of sources of inaccuracy in job analysis*; in M. Wilson, R. Harvey, G. Alliger and W. Bennett (Eds.), *The Handbook of Work Analysis: The Methods, Systems, Applications, and Science of Work Measurement in Organizations*, New York: Psychology Press/Taylor and Francis Group.
- [29] Khalidi M.A., Salter M.H. (2014) "Natural categories and human kinds: Classification in the natural and social sciences"; *The British Journal for the Philosophy Science*, pp. 1-7.

[30] Fetzer M., Tuzinski K. (2013) *Simulations for personnel selection*; New York: Springer.

[۳۱] سیف، علی اکبر (۱۳۸۶) روش‌های یادگیری و مطالعه؛ تهران: نشر دوران.