

Pengaruh *Anchoring* dalam Penilaian Kinerja: Sebuah Studi Eksperimen Kuasi dengan Variabel Tergantung Penilaian *Adaptive Performance*

Irmalia Nur Sukmananti¹ & Rahmat Hidayat²
Fakultas Psikologi Universitas Gadjah Mada

Abstract. The purpose of this study was to determine the effect of anchoring on the performance appraisal score in adaptive performance aspect. The hypothesis in this study was to examine whether there were differences in adaptive performance score. High anchoring group had higher adaptive performance score compared to low anchoring group. The method used in this study was quasi experiment with posttest-only design with two treatment groups. Participants involved in this research were 117 with age range from 22 to 49 years old that classified into high anchoring group (n = 70) and low anchoring group (n = 47). Questionnaire to examine the knowledge of the task was used to conduct a manipulation check of anchoring. Independent-samples T-test explained no significant effect of anchoring on adaptive performance score ($t = -0.100$; $p > 0.05$).

Keyword: adaptive performance; anchoring; performance; performance appraisal

Abstrak. Tujuan dari penelitian ini adalah untuk mengetahui pengaruh *anchoring* terhadap hasil penilaian kinerja pada aspek *adaptive performance*. Hipotesis yang diajukan dalam penelitian ini adalah terdapat perbedaan hasil penilaian kinerja aspek *adaptive performance* pada *high anchoring* dan *low anchoring*. Perbedaan hasil penilaian kinerja khususnya pada aspek *adaptive performance* yang dimaksudkan yaitu *high anchoring* memiliki skor penilaian yang lebih tinggi dibandingkan dengan *low anchoring*. Metode penelitian yang digunakan adalah eksperimen kuasi dengan desain penelitian berupa *the posttest-only design with two treatment groups*. Sebanyak 117 partisipan berusia 22 hingga 49 tahun dibagi ke dalam kelompok *high anchoring* (n = 70) dan kelompok *low anchoring* (n = 47). Kuesioner pemahaman mengenai tugas digunakan untuk melakukan cek manipulasi *anchor*. *Independent-Samples T Test* menunjukkan tidak adanya pengaruh yang signifikan antara hasil penilaian kinerja aspek *adaptive performance* dan *anchoring* ($t = -0,100$; $p > 0,05$).

Kata kunci: adaptive performance; anchoring; kinerja; penilaian kinerja

¹ Korespondensi dapat dilakukan melalui:
irmalianurs@gmail.com

² Atau melalui r.hidayat@ugm.ac.id

Kinerja merupakan salah satu kunci indikator perusahaan dalam memelihara performansi tim dan perusahaan. Armstrong & Baron (1998) menjelaskan bahwa kinerja adalah hasil dari pekerjaan yang memiliki hubungan kuat dengan tujuan strategi organisasi, kepuasan konsumen, dan memberikan kontribusi ekonomi. Kinerja karyawan tentunya membutuhkan penilaian secara berkelanjutan. Penilaian kinerja merupakan komponen penting bagi manajemen kinerja dalam membantu dan mendukung tujuan perusahaan. Namun dalam prosesnya, tidak jarang metode yang digunakan dalam penilaian kinerja mengalami berbagai masalah sehingga hasil yang dicapai kurang sesuai dengan yang diharapkan.

Dimensi atau aspek yang sering diukur dalam kinerja individu atau pegawai adalah *task performance*, *contextual performance*, *counterproductive work behavior*, dan *adaptive performance* (Koopmans *et al.*, 2011). *Adaptive performance* merupakan aspek yang masih baru. *Adaptive performance* didefinisikan sebagai sejauh mana individu beradaptasi pada perubahan di dalam sistem kerja atau peran kerja (Giffin, Neal, & Parker, 2007). Contoh nyata dari *adaptive performance* adalah mengatasi masalah secara kreatif; mempelajari tugas, teknologi, dan prosedur yang baru; dan mampu beradaptasi dengan individu lainnya, budaya, atau lingkungan fisik (Koopmans *et al.*, 2011). *Adaptive performance* merupakan aspek yang masih baru sehingga masih terdapat kelemahan dari aspek tersebut.

Penilaian terhadap *adaptive performance* lebih banyak pada penilaian yang bersifat subjektif sehingga dalam praktik penilaian kinerja dapat menimbulkan adanya bias penilaian.

Pengukuran subjektif seringnya memiliki ancaman kontaminasi dalam bentuk *rater bias* dan meningkatkan *random error* (Campbell, 1990). Selain itu, Stokes, Schneider, & Lyons (2010) menemukan kelemahan pada validitas konvergen pada dua tipe mayor asesmen *adaptive performance* (objektif dan subjektif) dan menjadi perhatian jika validitas konstruk pada *adaptive performance* tidak terlalu kuat. Oleh karena itu, dalam penelitian ini, peneliti tertarik untuk mengambil salah satu aspek yaitu *adaptive performance* yang nantinya akan menjadi aspek yang diukur dalam penilaian kinerja.

Seiring berjalannya waktu, masalah dalam penilaian kinerja mulai muncul. Evaluasi kinerja atau penilaian kinerja dalam *setting* organisasi menunjukkan beberapa masalah terkait dengan metodologi dan implementasi (Cardy & Dobbins, 1994), saat ini kebutuhan untuk kegunaan penilaian kinerja sebagai pengambilan keputusan (misalnya keputusan kompensasi dan promosi) masih dipertanyakan (Valle & Davis, 1999).

Masalah dalam penilaian kinerja tidak dapat terlepas dari adanya bias. Masalah dalam penilaian kinerja dirangkum dalam Davis & Werther (1989) yaitu bias langsung; bias tidak langsung; kompetensi; devolusi; *authoritarian*; informal, *incidental* dan *ongoing appraising*; pengumpulan informasi. Selanjutnya, beberapa peneliti seperti Schweiger & Sumners (1994), Longenecker (1997), Longenecker & Fink (1998), Rees & Porter (2003), dan Rees & Porter (2004) telah mempublikasikan artikel terkait kegagalan dan kesalahan dari penilaian kinerja. Mereka mengidentifikasi hasil yang berbeda. *Psychometric errors* adalah salah satu penyebab utama pada kegagalan penilaian kinerja (Appelbaum,

Roy, & Gilliland, 2011). *Error* seperti *leniency*, *halo effect*, *restriction of range*, *recency* dan *contrast*, adalah atribut dari predisposisi psikologis seorang penilai selama proses penilaian kinerja (Appelbaum, Roy, & Gililand, 2011).

Dari beberapa penelitian terdahulu, dapat disimpulkan bahwa masalah yang terjadi dalam penilaian kinerja tidak jauh dari penilai (*rater*). Beberapa penilai dapat memberikan hasil yang berbeda pada individu yang sama (Tsui & Ohlott, 2006; Zammuto, London, & Rowland., 1982). Seperti yang dikemukakan oleh Appelbaum et al (2011) bahwa error seperti *leniency*, *halo effect*, *restriction of range*, *recency* dan *contrast*, adalah atribut dari predisposisi psikologis seorang penilai selama proses penilaian kinerja. Duarte, Goodson, & Klich (1994) merangkum beberapa hal yang dapat memengaruhi hasil penilaian kinerja yaitu, tipe dari format rating dan instrumen yang digunakan, karakteristik penilai dan yang dinilai, proses kognitif dan pelatihan penilai, faktor kontekstual, dan tujuan penilaian.

Adapun penelitian terkait pengaruh *anchoring* pada *judgment* menjelaskan bahwa seseorang dapat dipengaruhi oleh *irrelevant* dan *uninformative information* yang bertindak sebagai *anchor* (Chapman & Johnson, 2002; Tversky & Kahneman, 1974). Dalam konteks penilaian kinerja, seorang penilai dalam menentukan keputusan dapat dipengaruhi oleh *anchor* dimana *anchor* dapat berupa informasi sebelumnya yang terkait maupun tidak terkait dengan apa yang akan dinilai.

Metode

Subjek penelitian adalah 204 mahasiswa Magister Manajemen UGM program

eksekutif. Penelitian ini menggunakan 2 kelompok, yaitu kelompok eksperimen 1 (*high anchoring*) dan kelompok eksperimen 2 (*low anchoring*). Subjek dibagi menjadi dua kelompok eksperimen, yaitu *high anchoring* dan *low anchoring*. Kedua kelompok diberi *booklet* yang berisikan skenario deskripsi pegawai, lembar penilaian kinerja, dan uji pemahaman yang bertindak sebagai *manipulation check*. Perlakuan yang diberikan adalah instruksi gaya *rating top down* dan instruksi gaya *rating down top*. *Post-test* yang diberikan berupa hasil dari tugas menilai pada lembar penilaian kinerja yang dikerjakan dari kedua kelompok. Cek manipulasi juga digunakan untuk mengetahui sejauh mana subjek mendalami peran dan tugas yang diberikan.

Penelitian ini menggunakan pendekatan kuantitatif dengan metode kuasi eksperimen. Jenis desain penelitian yang digunakan adalah *posttest-only design with two treatment groups*. Pengukuran penilaian kinerja aspek *adaptive performance* diambil dengan lembar penilaian kinerja. Format rancangan lembar penilaian kinerja berbeda-beda berdasarkan tujuan penilaiannya. Pada penelitian ini peneliti menggunakan teknik *Behaviourally Anchored Rating Scales* (BARS). *Behaviourally Anchored Rating Scales* (BARS) adalah teknik penilaian kinerja di mana penilai menilai pegawai berdasarkan beberapa jenis perilaku kerja yang mencerminkan dimensi kinerja dan membuat skalanya. Metode ini mendeksripsikan perilaku yang diharapkan sesuai dengan tingkat kinerja yang diharapkan. Penelitian ini akan mengukur salah satu dimensi kinerja yaitu *adaptive performance*.

Adapun tahapan proses pengumpulan data dalam penelitian dimulai dengan penngambilan data yang

dilakukan oleh peneliti sendiri dengan didampingi oleh petugas akademik MM UGM untuk mendatangi subjek penelitian yang berada di dalam ruang kelas-kelas yang telah dipilih sebelumnya. Peneliti menjelaskan instruksi pengisian kuisisioner kepada calon subjek dan apabila ada sesuatu yang kurang jelas, calon subjek dipersilakan untuk bertanya. Pengumpulan data dilakukan dengan cara membagikan *booklet* secara langsung oleh peneliti dibantu oleh rekan-rekan di sekitar calon subjek, dan setelah pengisian selesai, *booklet* dikumpulkan kepada ketua kelas untuk selanjutnya dikumpulkan pada bagian akademik, setelah itu peneliti akan mengambil *booklet* yang telah terkumpul di bagian akademik. Setelah data terkumpul, peneliti melakukan uji manipulasi terlebih dahulu. Hasil skor manipulasi nantinya akan menentukan data subjek tersebut dapat lolos atau tidak. Proses selanjutnya kemudian analisis data dari subjek yang telah lolos uji manipulasi berdasarkan kriteria yang telah ditentukan oleh peneliti.

Dalam penelitian ini, skala *adaptive performance* mengacu pada skala Pulakos (2000) dengan mengambil 4 (empat) indikator, yaitu (1) *problem solving*, (2) *new learning*, (3) *interpersonal adaptability*, (4) *cultural adaptability*.

Booklet skenario merupakan kumpulan dari beberapa tugas yang akan diberikan kepada subjek. *Booklet* skenario berisi panduan tugas, skenario tugas subjek yang akan berperan menjadi seorang atasan yang menilai bawahannya, deskripsi pegawai yang akan dinilainya, rubrik pedoman penilaian kinerja, petunjuk pengisian, lembar penilaian kinerja, dan uji pemahaman yang bertindak sebagai uji manipulasi. *Booklet* berisi tiga bagian utama, bagian pertama adalah deskripsi pegawai yang akan

dinilai oleh subjek yang berperan menjadi atasan; bagian kedua adalah instruksi cara menilai serta lembar penilaian kinerja yang terdiri dari beberapa butir pernyataan *adaptive performance*; bagian ketiga adalah *check manipulation* yang berupa lima pernyataan yang harus diisi oleh subjek. Peneliti melakukan uji coba sebelum dilakukan pengambilan data. Pelaksanaan uji coba memiliki tujuan untuk mengetahui prosedur pelaksanaan dan meminimalisasi kemunculan kendala *booklet* skenario di lapangan. Uji coba keterbacaan *booklet* skenario dilakukan pada dosen pembimbing skripsi, peneliti, 4 mahasiswa Magister PIO dan 2 alumni Magister PIO. Rubrik pedoman penilaian kinerja juga diberikan dalam *booklet*. Rubrik adalah suatu pedoman penskoran yang digunakan untuk menentukan tingkat kemahiran seseorang dalam mengerjakan sesuatu.

Sebelum menguji hipotesis, dalam penelitian eksperimen peneliti akan melakukan *manipulation check* terlebih dahulu. *Manipulation check* dilakukan terhadap kelompok eksperimen *high anchoring* dan kelompok eksperimen *low anchoring* dengan cara menguji pemahaman subjek terhadap tugas dan instruksi yang diberikan. Pengukuran ini akan menggunakan kuisisioner. Uji coba keterbacaan dilakukan kepada dosen pembimbing, peneliti, mahasiswa Magister PIO, dan peneliti sebelum diberikan kepada subjek.

Data penelitian ini akan dianalisis menggunakan *Independent-Sample T Test*. *Independent-Sample T Test* digunakan untuk membandingkan perbedaan skor *adaptive performance* antara kelompok *high anchoring* dan *low anchoring*. *Independent-Sample T Test* dilakukan dengan bantuan *IBM Statistical Package for Social Science (SPSS) for Windows Version 21.0*.

Interpretasi *Independent-Sample T Test* dapat dilihat dari signifikansinya, jika nilai $p < 0,05$ maka terdapat perbedaan yang signifikan antara dua kelompok. Sebaliknya, jika analisis menunjukkan hasil yang tidak signifikan yaitu $p > 0,05$, skor *adaptive performance* pada kedua kelompok tidak memiliki perbedaan yang signifikan (Field, 2013).

Hasil

Pengambilan data pada eksperimen ini dibagi menjadi dua kloter, jadwal masing-masing kloter berbeda-beda karena keterbatasan fasilitas eksperimen. Jadwal pengambilan data ditetapkan pada 10, 16, dan 17 Juni 2017. Pada tanggal 10 Juni 2017 pengambilan data dimulai pukul 13.00 WIB. Lokasi pelaksanaan pada tanggal tersebut di kampus MM UGM Yogyakarta. Pada tanggal 16 dan 17 Juni 2017, lokasi pelaksanaan di kampus MM UGM Jakarta. Pada tanggal 16 Juni 2017 dimulai pukul 18.00, sedangkan pada tanggal 17 Juni dimulai pukul 10.00. Proses pengambilan data berjalan lancar pada kedua lokasi tersebut.

Calon subjek pada penelitian ini terdiri dari 204 mahasiswa. Akan tetapi, booklet yang kembali hanya 146 (kelompok *high anchoring*= 87 dan kelompok *low anchoring*= 59). Total subjek yang mengikuti penelitian ini berjumlah 146 subjek, yang terdiri dari 76 laki-laki dan 70 perempuan. Usia subjek berkisar antara 22 tahun hingga 49 tahun. Total subjek yang lolos uji manipulasi berjumlah 117 subjek

Uji konsistensi internal yang telah dilakukan menunjukkan bahwa semua butir memiliki *total-item correlation* yang baik ($r = 0,254 - 0,434$). Uji realibilitas dapat dilakukan dengan menggunakan formula koefisien alpha (α). Koefisien *Cronbach's*

Alpha menunjukkan nilai 0,556 sehingga skala *Adaptive Performance* pada kelompok *low anchoring* dalam penelitian ini memiliki realibilitas sedang.

Uji konsistensi internal yang telah dilakukan menunjukkan bahwa semua butir memiliki *total-item correlation* yang baik ($r = 0,311 - 0,589$). Uji realibilitas dapat dilakukan dengan menggunakan formula koefisien alpha (α). Koefisien *Cronbach's Alpha* menunjukkan nilai 0,750 sehingga skala *Adaptive Performance* pada kelompok *high anchoring* dalam penelitian ini memiliki realibilitas yang cukup baik.

Sebelum melakukan uji hipotesis dengan *Independent-Samples T Test*, ada beberapa analisis yang dilakukan, salah satunya uji asumsi. Peneliti melakukan dua uji asumsi, yaitu uji normalitas dan uji homogenitas. Uji normalitas dilakukan untuk memastikan bahwa data mengikuti distribusi yang normal, sementara itu uji homogenitas dilakukan untuk mengetahui homogenitas varian data. Hasil menunjukkan bahwa persebaran data skor *adaptive performance* pada kelompok *high anchoring* adalah normal ($p = 0,115$; $p > 0,05$). Sedangkan data skor *adaptive performance* pada kelompok *low anchoring* adalah normal ($p = 0,362$; $p > 0,05$). *Levene Test for Equality of Variances* menunjukkan bahwa varian data adalah homogen ($p = 0,419$; $p > 0,05$). *Independent-Samples T Test* dapat diterapkan karena data bersifat normal dan homogen.

Pengujian hipotesis digunakan untuk membuktikan apakah hipotesis penelitian diterima atau ditolak. Hipotesis akan diterima bila hasil analisis menunjukkan perbedaan skor *adaptive performance* yang signifikan antara kelompok *high anchoring* dan kelompok *low anchoring*. Pengujian hipotesis penelitian ini menggunakan *Independent-Samples T Test*. Uji *independent-samples t test*

skor *adaptive performance* antara kelompok *high anchoring* dan *low anchoring* menunjukkan nilai signifikansi sebesar 0,920. Nilai signifikansi atau p yang kurang dari 0,05 menunjukkan adanya perbedaan yang signifikan, nilai p yang tertera di dalam tabel 8 menunjukkan bahwa $p > 0,05$ sehingga tidak ada perbedaan yang signifikan *adaptive performance* pada *high anchoring* dan *low anchoring*. Hipotesis pada penelitian ini adalah terdapat perbedaan hasil penilaian kinerja pada aspek *adaptive performance* antara *high anchoring* dan *low anchoring*. Tidak adanya perbedaan hasil skor dari aspek *adaptive performance* yang signifikan antara *high anchoring* dan *low anchoring* membuat hipotesis pada penelitian ini ditolak. Selanjutnya dilakukan pengujian *t test* pada masing-masing aitem menunjukkan hasil nilai signifikansi yang berkisar dari 0,044 hingga 0,962. Nilai signifikansi atau p yang kurang dari 0,05 menunjukkan adanya perbedaan yang signifikan, nilai p yang tertera di dalam tabel 7 menunjukkan bahwa aitem 1 sampai 5 memiliki $p > 0,05$ sehingga tidak ada perbedaan yang signifikan pada masing-masing aitem di kelompok *high anchoring* dan *low anchoring*. Hasil menunjukkan bahwa hanya aitem 6 yang memiliki perbedaan yang signifikan dengan $p < 0,05$, yaitu 0,044.

Diskusi

Penelitian ini bertujuan untuk mengetahui pengaruh *anchoring* pada hasil penilaian kinerja aspek *adaptive performance*. Hipotesis yang diajukan pada penelitian ini adalah terdapat perbedaan skor *adaptive performance* antara *high anchoring* dan *low anchoring*. Setelah dianalisis dengan menggunakan *Independent-Samples T Test* tidak ditemukan adanya

perbedaan yang signifikan. Temuan ini berbeda dengan penelitian-penelitian sebelumnya yang melihat bahwa terdapat adanya pengaruh *anchoring* dalam hasil kinerja. Thorsteinson *et al.* (2008) menemukan bukti dalam penelitiannya jika keseluruhan hasil dari *performance rating* yang dilakukan oleh siswa kepada *fictional sales representative* dan kepada instruktur mereka secara konsisten lebih tinggi pada kelompok yang dipaparkan *high anchor* daripada kelompok yang dipaparkan pada *low anchor* atau *no anchor*. Penelitian sebelumnya menunjukkan bahwa manipulasi dari *anchor* dapat memengaruhi hasil penilaian kinerja. Penelitian yang dilakukan tidak dapat menemukan pengaruh *anchor* terhadap hasil penilaian kinerja. Hal tersebut karena manipulasi *anchor* yang diberikan tidak kuat disadari oleh subjek walaupun hasil skor dari *manipulasi check* menunjukkan hasil yang mencukupi.

Faktor yang membuat penelitian ini tidak terbukti seperti pada pelaksanaan eksperimen yang menggunakan sistem klasikal sehingga perlakuan yang telah diberikan tidak dapat terkontrol dengan baik dan memungkinkan adanya interaksi antarsubjek yang dapat memengaruhi hasil dari skor penilaian kinerja pada aspek *adaptive performance*. Penelitian ini diberikan kepada subjek dengan memberikan tugas fiksi untuk menilai seorang karyawan fiksi yang telah disusun oleh peneliti. Hal ini bisa saja membuat subjek tidak terlalu bersungguh-sungguh dalam mengerjakan tugasnya. Thorsteinson *et al.* (2008) mengemukakan bahwa partisipan atau subjek mengetahui jika mereka memberikan rating pada karyawan fiksi sehingga mereka bisa jadi memiliki motivasi yang rendah untuk memperhatikan dengan seksama dan

detail semua informasi yang telah disediakan oleh peneliti.

Selain itu, peneliti menemukan adanya indikasi pengaruh dari diberikannya rubrik panduan penilaian kinerja pada rangkaian eksperimen. Tujuan awal peneliti untuk memasukan rubrik ke dalam rangkaian eksperimen yang seharusnya dapat mengontrol jalannya eksperimen. Namun setelah eksperimen dilakukan, peneliti melihat bahwa penyajian rubrik dapat berpengaruh pada hasil penilaian subjek. Hal ini karena penyajian rubrik diberikan sebelum lembar penilaian kinerja sehingga subjek dalam menilai kinerja akan berpatokan dengan rubrik yang sudah diberikan. Perlakuan *anchor* yang ada di awal menjadi lemah akibat dari disajikannya rubrik panduan tersebut. Penyesuaian (*adjustment*) pada umumnya tidak mencukupi (Tversky & Kahneman, 1974), hal tersebut dapat terjadi karena nilai yang baru tidak terlalu diterima dibanding dengan nilai yang sudah ada sebelumnya (Wilson Houston, Etling & Brekke, 1994). Tidak terbuktinya penelitian ini dapat dipengaruhi oleh penyesuaian subjek yang berdasar pada adanya nilai yang baru yaitu rubrik. Sehingga hal tersebut mencemari perlakuan eksperimen dengan melemahnya nilai awal yang telah diberikan sebagai perlakuan.

Faktor terakhir yang menyebabkan hipotesis tidak terbukti adalah adanya *recency effect*. *Recency effect* dapat diartikan sebagai bias keputusan yang terjadi ketika pengambil keputusan membobot informasi terakhir lebih besar dibandingkan informasi yang diterima sebelumnya. Hal ini berkaitan dengan urutan penyajian informasi yang disusun oleh peneliti sehingga informasi yang terakhir (rubrik pedoman penilaian

kinerja) menjadi lebih besar pengaruhnya daripada *anchor* yang telah disajikan di awal.

Kesimpulan

Kesimpulan dalam penelitian ini adalah *anchoring* tidak berpengaruh secara signifikan terhadap hasil penilaian kinerja pada aspek *adaptive performance*. Hasil penelitian tidak sesuai dengan hipotesis bisa jadi dikarenakan adanya indikasi pengaruh dari faktor lain yang dapat mencemari jalannya eksperimen.

Saran

Berdasarkan temuan, pembahasan, dan kelemahan penelitian, peneliti memberikan saran untuk penelitian berikutnya. Saran-saran tersebut antara lain penelitian selanjutnya sebaiknya melaksanakan eksperimen di ruang bebas distraksi, penelitian selanjutnya sebaiknya mempertimbangkan metode penilaian lain, misalnya menerapkan metode penilaian 360°. Penelitian ini hanya menerapkan metode penilaian dengan satu rater saja. Penelitian selanjutnya sebaiknya memperhatikan kembali faktor-faktor yang dapat berisiko mencemari eksperimen. Penelitian selanjutnya diharapkan dapat memasukan rubrik pedoman penilaian kinerja sebagai perlakuan, sehingga dapat terlihat bagaimana pengaruhnya. Penelitian selanjutnya sebaiknya meninjau kembali penerapan manipulasi cek.

Daftar Pustaka

Appelbaum, S. H., Roy, M., & Gilliland, T. (2011). Globalization of performance appraisal: theory and applications. *Journal of Management*

- History*, 4(3), 233-249. doi: [10.1177/1534484318798533](https://doi.org/10.1177/1534484318798533)
- Armstrong, Michael & Baron, A. (1998). *Performance management: The new realities (Developing practice)*. New York: Gardners Books.
- Campbell, J. P. (1990). Modeling the performance prediction problem in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (pp. 687-732). Palo Alto, CA, US: Consulting Psychologists Press.
- Cardy, R. L., & Dobbins, G. H. (1994). *Performance appraisal: Alternative perspectives*. Cincinnati: South-Western.
- Chapman, G. B., & Johnson, E. J. (2002). Incorporating the irrelevant: Anchors in judgments of belief and value. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 120-138). Cambridge, England: Cambridge University Press.
- Davis, K. & Werther, W. B. (1989). *Human resource and personnel management*. 5th edition. USA: McGraw-Hill, Inc.
- Duarte, N. T., Goodson, J. R., & Nancy, R. K. (1994). Effects of dyadic quality and duration on performance appraisal. *The Academy of Management Journal*, 37(3), 499-521. doi: [10.2307/256698](https://doi.org/10.2307/256698)
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. London: SAGE Publications Ltd.
- Griffin, M. A., Neal, A., & Parker, S. K. (2007). A new model of work role performance: Positive behavior in uncertain and interdependent contexts. *Academy of Management Journal*, 50(2), 327-347. doi: [10.5465/AMJ.2007.24634438](https://doi.org/10.5465/AMJ.2007.24634438)
- Koopmans L, Bernaards C. M, Hildebrandt, V. H, Schaufeli, W. B., De Vet, H. C. W, & Van der Beek, A. J. (2011). Conceptual frameworks of individual work performance—A systematic review. *Journal of Occupational and Environmental Medicine*, 53(8), 856-866. doi: [10.1097/JOM.0b013e318226a763](https://doi.org/10.1097/JOM.0b013e318226a763)
- Longenecker, C. O. (1997). Why managerial performance appraisals are ineffective: Causes and lessons. *Career Development International*, 2(5), 212-218. doi: [10.1108/13620439710174606](https://doi.org/10.1108/13620439710174606)
- Longenecker, C. O. & Fink, L. S. (1998). Training as performance appraisal improvement strategy. *Career Development International*, 3(6), 243-251.
- Pulakos, E. D., Arad, S., Donovan, M. A., & Plamondon, K. E. (2000). Adaptability in the workplace: Development of a taxonomy of adaptive performance. *Journal of Applied Psychology*, 85(4), 612-624.
- Rees, D. W., & Porter, C. (2003). Appraisal pitfalls and the training implications-part 1. *Industrial and Commercial Training*, 35(7), 280-284. doi: [10.1108/00197850310501677](https://doi.org/10.1108/00197850310501677)
- Rees, D. W., & Porter, C. (2004). Appraisal pitfalls and the training implications-part 2. *Industrial and Commercial Training*, 36(1), 29-34. doi: [10.1108/00197850410516094](https://doi.org/10.1108/00197850410516094)
- Schweiger, I. & Sumners, G. (1994). Optimizing the value of performance appraisals. *Managerial Auditing Journal*, 9(8), 3-7. doi: [10.1108/02686909410071124](https://doi.org/10.1108/02686909410071124)

- Stokes, C. K., Schneider, T. R., & Lyons, J. B. (2010). Adaptive performance: A criterion problem. *Team Performance Management: An International Journal*, 16(3/4), 212-230. doi: [10.1108/13527591011053278](https://doi.org/10.1108/13527591011053278)
- Thorsteinson, T. J., Breier, J., Atwell, A., Hamilton, C., & Privette, M. (2008). Anchoring effects on performance judgments. *Organizational Behavior and Human Decision Processes*, 107, 29-40. doi: [10.1016/j.obhdp.2008.01.003](https://doi.org/10.1016/j.obhdp.2008.01.003)
- Tsui, A. S., & Ohlott, P. (2006). Multiple assessment of managerial effectiveness: Interrater agreement and consensus in effectiveness models. *Personnel Psychology*, 41(4), 779-803. doi: [10.1111/j.1744-6570.1988.tb00654.x](https://doi.org/10.1111/j.1744-6570.1988.tb00654.x)
- Tversky, A., & D, Kahneman. (1974). Judgment under uncertainty: Heuristics and biases. *Science, New Series*, 185(4157), 1124-1131. doi: [10.1126/science.185.4157.1124](https://doi.org/10.1126/science.185.4157.1124)
- Valle, M., & Davis, K. (1999). Teams and performance appraisal using matrices to increase reliability and validity. *Team Performance Management*, 5(8), 238-244. doi: [10.1108/13527599910304912](https://doi.org/10.1108/13527599910304912)
- Wilson, T. D., Houston, C., Etling, K. M., & Brekke, N. (1996). A new look at anchoring effects: Basic anchoring and its antecedents. *Journal of Experimental Psychology: General*, 4, 387-402.
- Zammuto, R., London, M. & Rowland, K. (1982). Organizational and rater differences in performance appraisal. *Personnel Psychology*, 35, 643-658.