



BRANDAFARIN JOURNAL OF MANAGEMENT

Volume No.: 3, Issue No.: 40, Jul 2023

P-ISSN: 2717-0683 , E-ISSN: 2783-3801

Investigating employee performance evaluation systems in efficient organizations

Abbas Kangani

DBA graduate of Tarjoman Oloom Higher Education Institute

Abstract

Employee perceptions of the fairness of performance evaluations are critical to the success of any appraisal system. Research on performance management includes extensive studies on how bias emerges in the appraisal process. Despite this, there is no empirical evidence linking formal discrimination complaint filings – a key measure of bias - and performance appraisals. To close this gap, we conduct an empirical analysis using information on appraisal systems and discrimination complaints from the US federal government. Our findings suggest that agencies with better-designed appraisal systems will experience higher appraisal-related discrimination complaints, contrary to expectations. In particular, an expansion of training and voice opportunities are associated with an increase in appraisal discrimination complaints, while increased differentiation between performers is not consistently associated with complaints. For managers struggling with performance management, it may be better to spend time on improving performance-focused interpersonal communication rather than redesigning appraisal systems.

Keywords: Employee performance evaluation, rating system, efficient organization

بررسی سیستم های ارزیابی عملکرد کارکنان در سازمان های کارآمد

عباس کنگانی

دانش آموخته DBA موسسه آموزش عالی ترجمان علوم

چکیده

درک کارکنان از منصفانه بودن ارزیابی عملکرد برای موفقیت تمامی سیستم های ارزیابی مهم است. تحقیقات در مورد مدیریت عملکرد شامل مطالعات گسترده در مورد چگونگی پدیدار شدن سوگیری در فرآیند ارزیابی است. با وجود این، هیچ گونه شواهد تجربی در زمینه بررسی ارتباط شکایات های رسمی تبعیض (یک معیار کلیدی برای سوگیری) و ارزیابی عملکرد وجود ندارد. برای پر کردن این شکاف، یک تحلیل تجربی با استفاده از اطلاعات مربوط به سیستم های ارزیابی و شکایات تبعیض از دولت فدرال ایالات متحده انجام می دهیم. یافته های ما نشان می دهد سازمان هایی که سیستم های ارزیابی آنها کارآمدتر طراحی شده اند، برخلاف انتظارات، شکایات تبعیض بیشتری در مورد ارزیابی، تجربه خواهند کرد. به طور خاص، گسترش فرصت های آموزشی و ابراز عقیده با افزایش شکایات تبعیض در مورد ارزیابی همراه است، در حالی که افزایش تمایز بین مجریان ارزیابی همیشه با شکایات مرتبط نیست. برای مدیرانی که با مدیریت عملکرد سر و کار دارند، ممکن است بهتر باشد به جای طراحی مجدد سیستم های ارزیابی، برای بهبود ارتباطات بین فردی مبتنی بر عملکرد وقت بگذارند.

کلمات کلیدی: ارزیابی عملکرد کارکنان، سیستم رتبه بندی، سازمان کارآمد

سیستم‌های ارزیابی عملکرد کارکنان به‌عنوان ابزاری کلیدی در راستای مسئولیت‌پذیری کارکنان در بخش‌های دولتی، خصوصی و غیرانتفاعی عمل می‌کنند. ارزیابی‌ها اغلب اهداف متعددی را دنبال می‌کنند: ارائه مستندات و توجیه برای اقدامات آتی کارکنان، ایجاد انگیزه در کارکنان، شناسایی حوزه‌هایی که برای توسعه حرفه‌ای مورد نیاز است، و به منظور ابلاغ اولویت‌های سازمانی و موارد دیگر (دیلی، 2004؛ لوی و ویلیامز، 2004 مورفی و کلیولند، 1995). این اهداف چندگانه اغلب با هم تداخل دارند و منجر به تعارض بین فردی می‌شود.

در حالی که بخش خصوصی به طور فعال‌تر مزایای سیستم‌های ارزیابی عملکرد کارکنان را زیر سوال می‌برند، ولی اکثر سیستم‌های پرسنل دولتی به رتبه‌بندی عملکرد سالانه نیاز دارند. اصلاحات انجام شده در زمینه خدمات شهروندی توسط ایالت‌های آمریکا در دو دهه اخیر عمدتاً ناشی از این نگرانی است که دولت به ابزارهای قوی تری برای پاسخگو کردن کارکنان نیاز دارد (باتالیو و کاندری، 2009؛ کلو و سلدن، 2003). بسیاری از این اصلاحات بر تسهیل اخراج کارکنان دولتی و تقویت پیامدهای عملکرد قوی یا ضعیف تمرکز دارند. همچنین توجه فزاینده‌ای به ارزیابی عملکرد در دولت فدرال ایالات متحده شده است. در طی پانزده سال گذشته، دفتر مدیریت کارکنان ایالات متحده (OPM) استانداردهایی را برای ارزیابی قدرت سیستم‌های ارزیابی عملکرد (2006) ایجاد کرده است. اخیراً، OPM طرح "مدیریت عملکرد پلاس" را برای تمرکز بیشتر بر روابط بین مدیران و کارکنان، با تأکید بر بازخورد و مربیگری مداوم ارائه کرد (دفتر مدیریت پرسنل ایالات متحده 2016).

محققان در مورد چگونگی اندازه‌گیری اثربخشی سیستم‌های ارزیابی عملکرد اجماع کامل ندارند. انتخاب‌های بالقوه عبارتند از دقت رتبه‌بندی‌ها و مجموعه‌ای از واکنش‌های رفتاری، نگرشی و شناختی به رتبه‌بندی‌ها، مانند ادراکات منصفانه (لوی و ویلیامز، 2004؛ مورفی و کلیولند، 1995). در میان اقدامات بالقوه موجود، شکایات تبعیض مبنی بر رفتار سوگیرانه در فرآیند ارزیابی عملکرد می‌تواند به عنوان یکی از سیگنال‌های رفتاری برای ادراک منصفانه عمل کند. شکایات تبعیض آمیز در فرآیند ارزیابی می‌تواند مشکلات بی‌شماری از جمله رتبه‌بندی نادرست یا سوگیرانه، بکارگیری متناقض استانداردهای عملکرد و/یا تعارض بین فردی بین مدیر و کارمند را نشان دهد. بر این اساس، تحقیق ارائه شده در این مقاله این سوال را مطرح می‌کند: رابطه بین ساختار ارزیابی عملکرد و میزان ادعای تبعیض توسط کارکنان در فرآیند ارزیابی چیست؟

این مطالعه برای درک ارتباط بین ساختار ارزیابی و شکایت از تبعیض در ارزیابی، دو مجموعه داده از آژانس‌های فدرال ایالات متحده را ترکیب می‌کند. اطلاعات مربوط به ساختار ارزیابی آژانس از نمرات اختصاص داده شده توسط OPM به سیستم‌های ارزیابی آژانس که تحت چارچوب ارزیابی عملکرد و ابزار پاسخگویی (PAAT) درجه‌بندی شده‌اند، به دست می‌آید. نمرات PAAT با داده‌های مربوط به شکایات از تبعیض آژانس ترکیب می‌شود که طبق الزامات قانون اعلام و مقابله با تبعیض و پاداش کارمندان فدرال سال 2002 (قانون بدون ترس) به صورت عمومی گزارش می‌شود.

مقاله با مروری بر موضوعات مهم در تحقیقات مرتبط با ارزیابی عملکرد و سوگیری آغاز می‌شود. ما در ادامه، ادبیات مربوط به جنبه‌های خاص طراحی سیستم ارزیابی عملکرد را که ممکن است با ادراک سوگیری مرتبط باشد، بررسی می‌کنیم. پس از ارائه فرضیه‌ها، داده‌های ما با جزئیات بیشتری تشریح می‌شوند. سپس در ادامه مطالعه، تحلیل تجربی انجام خواهیم داد. ما با تأمل در مورد مفاهیم یافته‌های خود برای تحقیقات فعلی جمع‌بندی می‌کنیم و مراحل را برای مطالعات آینده ترسیم می‌کنیم.

مرور ادبیات

ادعای تبعیض در فرآیند ارزیابی یک رفتار هدفمند ناشی از برداشت منفی از ارزیابی است و به کارمند نشان می‌دهد که دلایلی مانند جنسیت، نژاد، سن، وضعیت معلولیت و غیره بر ارزیابی او تأثیر منفی داشته است. طبق ادبیات روانشناسی، این تصور از آسیب ممکن است ناشی از رتبه‌بندی نادرست و/یا سوگیرانه باشد. هر دو ممکن است یک کارمند را به ادعای تبعیض سوق دهند، که ادبیات موجود مطالعات دقیقی در مورد ماهیت مختلف هر یک ارائه می‌دهد.

دقت درک شده یک سیستم ارزیابی مبتنی بر فرآیند سنجش است که توسط مدیران برای یادآوری و مستندسازی اطلاعات عملکرد استفاده می‌شود (لندی، بارنز، و مورفی، 1978؛ لندی، بارنز-فارل، و کلیولند، 1980). این فرآیند شامل «تکرار ارزیابی، شناسایی اهداف، و آگاهی از عملکرد کارکنان و وظایف شغلی آنها می‌باشد» (لندی و همکاران، 1980، ص 355). محققان برای بهبود دقت در مستندسازی عملکرد، استراتژی‌های متعددی را شناسایی کرده‌اند که مدیران می‌توانند از آن استفاده کنند، از جمله نوشتن یادداشت‌هایی درباره عملکرد به‌طور منظم و ذخیره‌سازی آن‌ها به‌صورت سازمان‌یافته به طوری که به راحتی بتوان آن‌ها را برای مستندسازی ارزیابی پیدا کرد (فلدمن، 1981).

همچنین ارزیابی‌ها و تعاملات مکرر با کارکنان تضمین می‌کند که هم سرپرست و هم کارمند از اهداف یکدیگر آگاه هستند، که می‌تواند بر نمره ارزیابی نیز تأثیر بگذارد (وانگ، وانگ و کوانگ، 2010).

دومین دلیل کلیدی که یک کارمند ممکن است یک ادعای تبعیض در رابطه با ارزیابی عملکرد خود ارائه کند این است که او معتقد است رتبه بندی به دلیل برخورد سوگیرانه توسط سرپرست، نادرست است. تحقیقات فراوانی در مورد روش‌هایی که در آن سوگیری، به دلیل نژاد/قومیت، جنسیت، سن و عوامل دیگر رتبه بندی‌ها را تحریف می‌کند، وجود دارد (کاکس و انکومو، 1986؛ هیلن و ایگلی، 2008؛ لسللی، کینگ، برادلی، و هبل، 2008). مدل برازش کلیشه‌ای نشان می‌دهد که سوگیری در ارزیابی کارکنان یا سایر تصمیمات کارکنان، زمانی که ارزیاب‌ها کلیشه‌ها را به یک امتیاز فردی نسبت می‌دهند، تأثیر می‌گذارد (دیپو، 1985). به عنوان مثال، اگر یک ارزیاب به دلیل سفیدپوست بودن یک فرد او را سزاوار مدیریت بداند، به احتمال زیاد یک فرد سفیدپوست را به آن موقعیت ارتقا خواهد داد، حتی اگر یک اقلیت قومی واجد شرایط تر باشد. ارزیاب‌ها رتبه‌ها را تا جایی ارزیابی می‌کنند که تصورات آنها از رتبه با کلیشه شغلی منطبق است، که منجر به تبعیض می‌شود (دیپ بوی، 1985). لندی (2008) مدل برازش کلیشه‌ای دیپ بوی (1985) را بیشتر توسعه داد و خاطر نشان می‌کند که بسیاری از این کلیشه‌ها منفی هستند، به ویژه هنگامی که در مورد زنان و اقلیت‌ها به کار می‌روند. علاوه بر این، ماهیت ذهنی بسیاری از تصمیمات کارکنان، از جمله ارزیابی‌ها، امکان بروز سوگیری را فراهم می‌کند. علاوه بر این، سازمان‌ها اغلب فاقد شرایطی هستند که از تفکر مغرضانه جلوگیری می‌کند (لندی، 2008).

روش‌هایی وجود دارد که سازمان‌ها می‌توانند برای طراحی سیستم‌های ارزیابی خود انجام دهند تا با کاهش فرصت‌های عدم دقت یا سوگیری، ادراک منصفانه را تقویت کنند. این موارد در زیر به تفصیل مورد بحث قرار می‌گیرند و به دنبال آن فرضیه‌هایی برای این مطالعه ارائه می‌شوند. به عنوان مثال، تعیین شفاف عملکرد مورد انتظار به کارکنان اجازه می‌دهد تا بدانند چه رفتارهایی بیشتر ارزش دارد و چگونه باید برای دستیابی به آنها تلاش کنند. دخیل کردن کارکنان در طول فرآیند ارزیابی می‌تواند باعث افزایش سطح اعتماد کارکنان به رتبه‌بندی نهایی شود. آموزش مدیران و کارکنان در مورد الزامات فرآیند ارزیابی، و آموزش هنجارها و انتظارات بنیادی سیستم، می‌تواند به کاهش تعارض و افزایش اطمینان از بکارگیری مداوم استانداردها کمک کند. در نهایت، تشخیص تفاوت در سطوح عملکرد در بین کارکنان، موفقیت کارکنان قوی را تایید می‌کند. با این حال، هنگامی که به افراط کشیده می‌شود، تمایز بین کارکنان ممکن است منعکس کننده سوگیری ارزیاب باشد تا تفاوت‌های واقعی در عملکرد.

تعیین شفاف انتظارات

استانداردهای شفاف برای عملکرد با ادراک منصفانه بیشتر از سیستم ارزیابی مرتبط است (دیکسون، 1993؛ اردوغان، کرایمر، و لیدن، 2001؛ فولگر، کونوفسکی، و کروپانزانو، 1992). شفافیت زمانی افزایش می‌یابد که کارکنان بتوانند ببینند که چگونه کار آنها به تحقق اهداف سازمانی کمک می‌کند. مطالعات مکرر در نظریه انگیزش نشان می‌دهد که اهداف شفاف، اثرات انگیزشی بیشتری نسبت به اهداف مبهم دارند (لاتام، 2012). تعیین شفاف اهداف به ویژه در بخش دولتی می‌تواند چالش برانگیز باشد، اما همچنین بسیار مهم است، زیرا هدف مبهم با کاهش بهره‌وری سازمانی مرتبط است (چون و رابینی، 2005؛ پاندی و رایت، 2006؛ توبر و کالینز، 2000).

یکی از راه‌هایی که شرکت‌ها می‌توانند اهداف عملکردی شفاف‌تری را تعیین کنند، ایجاد یک «افق دید» است که به کارکنان نشان می‌دهد که چگونه عملکرد آنها با عملکرد برنامه مرتبط است. مشابه با تعیین اهداف شفاف، همسویی تلاش کارکنان با عملکرد سازمانی به کارکنان کمک می‌کند تا کار را اولویت بندی کنند و به دنبال دانش مورد نیاز برای دستیابی به اهداف باشند (هیت، بیرمن، شیمیزو و کوچار، 2001؛ لاک و لاتام، 2002). همسویی هدف زمانی اتفاق می‌افتد که اهداف و نیات افراد با اهداف و نیات سازمان همسو باشد (آیرز، 2013؛ روبین، 2015). با افزایش همسویی بین ارزیابی کارکنان و اهداف سازمانی، کارکنان رضایت بیشتری را از سیستم‌های ارزیابی نشان می‌دهند (دکرامر، اسمولدرز و وان در استراتن، 2013). این منجر به فرضیه زیر می‌شود:

H1: با افزایش شفافیت انتظارات عملکردی، میزان شکایات تبعیض ارزیابی کاهش می‌یابد.

مشارکت کارکنان

مدت‌هاست که مشارکت کارکنان در ارزیابی اهمیت دارد: "مشارکت کارکنان در ارزیابی یکی از عواملی است که به طور گسترده مورد تحقیق قرار گرفته است... تحقیقات نشان می‌دهد که اجازه دادن به کارکنان برای مشارکت در فرآیند ارزیابی با واکنش‌های مثبت کارکنان نسبت به سیستم ارزیابی مرتبط است" (کاویلی، کپیینگ، و لوی، 1998، ص. 616). مشارکت مفید است زیرا به

کارکنان درک بهتری از این موضوع که کدام یک از جنبه‌های عملکرد و استانداردهای مورد انتظار رتبه‌بندی می‌شوند، می‌دهد (سیلورمن و وکسلی، 1984).

نکته مهم این است که مشارکت می‌تواند در چندین نقطه از فرآیند ارزیابی رخ دهد. بسته به مطالعه، مشارکت به‌عنوان مشارکت در طراحی سیستم، تعیین هدف، خودارزیابی و/یا ارائه نمونه‌هایی از عملکرد به ارزیاب، مشارکت در گفتگوهای بازخورد یا اعمال نفوذ بر رتبه‌بندی نهایی عملیاتی شده است (کاولی و همکاران، 1998؛ گرینبرگ، 1986؛ رابرتز، 2003؛ رابرتز و رید، 1996). مدیرانی که کارکنان را در تعیین اهداف و انتظارات مشارکت می‌دهند، متوجه می‌شوند که کارکنان انتظارات را بهتر درک می‌کنند، اطمینان بیشتری برای دستیابی به انتظارات دارند و عملکرد بهتری دارند (رابرتز و رید، 1996). وقتی مشارکت همه‌جانبه باشد، مشارکت کارکنان در تعیین اهداف، بحث در مورد عملکرد واقعی، و ارائه مثال‌های کاری به مدیران برای اطلاع از رتبه‌بندی‌های ارزیابی، همگی با درک عدالت رویه‌ای از فرآیند ارزیابی مرتبط هستند (گرینبرگ، 1986؛ رابرتز و رید، 1996). در مجموع، فرضیه زیر را پیشنهاد می‌کند:

H2: با بهبود فرصت‌های مشارکت در فرآیند ارزیابی، میزان شکایات تبعیض ارزیابی کاهش می‌یابد.

آموزش

آموزش در مورد سیستم‌های ارزیابی عملکرد باید یکی از بخش‌های توسعه شایستگی‌های مدیریتی باشد. آموزش‌های ارزیابی ارائه شده به ارزیابان بیشتر از آموزش‌های ارزیابی ارائه شده به کارکنان مورد توجه قرار گرفته است. به طور خاص، آموزش ارزیابی ارائه شده به ارزیاب‌ها به طور کلی بر بهبود دقت رتبه‌بندی تمرکز دارد. یک متا‌آنالیز اخیر نشان می‌دهد که آموزش ارزیاب‌ها دقت ارزیابی را بیشتر از تغییرات در ابزار ارزیابی بهبود می‌بخشد، و متداول‌ترین تکنیک آموزش ارزیاب‌ها، آموزش چارچوب سنجش است (روچ، وهر، میشر، و کیشچینسکا، 2012). آموزش چارچوب سنجش از مثال‌ها، تمرین و بازخورد استفاده می‌کند تا به ارزیاب‌ها کمک کند تا درک مشترکی از استانداردهای عملکرد ایجاد کنند (آروی و مورفی، 1998؛ روچ و همکاران، 2012). این آموزش تضمین می‌کند که مدیران سطوح مشابهی از عملکرد را مطابق با رتبه‌های فردی ارزیابی می‌کنند، که نشان داده شده است که مولفه کلیدی ادراکات عدالت رویه‌ای در ارزیابی است (گرینبرگ، 1986). آموزش مدیران در ارزیابی عملکرد، علاوه بر بهبود دقت، رضایت از سیستم ارزیابی، حل تعارضات محل کار و کاهش تحریف رتبه‌بندی را افزایش می‌دهد (تیلور، تریسی، نرارد، هریسون و کارول، 1995).

محدود مطالعاتی که آموزش ارزیابی ارائه شده به کارکنان را بررسی می‌کنند، آموزش را به طور خاص به عنوان کمک به درک منصفانه بودن ارزیابی ارائه می‌کنند. این مطالعات استدلال می‌کنند که آموزش کارکنان به آن‌ها کمک می‌کند تا درک کنند که سیستم چگونه کار می‌کند، استانداردهای عملکرد بهتر را درک کنند، و بفهمند چه اطلاعات عملکردی برای ارائه به ارزیاب‌ها برای بهبود دقت رتبه‌بندی، مهم است (فولگر و همکاران، 1992). آگاهی از معیارهای ارزیابی عملکرد، از جمله منصفانه بودن سیستم ارزیابی کلی، رضایت از سیستم ارزیابی، و رضایت از رتبه، ادراک منصفانه بودن ارزیابی را افزایش می‌دهد (اردوغان و همکاران، 2001؛ تیلور و همکاران، 1995). در مجموع، این تحقیق در مورد آموزش ارزیابی برای مدیران و کارکنان نشان می‌دهد:

H3: با افزایش آموزش سیستم‌های ارزیابی عملکرد که در اختیار کارکنان و مدیران قرار می‌گیرد، میزان شکایات تبعیض ارزیابی کاهش می‌یابد.

تمایز بین اجرا کنندگان

اگر یک سازمان قصد استفاده از سیستم‌های ارزیابی خود را داشته باشد تا تصمیمات اداری مانند افزایش حقوق، ترفیع، یا نظم را اطلاع‌رسانی سازد، مهم است که رتبه‌بندی‌های اختصاص داده شده به کارکنان، به مدیران این امکان را ایجاد کند که بین سطوح عملکرد آنها تمایز قائل شوند. متأسفانه ده‌ها شواهد تجربی نشان می‌دهند که هنگامی که رتبه‌بندی‌ها برای آگاهی دادن به تصمیمات مدیریتی با پیامدهای مثبت یا منفی بسیار مهم استفاده می‌شوند، ارزیاب‌ها به طور کلی تمایل به افزایش رتبه‌بندی‌ها یا صدور رتبه‌بندی‌های نادرست دارند (جوهر و ویلیامز، 1997؛ مورفی و کیولند، 1995). ممکن است سازمان‌ها برای پرداختن به این موضوع، سیاست‌های سیستم ارزیابی را برای تشویق تمایز در هر دو رتبه‌بندی اختصاص داده شده و پیامدهای رتبه‌بندی‌ها انتخاب کنند. چنین سیاست‌هایی براساس نظریه‌های انتظار و عدالت بنا نهاده شده‌اند. نظریه برابری نشان می‌دهد که افراد تصمیمات ارزیابی را زمانی منصفانه‌تر می‌بینند که پیامدهای مثبت بیشتری برای عملکرد بالا دریافت کنند (بلمه، بالدوین و روبین، 2009؛ لا‌تام، 2012). نظریه انتظار نیز به همین شکل نشان می‌دهد که افراد عملکرد خود را با همتایان خود مقایسه می‌کنند. هنگامی که یک فرد

درک می کند که در سطح بالاتری عمل می کند، اما می بیند که همتای او با عملکرد پایین تر امتیاز و یا نتیجه مشابهی دریافت می کند، به این نتیجه می رسید که سیستم ارزیابی کمتر منصفانه است (بلوم و همکاران، 2009؛ لاتام، 2012)

ممکن است سازمان ها به توزیع رتبه بندی های عملکرد خود توجه داشته باشند زیرا ممکن است متوجه شوند که رتبه بندی ها به طور دقیق عملکرد را نشان نمی دهند. برای پرداختن به این موضوع، ممکن است سازمان ها یک سیستم رتبه بندی نسبی (بر خلاف سیستم رتبه بندی مطلق) و یا یک سیستم رتبه بندی توزیع اجباری (FDRS) را اجرا کنند. یک سیستم رتبه بندی نسبی از مدیران می خواهد تا کارمندان را در مقابل یکدیگر رتبه بندی کنند و اغلب آن ها را در فهرستهای منظم قرار دهند، در حالی که یک سیستم رتبه بندی مطلق عملکرد کارکنان را در برابر عناصر و استانداردهای از پیش تعیین شده ارزیابی می کند (روچ، استرنبورگ و کاپوتو، 2007). FDRS نیازمند این است که ارزیابان با دسته بندی کارکنان به دسته های عملکرد از پیش تعیین شده، آن ها را ارزیابی کنند (دومینیک، 2009). محققان در حال حاضر در مورد تاثیر مثبت و منفی این رویکردهای رتبه بندی بحث می کنند (همانطور که در مطالعات دومینیک، 2009؛ مون، اسکالز و لاتام، 2016 خلاصه شده است). در کوتاه مدت، رتبه بندی اجباری ممکن است انگیزه کارمند و عملکرد وظیفه را بهبود بخشد و ممکن است متقاضیان با عملکرد بالا را برای پیوستن به سازمان جذب کند زیرا آن ها می بینند که عملکرد بالا شناخته شده است (بلوم و همکاران، 2009). برعکس، رتبه بندی نسبی تحت طرح FDRS ممکن است منجر به ادراک عدالت رویه ای کارکنان در مقایسه با رتبه بندی مطلق در برابر استانداردهای از پیش تعریف شده شود (روچ و همکاران، 2007). سایر انتقادات حاکی از آن است که استفاده از رتبه بندی های اجباری و منصفانه کمتر دشوار و منصفانه است (شلايچر، بل و گرین، 2008) و منجر به رتبه بندی هایی می شود که دقت کمتری دارند (پارک، 2014).

علاوه بر سوق دادن مدیران به تغییر توزیع رتبه بندی ها، ممکن است سازمان ها انتخاب کنند که رتبه بندی ها را به تصمیمات مدیریتی مثبت یا منفی مرتبط سازند. برای به رسمیت شناختن مجریان قوی، سازمان ها معمولاً طرح های پرداخت برای عملکرد را در نظر می گیرند. متأسفانه طرح های پرداخت برای عملکرد در دولت به ندرت باعث افزایش مورد انتظار در انگیزه و عملکرد می شوند. در دولت فدرال آمریکا، پرداخت برای عملکرد انگیزه مدیر عمومی را افزایش نداد و عملکرد را در مطالعات متعدد بهبود نداد (پری، انگیرز و جون، 2009). تلاش های اخیر در دولت های ایالات متحده منجر به درک طرفداری و وعده های شکست خورده شده است (باتالیو و کاندیری، 2009؛ کلو و نیگرو، 2002). فراتر از تجربه بخش دولتی، استفاده از رتبه بندی ها برای تصمیمات اجرایی ممکن است منجر به رتبه بندی های مغرضانه در برابر یک گروه خاص شود. یک مطالعه نشان داد که رتبه بندی هایی که به زنان اختصاص داده می شود تنها زمانی مغرضانه هستند که از این رتبه بندی ها برای آگاهی از تصمیمات مربوط به پرداخت مزایا استفاده شود (دابینز، کاردی و تروکسیلو، 1988). از آنجا که تمایز بین اجرا کنندگان ممکن است منجر به ترجیح و یا تبعیض شود، ما فرضیه زیر را پیشنهاد می کنیم:

فرضیه 4. زمانی که تمایز بین سطوح عملکرد افزایش می یابد، نرخ ارزیابی شکایات تبعیض افزایش خواهد یافت.

داده ها و روش ها

دو مجموعه داده در سطح سازمان ادغام می شوند تا رابطه بین ساختار ارزیابی عملکرد و شکایات تبعیض ناشی از فرآیند ارزیابی را تجزیه و تحلیل کنیم. برای متغیرهای مستقل، داده های مربوط به سیستم های ارزیابی عملکرد نمایندگی فدرال توسط OPM به یکی از نویسندگان در سال 2011 ارائه شد. برای متغیر وابسته، اطلاعات مربوط به شکایات تبعیض نمایندگی از گزارش No FEAR Act نمایندگی محاسبه می شود. تحت این قانون، سازمان ها ملزم به گزارش عمومی بر مبنای سالانه تعداد شکایات تبعیض آمیزی هستند که در سال مالی دریافت کرده اند، از جمله اساس تبعیض (رفتار ناعادلانه به دلیل جنسیت، نژاد، سن و غیره شاکلی) و شیوه ای که در آن فرد معتقد است آسیب دیده است (فرد از ارتقا محروم شده بود، اگر به طور نامناسب تنبیه می شد، ارزیابی آن ها ناعادلانه بود و غیره). نهادها مجموع نرخ های شکایت را گزارش می کنند. ما نمی دانیم کدام کارمندان فردی شکایات را ثبت کرده اند و نمی توانیم تعیین کنیم که کدام بخش از شکایات تبعیض مربوط به ارزیابی، به تعصب در قالب تبعیض نژادی، جنسیتی، سنی و غیره متهم می شوند. در نتیجه واحد تحلیل، سازمان است.

متغیرهای مستقل

OPM با استفاده از ده معیار شناسایی شده در PAAT، سیستم های ارزیابی عملکرد نمایندگی را ارزیابی می کند. این معیارها عبارتند از: هم تراز، تمرکز نتایج، اقدامات معتبر، تمایز در سطوح عملکرد، پیامدها، مشارکت کارکنان، بازخورد و گفتگو و آموزش. OPM شامل دو معیار اضافی مربوط به مسئولیت هایش در سیستم های مدیریت کارکنان سازمان حساسی از جمله ارزیابی و راهنمایی سازمانی و نظارت و پاسخگویی است. براساس تعریف OPM و اندازه گیری هر معیار، متغیرهای مورد استفاده در این تحلیل براساس تناسب آن ها با نظریه ارائه شده در بالا انتخاب شدند.

هر معیار شامل سوالات/مولفه های متعددی است و OPM نقاط خاصی را به هر مولفه از هر معیار اختصاص می دهد. OPM به صورت ترکیبی، ۱۰ امتیاز را به هر یک از ۱۰ معیار اختصاص می دهد، به طوری که یک سازمان با یک سیستم ارزیابی عملکرد برجسته می تواند امتیاز بالایی از ۱۰۰ را کسب کند. OPM به عنوان سیستمی مشخص می شود که ۸۰ امتیاز یا بالاتر را از ۱۰۰ امتیاز ممکن کسب می کند.

سازمان برای ثبت تلاش های خود در مورد هر یک از این ده معیار، به پرسش های متعدد پاسخ می دهند و اسناد گسترده ای را برای OPM فراهم می کنند. OPM برای هر معیار براساس ارزیابی آن ها از شواهد ارائه شده توسط سازمان ها در مقابل معیارهای دقیق امتیاز تعیین می کند. مجموعه داده شامل نقاط اختصاص داده شده به هر جز از هر معیار است. به عنوان مثال، معیار همترازی شامل چهار مولفه است که هر کدام به طور جداگانه امتیاز دهی شده اند و مجموع نقاط بالقوه ماکزیمم برابر با ده امتیاز است. علاوه بر این، چندین ارزیاب مدیریت پروژه، موارد ارسالی PAAT را برای به دست آوردن امتیاز نهایی برای ارزیابی سیستم مورد ارزیابی قرار دادند. مجموعه داده ها شامل هیچ علامتی بر روی قابلیت اطمینان بین ارزیاب نمی شود.

همانطور که در مطالعه روین (۲۰۱۵) نشان داده شد، داده های PAAT محدودیت های زیادی دارند. اول اینکه استفاده از سیستم ارزیابی PAAT کاملاً داوطلبانه است و توسط خود سازمان آغاز می شود. بسیاری از سیستم های ارزیابی سازمان فدرال از طریق PAAT ارزیابی نشده اند و بنابراین در مجموعه داده در نظر گرفته نشده اند، در حالی که بقیه چندین بار تحت چارچوب PAAT ارزیابی شده اند. برای تجزیه و تحلیل انجام شده در اینجا، برای آن دسته از سازمان هایی که چندین بار ارزیابی شده اند، از جدیدترین امتیاز PAAT استفاده می کنیم. با این وجود با توجه به سن داده ها، این امکان وجود دارد که امتیازات مورد استفاده در این تجزیه و تحلیل نشان دهنده سیستم های فعلی مورد استفاده نباشد. مهم تر اینکه ماهیت داوطلبانه ارزیابی PAAT به این معنی است که ما نمی توانیم فرض کنیم که مشارکت در بازبینی PAAT یک فرآیند کاملاً تصادفی است. دوم اینکه مجموعه داده تنها شامل امتیازات اختصاص یافته توسط OPM برای معیار فردی است؛ مجموعه داده ها، سازمان های اسناد و مدارک گسترده ای که به OPM ارائه شده اند را شامل نمی شود.

متغیرهای مستقل از معیار PAAT براساس تناسب آن ها با نظریه و از طریق تحلیل عاملی تاییدی به دست آمدند. وضوح متغیر با محاسبه مجموع امتیازات کل از معیار PAAT در مورد (الف) هم تراز، (ب) تمرکز بر نتایج و (ج) اندازه گیری معتبر شکل گرفت. هم تراز درجه ای را در نظر می گیرد که طرح های عملکرد کارکنان یک خط دید بین اهداف فردی و اهداف سازمانی ایجاد می کند، در حالی که نتایج معیار تمرکز بر این موضوع است که آیا کارکنان در واقع مسئول سهم خود در دستاوردهای اهداف سازمانی هستند یا خیر (OPM، 2006). معیار سنجش معتبر درجه ای را در نظر می گیرد که برنامه ارزیابی تضمین می کند که معیارهای عملکرد برای کارمندان خاص، قابل اندازه گیری و عینی هستند. ما این موارد را با یک متغیر به نام وضوح ترکیب کردیم، زیرا کارکنان سازمان هایی که در هر سه حوزه موفق هستند باید درک بهتری از انتظارات برای عملکرد فردی خود داشته باشند.

مدل های تجربی، تلاش های سازمان برای ایجاد تمایز بین سطوح عملکرد با معیارهای چندگانه را به کار می گیرند. متغیر تمایزها به عنوان یک معیار جهانی تمایز، مجموع را از معیار OPM محاسبه می کند (الف) تمایزهای بین سطوح عملکرد و (ب) پیامدهای مبتنی بر عملکرد. از سازمان ها خواسته می شود تا توزیع رتبه بندی های عملکرد را برای آخرین دوره ارزیابی گزارش دهند، با نقاط بیشتری که به سازمان هایی با توزیع گسترده تر داده می شود. برای مستند کردن نتایج برای رتبه بندی های ارزیابی عملکرد، سازمان ها استفاده از نتایج مثبت و منفی ناشی از رتبه بندی ها، از جمله پاداش نقدی و زمان مرخصی قرار دادن کارمندان در دوره های بهبود عملکرد را گزارش می کنند. برای یک تحلیل دقیق تر، سه شکل مختلف تمایز عملکرد را نیز به صورت جداگانه در نظر می گیریم. به بیان دقیق تر، در یک مدل جایگزین، متغیر تمایز جهانی را حذف می کنیم و به جای آن سه متغیر را در نظر می گیریم: (الف) رتبه بندی هایی که از امتیاز OPM اختصاص داده شده به توزیع رتبه بندی های عملکرد استفاده می کنند، با یک توزیع گسترده تر

که نقاط بالاتر را دریافت می کند؛ (ب) پاداش هایی که از امتیاز OPM برای ارزیابی میزان ارائه انواع جوایز مثبت برای عملکرد بالا توسط سازمان ها استفاده می کنند؛ و (ج) پیامدهای عملکرد ضعیف که شامل امتیاز OPM PAAT برای پیامدهای رسمی تحمیل شده بر عملکرد ضعیف، با پیامدهای منفی بیشتر است که امتیاز سازمان را بیشتر می کند. جالب توجه است که در طرح امتیاز دهی کلی PAAT، سازمان ها امتیاز بیشتری برای ارائه شناخت مثبت کسب می کنند تا برای تحمیل پیامدهای منفی.

متغیرهای ابراز عقیده و آموزش با استفاده از یک معیار از PAAT برای هر متغیر به دست آمدند. متغیر ابراز عقیده به عنوان مجموع نقاط اختصاص داده شده توسط OPM برای معیار مشارکت کارکنان محاسبه می شود که درجه مشارکت کارکنان در طراحی کلی برنامه ارزیابی و سطح مشارکت کارکنان در تنظیم استانداردهای عملکرد فردی آن ها را در نظر می گیرد (OPM, 2006). متغیر آموزش نشان دهنده مجموع امتیاز اختصاص داده شده توسط OPM برای آموزش معیار آن است. این معیار از سازمان ها می خواهد تا گزارش دهند که آیا سیستم ارزیابی نیازمند آموزش کارمندان و مدیران در فرآیند ارزیابی است یا خیر و اینکه آیا آن ها در واقع در دو سال قبل از تکمیل PAAT آموزش دیده اند.

تحلیل عامل تاییدی با چرخش واریماکس ترکیب معیارهای PAAT را برای وضوح و تمایز متغیرها ارزیابی کرد و از ایجاد دو متغیر حمایت کرد. ساختار دو متغیره به ترتیب ۷۸٪ و ۳۱٪ از واریانس را توضیح می دهد. علاوه بر این، دو ساختار متغیر، تناسب قابل قبولی به دست آوردند ($p < .01$, $\chi^2(10) = 247.78$). متغیر وضوح ضریب پایایی مقیاس آلفای کرونباخ 0.92 را تولید کرد. متغیر پیامدها ضریب پایایی مقیاس آلفای کرونباخ 0.73 را تولید کردند. هر دو متغیر دارای درجه بالایی از سازگاری در مدل هستند. ما در ابتدا ترکیب معیارهای OPM (الف) بازخورد و گفتگو و (ب) مشارکت کارکنان در یک متغیر ابراز عقیده ترکیبی را در نظر گرفتیم، اما ساختار عاملی چنین ترکیبی را پشتیبانی نمی کند. در نتیجه ابراز عقیده با استفاده از مشارکت کارمند معیار واحد OPM ارزیابی می شود زیرا تعریف OPM از این معیار با نظریه مورد بحث در گذشته سازگارتر است.

متغیرهای وابسته

اطلاعات مربوط به شکایات تبعیض سازمانی از گزارش No FEAR Act سازمان، برای متغیرهای وابسته اندازه گیری می شود. از سال ۲۰۰۳، سازمان ها ملزم به گزارش سالانه مبانی و مسائل مربوط به شکایات تبعیض هستند. این گزارش ها شامل اطلاعاتی هستند از جمله تعداد شکایات تبعیض در مورد رفتار مغرضانه در فرآیند ارزیابی عملکرد، کل شکایات ثبت شده در طول سال و اینکه آیا کارمندان ادعا می کنند که آن ها براساس جنسیت، نژاد و غیره مورد تبعیض قرار گرفته اند (برای اطلاعات بیشتر به روبین و کلوگ، 2012 مراجعه کنید).

متغیر وابسته که از داده های No FEAR Act توسعه یافته، نسبتی است که با تقسیم تعداد شکایات تبعیض ارزیابی به تعداد کارمندان در آن سازمان اندازه گیری می شود. تعداد کارمندان با استفاده از آمار کارکنان سپتامبر از داده های OPM Fedscope جمع آوری شد. ما به جای تعداد شکایات تبعیض ارزیابی از یک نسبت استفاده می کنیم زیرا ۵۰ شکایت در یک سازمان کوچک مانند بنیاد علوم ملی به معنای چیزی بسیار متفاوت از یک سازمان بزرگ مانند اداره بهداشت سربازان قدیمی است. با هر دو داده سیستم شکایت و ارزیابی حاصل از گزارش های سازمان، سطح تجزیه و تحلیل ما سازمان فدرال است؛ در مجموع ۷۹ سازمان در نمونه آماری وجود دارد.

علاوه بر نسبت شکایات به کارمندان، دو متغیر وابسته جایگزین را برای ارزیابی تاثیر طراحی سیستم ارزیابی بر شکایات تبعیض مربوط به ارزیابی عملکرد توسعه می دهیم. این نسبت به عنوان یک متغیر ساختگی در نظر گرفته می شود که زمانی که یک سازمان هیچ شکایت تبعیض مربوط به ارزیابی را تجربه نکند نمره ۰ را دریافت می کند، در صورتی که هر گونه شکایت تبعیض مربوط به ارزیابی را دریافت کند، برابر با یک می شود. یک متغیر وابسته سوم توسعه داده شده است که در آن، در صورتی که یک سازمان هیچ شکایتی را تجربه نکرده باشد، نمره صفر را دریافت می کند؛ اگر نرخ شکایت از صفر بیشتر یا مساوی یا کمتر از میانه نرخ شکایت باشد، امتیاز ۱ خواهد بود؛ یا در صورتی که میزان شکایت بیش از حد متوسط باشد، نمره ۲ خواهد بود. براساس این متغیرهای مستقل، ما روابط فرض شده خود را با استفاده از (۱) رگرسیون حداقل مربعات معمولی (OLS) با استفاده از متغیر وابسته نسبت، (۲) یک مدل لجیت با متغیر وابسته ساختگی و (۳) یک مدل لجیت چند جمله ای با متغیر وابسته سه طبقه تحلیل می کنیم. حالت های لجیت و لجیت چند جمله ای با استفاده از احتمالات پیش بینی شده تفسیر می شوند.

در تحلیل ما، توصیف نحوه مدیریت زمان یک مسئله حیاتی است. سازمان ها PAAT را در سال های مختلف و در فواصل مختلف تکمیل کردند. همانطور که در بالا اشاره شد، این تجزیه و تحلیل تنها شامل آخرین امتیازات PAAT است. در این مدل ها، امتیازات

PAAT که در سال t توسعه داده شدند، با شکایات تبعیض در سال t+1 مرتبط بودند. در نتیجه تاریخ متغیرهای وابسته در میان سازمان‌ها متفاوت است، اما ساختار مشابهی را دنبال می‌کنند. برای مثال، امتیاز PAAT ۲۰۰۸ از اداره بزرگراه فدرال با داده‌های تبعیض از سال ۲۰۰۹ جفت می‌شود؛ داده‌های PAAT دفتر ثبت امتیاز و علائم تجاری ۲۰۰۹ با داده‌های تبعیض از سال ۲۰۱۰ جفت می‌شوند و غیره. در نتیجه این امر به ما اجازه می‌دهد تا از مشکلات مرتبط با تعصب منبع یکسان اجتناب کنیم و هر گونه رابطه را به صورت زیر تفسیر کنیم: یک مولفه ساختاری از سیستم ارزیابی در سال X باعث افزایش/کاهش در شکایات تبعیض در سال آینده می‌شود.

تجزیه و تحلیل

با بررسی آمار توصیفی متغیرهای کلیدی، تجزیه و تحلیل را شروع می‌کنیم (جدول ۱ را ببینید). سازمان‌های موجود در نمونه به طور متوسط حدود ۱۵۰۰۰ کارمند را شامل می‌شوند. همانطور که انتظار می‌رود، بخش بسیار کوچکی از کارمندان فدرال، شکایات تبعیض آمیزی را در ارتباط با فرآیند ارزیابی تنظیم می‌کنند. بیست و یک موسسه هیچ شکایت تبعیض آمیزی که به ارزیابی‌ها مربوط باشد را گزارش نکرده‌اند. جای تعجب نیست؛ این امر به سادگی به این معنی است که شکایات کارکنان حاکی از رفتار ناعادلانه در سایر بخش‌های فرآیند کارکنان است. آمار توصیفی در جدول ۱، به ویژه خطاهای استاندارد و تفاوت‌های بین مقدار میانگین و میانه متغیر وابسته نشان می‌دهد که این متغیر دارای چولگی بالایی است، با تعداد محدودی از مقادیر بسیار بالا. ما تصمیم گرفتیم که این مشاهدات را از نمونه حذف نکنیم به طوری که "انتخاب روی متغیر وابسته" و در نتیجه نتایج خود را تعدیل نکنیم. در عوض برای مدیریت این انحراف در متغیر وابسته، از خطاهای استاندارد قوی در همه مدل‌ها استفاده می‌کنیم. در مدل OLS همه متغیرها را با محاسبه لگاریتم طبیعی تبدیل می‌کنیم.

امتیاز کلی PAAT که توسط OPM تعیین می‌شود، می‌تواند حداکثر ارزش بالقوه ۱۰۰ امتیاز را داشته باشد که در آن OPM یک سیستم ارزیابی موفق را به عنوان سیستمی تعریف می‌کند که به امتیاز ۸۰ امتیاز یا بالاتر دست می‌یابد. در این نمونه، میانگین امتیاز PAAT برابر ۶۴ است. کمتر از نیمی از نمونه (۳۳ نمایندگان) امتیاز PAAT ۸۰ امتیاز یا بالاتر کسب می‌کنند. در امتیازات PAAT، تغییرات جالبی را بین جنبه‌های مختلف طراحی سیستم ارزیابی می‌بینیم. به طور متوسط، سیستم‌های سازمان در ارائه فرصت‌های ابراز عقیده قوی تر هستند، به عنوان مثال کارکنان در سراسر فرآیند ارزیابی در نظر گرفته می‌شوند. جای تعجب نیست که سازمان‌ها بیشتر در ایجاد تمایز بین کارمندان تلاش می‌کنند. تلاش‌های سازمان برای ایجاد تمایز به طور متوسط ۱۲ امتیاز از ۲۰ امتیاز را کسب می‌کند.

جدول ۱:

	میانگین	انحراف استاندارد	میانه	Min.	Max.	N
شکایات ارزیابی	.0008	.0009	.0005	0	.0053	79
شفافیت و وضوح	17.620	9.110	19	0	30	79
ابراز عقیده	7.614	2.431	8	0	10	79
آموزش	6.237	2.910	7	1	10	79
تمایزها	11.536	3.752	12	0	17	79
رتبه بندی ها	5.592	2.273	6	0	8	79
پاداش ها	3.163	1.044	3	0	4	79
پیامدهای عملکرد ضعیف	1.740	.649	2	0	2	79
مجموع امتیاز PAAT OPM	63.883	22.108	71.4	12	93	79
کارمندان آژانس	15,051	33,568	3,106	46	251,392	79

جدول 2: همبستگی

آموزش	ابراز عقیده	تمایزها	شفافیت و وضوح	شکایات ارزیابی
شکایات ارزیابی				1
شفافیت و وضوح			1	.045
تمایزها		1	.492*	.019
ابراز عقیده	1	.381*	.633*	.151
آموزش	.680*	.295*	.668*	.243*

نکات: همه همبستگی ها با متغیرهای تبدیل شده با استفاده از لگاریتم طبیعی محاسبه می شوند. * $p < .05$.

تحلیل های دو متغیره حداقل پشتیبانی را برای فرضیه ها فراهم می کنند (جدول ۲ را ببینید). تفاوت میانگین آزمون نشان می دهد که در نرخ شکایت بین سازمان هایی که طبق استانداردهای ارزیابی OPM رتبه بندی شده اند، و سازمان هایی که هدف OPM را ۸۰ امتیاز در PAAT برآورده می کنند یا از آن فراتر می روند، تفاوتی وجود ندارد ($t = -.834$). این همبستگی ها نیز نشان دهنده نقاط ضعف در فرضیه ها هستند. به بیان دقیق تر، شکایات ارزیابی متغیر وابسته به ازای هر کارمند، همبستگی مثبت و معناداری را تنها با نمره آموزش نشان می دهد. این در جهت مخالف فرض شده است، اما نشان می دهد که وقتی کارکنان سیستم ارزیابی را درک می کنند، زمانی می توانند بهتر قضاوت کنند که با آن ها ناعادلانه رفتار شده است.

جدول 3: مدل های رگرسیون OLS

	Model 1 OLS	Model 2 OLS	Model 3 OLS
Clarity, ln	.000 [.000]	.000 [.000]	.000 [.000]
Voice, ln	.000 [.000]	.000 [.000]	.000 [.000]
Training, ln	.001** [.000]	.001** [.000]	.001** [.000]
Distinctions, ln		.000 [.000]	
Awards, ln			-.001* [.000]
Rating distrib, ln			.000 [.000]
Conseq for low perf, ln			.000 [.000]
Adjusted R ²	.05	.04	.03
n	79	79	79
F-Statistic	3.38	3.16	3.51
Probab > F	.022	.019	.004

نکات: خطاهای استاندارد قوی Huber/White در پراتر ارائه شده است.

* $p < .05$; ** $p < .10$.

نتایج مدل های OLS بسیار کم هستند؛ آن ها تنها مقدار کمی از تغییرات را بررسی می کنند (جدول ۳ را ببینید). ما سه مدل OLS مختلف را به طور تجربی ارزیابی می کنیم. مدل ۱ رابطه بین نسبت شکایات تبعیض ثبت شده و شفافیت، ابراز عقیده و آموزش را ارزیابی می کند. این سه متغیر مستقل، ویژگی های کلیدی سیستم ارزیابی را قبل از تخصیص رتبه ها نشان می دهند و در نتیجه، هر رابطه ای با متغیر وابسته می تواند به راحتی به عنوان یک نتیجه عدالت رویه ای توصیف شود. مدل ۲ معیار تمایزات جهانی، پیامدها، برای ارزیابی تاثیر اضافه شده پیامدهای سیستم ارزیابی بر شکایات تبعیض را اضافه می کند. در نهایت مدل ۳ نتایج را کاهش می دهد و در عوض شامل سه معیار دقیق ر پاداش، رتبه بندی و نتایج برای اجرا کنندگان ضعیف است. برای ارزیابی این که آیا این سه متغیر تمایز به طور مشترک برابر با صفر هستند یا خیر، از یک آزمون والد استفاده می شود، یعنی آیا آن ها براساس مدل ۱ بهبود می یابند. ما علاوه بر این دو آزمون نسبت احتمال را نیز انجام می دهیم تا تعیین کنیم آیا مدل ۲ بهتر از مدل ۱ است و آیا مدل ۳ با جزئیات بیشتر بهتر از مدل ۱ است.

متغیر آموزش تنها متغیری است که در هر سه مدل معنادار است، در حالی که در دسترس بودن پاداش تنها در مدل سوم معنادار است. در مدل ۲، آموزش مهم باقی می ماند حتی زمانی که پیامدها در مدل لحاظ می شوند. نتایج مدل ۳ مشابه با مدل ۱ است حتی زمانی که در برابر توزیع رتبه بندی های عملکرد، جوایز و نتایج برای اجرا کنندگان ضعیف اندازه گیری می شود. در هر سه مدل، با افزایش

آموزش، احتمال شکایت ارزیابی یک کارمند تقریباً یک درصد افزایش می یابد. در مدل ۳، یک درصد افزایش در پاداش های مربوط به عملکرد با یک درصد کاهش در شکایات تبعیض مربوط به ارزیابی همراه است. علیرغم مربعات R تعدیل شده کوچک هر یک از این سه مدل، آماره F نشان می دهد که متغیرهای مستقل در پیش بینی شکایات تبعیض مربوط به ارزیابی بهتر از شانس تصادفی هستند. آزمون های پس از تخمین نشان می دهند که چند خطی بودن در هیچ یک از سه مدل مشکلی ندارد (در مدل 1، $VIF = 2.12$ ، در مدل 2، $VIF = 2.02$ ، در مدل 3، $VIF = 2.31$). مدل ها توزیع نرمال را با کشش 0 نشان می دهند. دو آزمون، نسبت درستنمایی مدل 1 را با مدل 2 و مدل 1 را با مدل 3 مقایسه کردند. نتایج آزمون نشان می دهد که هیچ یک از مدل ها بهتر از بقیه نیست (مدل 3 در مقابل مدل 1: $p > \chi^2 = .560$ (مدل 2 در مقابل مدل 1: $p > \chi^2 = .9490$)).

به عنوان یک روش جایگزین برای ارزیابی رابطه بین نرخ شکایت تبعیض ارزیابی و طراحی سیستم ارزیابی، ما متغیر وابسته را به سه دسته تقسیم کرده و برای در نظر گرفتن احتمال نهادی که سطوح مختلفی از شکایات تبعیض مربوط به ارزیابی را تجربه می کند، از تحلیل لجیت چند جمله ای استفاده می کنیم. رده پایه صفر است، در نتیجه این مدل احتمال مشاهده نرخ های شکایت بالا یا پایین در مقابل عدم شکایت را ارزیابی می کند. همانطور که قبلاً اشاره شد، متغیر وابسته با استفاده از میانه بازایی شد، به طوری که در صورتی که از سازمان شکایت ارزیابی دریافت نشود برابر با صفر است، امتیاز یک به عنوان نرخ شکایت پایین است که بیشتر از صفر اما کمتر یا مساوی با میانه تعریف می شود، و دسته دو به عنوان یک نرخ شکایت بالا، با نرخ شکایت بالاتر از میانگین تعریف می شود. بر اساس کد گذاری مجدد مبتنی بر میانه، ۲۷ درصد از سازمان ها در نمونه هیچ شکایتی ندارند، درصد سازمان های با شکایت کم، ۳۳ درصد و سازمان های با شکایت زیاد، ۴۱ درصد است. به خاطر سازگاری، مدل لجیت چند جمله ای شامل تمام متغیرهای مورد استفاده در مدل ۳ برای ساده کردن بحث در مورد نتایج است (به مدل های ۴ الف و ۴ ب در جدول ۴ مراجعه کنید). مدل لجیت چند جمله ای، آمار والد ۲۶.۹۴ و مربع R تعدیل شده 0.26 را نشان می دهد.

جدول 4. مدل های لاجیت و لاجیت چند جمله ای.

	Model 4a	Model 4b	Model 5
	لاجیت چند جمله ای	لاجیت چند جمله ای	لاجیت
	شکایت صفر در برابر شکایت کم	شکایت صفر در برابر شکایت زیاد	
شفافیت	.013 [.063]	-.043 [.048]	-.033 [.046]
ابراز عقیده	.124 [.191]	.133 [.170]	.165 [.156]
آموزش	.293 [.206]	.363** [.143]	.350** [.144]
پاداش ها	-1.073* [.624]	-.826** [.358]	-.842** [.386]
توزیع رتبه بندی	.344 [.263]	.374* [.206]	.362* [.201]
عواقب عملکرد ضعیف	-.834 [.769]	-.581 [.691]	-.522 [.660]
n	79	79	79

نکات: خطاهای استاندارد قوی Huber/White در پرانتز ارائه شده است.

* $p < .05$; ** $p < .10$.

احتمالات پیش بینی شده برای مدل های لجیت چند جمله ای به ما اجازه می دهد تا احتمال مشاهده یک نتیجه خاص (شکایات صفر، کم، یا بالا) را زمانی که متغیرهای مستقل در مقادیر خاص تنظیم می شوند، تعیین کنیم (جدول ۵ را ببینید). یک سازمان با یک سیستم ارزیابی متوسط دارای احتمال ۰.۵۶۱ تجربه شکایات تبعیض مربوط به ارزیابی بالا، و احتمال ۰.۲۰۹ تجربه شکایات تبعیض مربوط به ارزیابی صفر است. در مقایسه، یک سازمان با یک سیستم ارزیابی قوی دارای یک احتمال ۰.۰۷۲ از تجربه شکایات صفر در مقایسه با احتمال ۰.۶۴۴ از تجربه شکایات بالا است. هنگام مقایسه احتمالات، احتمال دریافت شکایات بالا به جای شکایات صفر یا کم برای هر سه سناریوی آزمون شده بیشتر است. در ضمن یک سازمان احتمال بیشتری برای دریافت شکایات صفر با یک سیستم ارزیابی ضعیف دارد. برعکس، یک سازمان احتمال بیشتری برای دریافت نرخ بالایی از شکایات با یک سیستم ارزیابی قوی دارد.

جدول 5. احتمالات پیش بینی شده: سیستم های ارزیابی.

	MNLogit	MNLogit	MNLogit	Logit
	احتمال شکایت صفر	احتمال شکایت کم	احتمال شکایت زیاد	احتمال هر شکایتی
سیستم ارزیابی متوسط (متغیرها بر اساس میانگین تنظیم می شوند)	.209	.230	.561	.787
سیستم ارزیابی قوی (متغیرها بر اساس مقدار بیشینه تنظیم می شوند)	ns	.284	.644	.933
سیستم ارزیابی ضعیف (متغیرها بر اساس مقدار کمینه تنظیم می شوند)	.286	.349	.365	.558

توجه: ns نشان می دهد که احتمال پیش بینی شده از نظر آماری در سطح $p < 0.05$ معنی دار نیست.

جدول 6. احتمالات پیش بینی شده: ویژگی های ارزیابی فردی.

		MNLogit	MNLogit	MNLogit	Logit
		احتمال شکایت صفر	احتمال شکایت کم	احتمال شکایت زیاد	احتمال هر شکایتی
شفافیت	High	ns	.336	.406	.710
	Low	ns	ns	.754	.870
ابراز عقیده	High	.162	.240	.598	.846
	Low	ns	ns	ns	ns
آموزش	High	ns	ns	.709	.933
	Low	.637	ns	ns	.345
پاداش ها	High	.359	ns	.481	.646
	Low	.014	ns	ns	.982
توزیع رتبه بندی	High	ns	.249	.652	.899
	Low	.671	ns	ns	ns
عواقب	High	.239	.211	.550	.764
	Low	ns	ns	.564	.902

نکات: ns نشان می دهد که احتمال پیش بینی شده از نظر آماری در سطح $p < 0.05$ معنی دار نیست. High نشان می دهد که متغیر مستقل در حداکثر مقدار خود تنظیم شده است. Low در حداقل مقدار آن تنظیم شده است.

احتمالات پیش بینی شده برای متغیرهای مستقل فردی جزئیات بیشتری را فراهم می کنند (جدول 6 را ببینید). برای توسعه آمار ارائه شده در جدول 6، هر متغیر در بالاترین یا پایین ترین مقدار خود تنظیم شد و تمام متغیرهای دیگر را در میانگین خود نگه داشت. هنگام بررسی این یافته ها در جدول 6، لازم به ذکر است که متغیرهای مستقل فردی در حداقل مقادیر خود به ندرت احتمالات آماری معناداری را نشان می دهند. نرخ های بالا برای توزیع ابراز عقیده، آموزش، جوایز و رتبه بندی، همگی با احتمالات بالاتری از تجربه نرخ بالای شکایات ارتباط دارند. این برخلاف فرضیات است. از سوی دیگر، سازمان هایی که وضوح و پیامدهای پایینی را تجربه می کنند، در مقایسه با سطوح بالایی از وضوح و عواقب، احتمال بیشتری برای تجربه نرخ بالای شکایات دارند. این امر با فرضیه ها سازگار است.

برای رویکرد سوم برای تجزیه و تحلیل رابطه بین شکایات تبعیض مربوط به ارزیابی و طراحی سیستم های ارزیابی، متغیر وابسته را به یک متغیر ساختگی بازایی کردیم که نمره صفر به سازمان هایی با شکایت تبعیض صفر اختصاص داده شده و نمره ۱ به سازمان هایی با هر گونه شکایت تبعیض مربوط به ارزیابی اختصاص داده شده است. هفتاد و سه درصد از ادارات شکایاتی دریافت کرده اند، در حالی که ۲۷ درصد از ادارات هیچ شکایتی دریافت نکرده اند. رابطه بین شکایات تبعیض و طراحی سیستم ارزیابی با یک مدل لجوجیت ارزیابی می شود که شامل تمام متغیرهای مورد استفاده در مدل 3، OLS برای تفسیر ساده است. مدل لجوجیت، آمار والد ۲۵.۴۱ و مربع R تعدیل شده 0.24 را نشان می دهد. علامت و معناداری در مدل لجوجیت (مدل ۵ در جدول ۴) با نتایج در مدل لجوجیت چند جمله ای صفر در مقابل شکایات زیاد (مدل ۴ ب در جدول ۴) مطابقت دارد.

احتمالات پیش بینی شده مجدداً نشان می دهند که سیستم های ارزیابی قوی تر به طور کلی شکایات تبعیض مربوط به ارزیابی را تجربه خواهند کرد (ستون آخر را در جدول ۵ ببینید). به طور خاص یک سازمان با یک سیستم ارزیابی متوسط دارای یک احتمال ۰.۷۹ برای دریافت برخی از شکایات است، در حالی که یک سازمان با یک سیستم ارزیابی ضعیف دارای یک احتمال ۰.۵۶ برای دریافت برخی از شکایات تبعیض مربوط به ارزیابی است. در مقایسه، یک سازمان با یک سیستم ارزیابی قوی دارای احتمال 0.93 برای دریافت برخی شکایات تبعیض است. احتمالات پیش بینی شده متغیرهای مستقل فردی، اطلاعات اضافی را ارائه می دهند (جدول ۶ را ببینید). سازمان هایی با سیستم های ارزیابی که دارای وضوح پایین، جوایز کم و یا عواقب کم هستند، در مقایسه با نرخ های بالای وضوح، جوایز و یا عواقب، احتمال بیشتری برای تجربه برخی از شکایات تبعیض دارند. این موضوع با فرضیه ها سازگار است و وضوح و نتایج یافته ها با نتایج مدل لوجیت چند جمله ای سازگار است. ابراز عقیده بالا با احتمال قابل توجهی با دریافت بیشتر شکایات تبعیض همراه است که با فرضیه سازگار نیست، اما با یافته های لوجیت چند جمله ای مطابقت دارد. سازمان هایی که به طور گسترده آموزش را در سیستم های ارزیابی خود اجرا می کنند، به احتمال قابل توجهی متوجه می شوند که یک سازمان هر گونه شکایت تبعیض را دریافت می کند که با این فرضیه سازگار نیست، اما هم با لوجیت چند جمله ای و هم با یافته های OLS مطابقت دارد. در نهایت، سازمان هایی با توزیع رتبه بالا احتمال بیشتری برای دریافت شکایات تبعیض مربوط به ارزیابی دارند که با فرضیه سازگار است و به طور کلی با یافته های لوجیت چند جمله ای مطابقت دارد.

بحث و نتیجه گیری

تحقیقات موجود نشان می دهد که سیستم های ارزیابی عملکردی که به صورت کارآمد طراحی شده اند، ادراک کارکنان را از منصفانه بودن ارزیابی افزایش می دهند. اگر بپذیریم که میزان شکایات تبعیض مرتبط با ارزیابی، یکی از شاخص های ادراک انصاف هستند، بر این اساس نتایج این مطالعه نشان می دهد که سازمان هایی با سیستم های ارزیابی عملکردی که به صورت کارآمد طراحی شده اند، برخلاف ادبیات موجود، شکایات تبعیض بیشتری را تجربه خواهند کرد. در هر سه مدل مورد مطالعه، افزایش آموزش با نرخ شکایت بالاتر همراه است که با تحقیقات موجود و فرضیه های پیشنهادی در تضاد است. علاوه بر این، نتایج مدل لاجیت و لاجیت چند جمله ای به طور کلی نشان می دهند که سیستم های ارزیابی قوی تر شکایات از تبعیض مرتبط با ارزیابی بیشتری را تجربه می کنند. ابراز عقیده (لحن)، که آزمایش شده ترین ویژگی سیستم های ارزیابی می باشد، در مدل OLS معنی دار نبود، اما بر خلاف تحقیقات موجود در هر دو مدل لاجیت و لاجیت چند جمله ای، فرصت های بیشتر برای ابراز عقیده با نرخ شکایت بیشتر مرتبط بود. شفافیت که میزان درک کارکنان از انتظارات عملکردی را ارزیابی می کند، در مدل OLS معنی دار نبود، اما در هر دو مدل لاجیت چند جمله ای و لاجیت شفافیت پایین با احتمال بیشتر شکایات مرتبط بود، که با فرضیه مطابقت دارد.

تأثیرات ایجاد تمایز در سطوح عملکرد بر شکایات تبعیض در میان مدل ها متناقض بود. جوایز اثرات متفاوتی بر شکایات بین مدل های OLS (منفی)، لوجیت چند جمله ای (مثبت) و لوجیت (منفی) نشان دادند. سازمان های با توزیع گسترده تر در رتبه بندی ها به احتمال زیاد شکایات را به طور کلی و شکایات بیشتری را به ترتیب در مدل های لوجیت و لوجیت چند جمله ای تجربه می کنند که با فرضیات سازگار است. در نهایت، نرخ پایین پیامدها برای اجرا کنندگان ضعیف با احتمالات بیشتر شکایات بالا و تجربه هر گونه شکایت همراه است که با این فرضیه سازگار است.

محققان روانشناسی، تحقیقات در زمینه ارزیابی ها را تغییر داده اند تا بیشتر بر تنظیمات بین فردی و زمینه ای تمرکز کنند که در آن ارزیابی عملکرد اجرا می شود. این مطالعات نشان می دهند که روابط بین ارزیابان و نرخ ها به طور قابل توجهی برای درک فرآیند ارزیابی مهم است، اما این تغییر در تحقیق نشان نمی دهد که ساختار سیستم ارزیابی کاملاً از این ادراکات جدا است. در حالی که ما انتظار روابط بزرگ را نداشتیم، ضعف این ارتباطات غیر منتظره بود.

تا آنجا که ما در بررسی خود از دانش موجود قادر به تعیین بوده ایم، برای این که طراحی نظام ارزشیابی تا چه حد باید تأثیرگذاری نظام را در مقابل عوامل روان شناختی توضیح می دهد، هیچ انتظار استاندارد وجود ندارد. نتایج ممکن است به دلیل مشکلات اندازه گیری و نمونه کوچک باشد؛ هفتاد و نه مشاهده برای مدل های رگرسیون OLS کافی نیست، اما برای مدل های لوجیت چند جمله ای و لوجیت کم است. در ضمن تعصب در انتخاب نمونه ممکن است نتایج را تحت تأثیر قرار دهد. تعیین تعداد برنامه های ارزیابی عملکرد در دولت فدرال ایالات متحده دشوار است، بنابراین در حال حاضر نمی توان پیش بینی کرد که چرا یک سازمان PAAT تکمیل شده را ارائه می دهد. علی رغم این مشکلات، ما باید در رد سریع این نتایج محتاط باشیم. نمونه شامل سازمان هایی با مأموریت ها و مشاغل مختلف و توزیع گسترده ای در تعداد کارمندان است. دو پنجم سازمان های نمونه استاندارد OPM یک سیستم ارزیابی پیچیده را رعایت کردند یا از آن فراتر رفتند.

با حرکت رو به جلو، نتایج چالش های جالبی را برای محققان علاقه مند به درک چگونگی تأثیر سیستم های ارزیابی عملکرد بر فرهنگ سازمان، به ویژه فرهنگ انصاف در یک سازمان ارائه می دهد. اول، محققان معیارهای متعددی را برای ارزیابی اثربخشی سیستم های ارزیابی شناسایی کرده اند. یکی از این اقدامات، برداشت از مطلوبیت سیستم است و تنظیم شکایات تبعیض، یک خروجی

رفتاری محدود از آن ادراکات عدالت است. ممکن است با ورود به یک سازمان، با سیستم های ارزیابی چندگانه و ارزیابی مجموعه ای گسترده تر از معیارهای اثربخشی، بیشتر یاد بگیریم. پیدا کردن یک سازمان فدرال ایالات متحده با سیستم های ارزیابی چندگانه نباید دشوار باشد؛ سیستم های ارزیابی در قراردادهای چانه زنی جمعی گنجانده شده اند، به طوری که یک سازمان با واحدهای چانه زنی چندگانه به احتمال زیاد تغییرات جالبی در طراحی سیستم خواهد داشت. دوم، اگر ما بر این باور باشیم که ساختار دارای ارتباط حداقلی با پیامدها است، ارزیابی اثربخشی ارزیابی در دو سازمان که به نظر می رسد طراحی سیستم ارزیابی مشابهی دارند، جالب خواهد بود. اگر سیستم مشابه باشد، اما نتایج متفاوت باشند، می توانیم استنباط کنیم که پیاده سازی، زمینه و روابط بین فردی بیش از طراحی اهمیت دارند. در نهایت، همانطور که سازمان ها سیستم های ارزیابی جدید را اجرا می کنند، آگاهی از اینکه آیا طراحی جدید انگیزه، انصاف، رفتارهای شهروندی، تعهد و غیره را بهبود می بخشد یا خیر، می تواند آموزنده باشد.

PAAT تحت طرح OPM علائمی ارسال می کند که نشان می دهد سازمان های فدرال ایالات متحده باید تفاوت های بیشتری را در هنگام رتبه بندی کارکنان به وجود آورند و این که رتبه بندی ها باید پیامدهای مکرر و قوی تری برای کارمندان داشته باشند. این تلاش ها منجر به سرخوردگی قابل توجه در سازمان هایی می شود که به طور صریح سیستم های پرداخت برای عملکرد را اتخاذ کرده اند و نیاز به تمایز در رتبه بندی ها دارند. به عنوان مثال، کنگره ایالات متحده فشار قابل توجهی را بر خدمات اجرایی ارشد اعمال کرده است تا اطمینان حاصل کند که تعداد کمی از مدیران ارشد رتبه برتر را به خود اختصاص داده اند. نارضایتی گسترده از سیستم ارزیابی و پرداخت برای عملکرد، و سیگنال های اولیه که توزیع رتبه بندی ها تبعیض آمیز بود، نقش بزرگی در لغو اصلاحات کارکنان غیرنظامی وزارت دفاع آمریکا ایفا کرد. در نهایت، تلاش ها برای اجرای یک طرح پرداخت به ازای عملکرد در کمیسیون اوراق بهادار و بورس ایالات متحده و اداره حفاظت مالی مصرف کننده ایالات متحده نیز به دلیل تبعیض سیستماتیک در برابر زیر گروه های کارمندان با شکست مواجه شد. همانطور که سازمان ها به فشار OPM، کنگره و دیگران برای تجدید نظر در سیستم های ارزیابی پاسخ می دهند، نتایج در اینجا نشان می دهند که ما ممکن است از پرسش زیر بهره ببریم: آیا ما باید بر اصلاح طراحی سیستم های ارزیابی تمرکز کنیم، یا باید به جای آن بر نحوه پیاده سازی سیستم ها تمرکز کنیم؟ نتایج در اینجا نشان می دهد که ما بهتر است بر دومی تمرکز کنیم نه اولی.

منابع و مراجع

Ayers, R. S. (2013). Building goal alignment in federal agencies' performance appraisal programs. *Public Personnel Management, 42*(4), 495–520.

Cawley, B. D., Keeping, L. M., & Levy, P. E. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology, 83*(4), 615–633.

Daley, R. (2004). Designing effective performance appraisal systems. In S. Condrey (Ed.), *Handbook of human resource management in government* (2nd ed., pp. 499–527).

Dobbins, G. H., Cardy, R. L., & Truxillo, D. M. (1988). The effects of purpose of appraisal and individual differences in stereotypes of women on sex differences in performance ratings: A laboratory and field study. *Journal of Applied Psychology, 73*(3), 551.

Kellough, J. E., & Nigro, L. G. (2002). Pay for performance in Georgia state government employee perspectives on GeorgiaGain after 5 years. *Review of Public Personnel Administration, 22*(2), 146–166.

Landy, F. J. (2008). Stereotypes, bias, and personnel decisions: Strange and stranger. *Industrial and Organizational Psychology, 1*(4), 379–392.

Rubin, E., & Kellough, J. (2012). Does civil service reform affect behavior? Linking complaints, alternative personnel systems, and procedural justice perceptions. *Journal of Public Administration Research and Theory, 22*(1), 121–141.