

Jurnal Akuntansi dan Keuangan Indonesia
Volume 11 Nomor 1, Juni 2014

BAGAIMANA MENGURANGI BIAS KEMURAHAN HATI DALAM PENILAIAN KINERJA SUBJEKTIF? SEBUAH PENDEKATAN EKSPERIMEN¹

Monika Handojono

Politeknik Negeri Ambon
maden_acc@yahoo.com

Mahfud Sholihin

Universitas Gadjah Mada
mahfud@gadjahmada.edu

Abstract

The literature shows that leniency has negative impact on employee performance and firm productivity. However, there has been limited empirical research of how to mitigate the bias. This study examines if leniency is mitigateable by availability of subjective evaluation criteria and group rater. Using experimental method with sixty-nine undergraduate students as subjects, we find that process within group is effective to reduce individual subjective rating, even when there was no subjective criteria. However, we find group rating does not have effect on rating accuracy. Additionally, consistent with general assumption of leniency, we find that inflated rating is affected by altruistic traits of raters. With the inherent limitations associated with experimental method, our finding suggests that, in order to have subjective rating accurately, firm should promote sound rating process by defining more relevant criteria to complement the evaluation process.

Keywords: leniency, subjective evaluation, subjective evaluation criteria, group rater

Abstrak

Literatur menunjukkan bahwa bias kemurahan hati berdampak negatif terhadap kinerja karyawan dan produktivitas perusahaan. Meskipun demikian, hanya terdapat sedikit penelitian empiris mengenai cara memitigasi bias ini. Penelitian ini menguji apakah bias kemurahan hati dapat dimitigasi melalui ketersediaan kriteria evaluasi subjektif dan penggunaan grup penilai. Dengan menggunakan pendekatan eksperimen yang diikuti oleh 69 mahasiswa S1 sebagai subjek, kami menemukan bahwa proses yang terjadi dalam grup efektif untuk menurunkan penilaian kinerja subjektif yang diberikan secara individual, bahkan ketika tidak tersedia kriteria subjektif. Namun, kami juga menemukan bahwa kriteria subjektif tidak berpengaruh terhadap keakuratan penilaian kinerja. Selain itu, konsisten dengan asumsi umum mengenai bias kemurahan hati, kami menemukan bahwa penilaian kinerja yang ditinggikan dipengaruhi oleh kepribadian altruistik penilai. Dengan berbagai keterbatasan yang melekat pada desain eksperimen, temuan kami menyarankan bahwa untuk menghasilkan penilaian kinerja subjektif yang akurat, perusahaan harus mendorong proses penilaian yang lebih baik melalui penetapan kriteria yang lebih relevan dalam mendukung proses evaluasi kinerja.

Kata kunci: bias kemurahan hati, penilaian kinerja subjektif, kriteria penilaian kinerja subjektif, grup penilai

¹ Artikel ini merupakan bagian dari penulisan tesis S2 di Universitas Gadjah Mada yang berjudul "Apakah Bias Kemurahan Hati Dapat Dikurangi? Sebuah Pendekatan Eksperimen"

PENDAHULUAN

Bias kemurahan hati (*leniency*) telah mengundang perhatian beberapa peneliti dalam beberapa tahun belakangan ini. Beberapa penelitian bertema bias kemurahan hati yang mengemuka berkisar terutama pada determinan (Breur et al. 2013; Bol 2011; Slaughter dan Greguras 2008; Varma dan Pichler 2007; Moers 2005; Kingstrom dan Mainstone 1985) dan upaya mengurangi bias kemurahan hati (Chen 2012; Golman dan Bhatia 2012; Giebe dan Gürtler 2012; Grund dan Przemec 2012). Asumsi umum yang mendasari penelitian-penelitian tersebut adalah kepribadian manajer yang altruistik yang menyebabkan manajer tersebut bermurah hati. Altruistik menunjukkan kecenderungan individu untuk melakukan tindakan-tindakan secara sukarela dengan tujuan untuk memberikan manfaat bagi orang lain bukan bagi dirinya sendiri. Sayangnya, belum ada penelitian yang secara empiris menguji dampak kepribadian altruistik terhadap bias kemurahan hati dalam penilaian kinerja. Oleh karena itu, penelitian ini menguji dampak kepribadian altruistik terhadap bias kemurahan hati.

Bias kemurahan hati merupakan salah satu bias yang terjadi dalam penilaian kinerja subjektif (Breuer et al. 2013; Giebe dan Gürtler 2012; Golman dan Bhatia 2012; Bol 2011; Moers 2005). Berbeda dengan penilaian kinerja objektif yang memiliki ukuran-ukuran yang objektif dan terverifikasi, penilaian kinerja subjektif lebih mengandalkan *judgement* penilai sehingga sangat rentan terhadap bias-bias keperilakuan (Bol 2008; Gibbs et al. 2004; Prendergast dan Topel 1996). Dalam literatur ekonomi, bias kemurahan hati dianggap berbahaya bagi kinerja jangka panjang perusahaan karena memberikan dampak negatif terhadap usaha dan produktivitas karyawan serta mendistorsi upah (Golman dan Bhatia 2012). Selain argumen teoretis, dukungan terhadap dampak negatif bias kemurahan hati juga ditunjukkan melalui bukti empiris. Hasil eksperimen laboratorium Berger et al. (2010) menunjukkan bahwa penilaian kinerja yang diberikan oleh penilai dengan murah hati

menghasilkan penurunan kinerja partisipan dalam jangka panjang. Dengan demikian, hal tersebut menunjukkan bahwa bias kemurahan hati harus dimitigasi.

Dalam literatur mengenai upaya mengurangi bias dalam penilaian kinerja, penilaian grup ditemukan dapat mengurangi *performance-cues bias*. Hal ini disebabkan penggunaan beberapa penilai akan mengurangi informasi yang berlebihan ketika digunakan sebagai dasar penilaian sehingga dapat memperbaiki kesalahan dan bias (Martell dan Leavitt 2002). Selain menggunakan grup penilai, melengkapi aktivitas penilaian dengan menyediakan kriteria kinerja subjektif kemungkinan dapat mengurangi bias kemurahan hati (Golman dan Bhatia 2012). Gagasan ini sejalan dengan uraian yang dikemukakan oleh Giebe dan Gürtler (2012). Dalam studi ini, kriteria kinerja subjektif yang dipakai diadaptasi dari Bellavance et al. (2013).

Dengan demikian, penelitian ini menguji apakah penggunaan kriteria kinerja subjektif dan aktivitas grup penilai dapat digunakan untuk memitigasi bias kemurahan hati. Pertimbangan yang melandasi usulan ini bersumber dari kondisi yang ditemukan pada eksperimen yang dilakukan oleh Bol dan Smith (2011). Ketiadaan kriteria maupun arahan yang diberikan kepada partisipan mengenai ukuran subjektif yang digunakan untuk evaluasi, kemungkinan menyebabkan kognisi partisipan terdistorsi oleh informasi yang sebenarnya tidak relevan dengan tugas yang dinilai. Akibatnya penilai memberikan penilaian kinerja yang tinggi. Ada kemungkinan hal ini akan berbeda bila kognisi partisipan yang akan memberi penilaian kinerja diarahkan pada informasi-informasi yang lebih relevan dengan penilaian.

Alasan lainnya yang mendasari usulan penggunaan grup penilai untuk mengurangi bias kemurahan hati berasal dari kondisi yang ditemui pada beberapa penelitian sebelumnya (Bol dan Smith 2011; Liedtka et al. 2008). Ternyata, dalam semua studi tersebut, penilaian kinerja dilakukan oleh partisipan yang berperan secara individual. Kemungkinan penggunaan partisipan individual yang tidak

memiliki keahlian dan pengalaman yang cukup dapat memengaruhi hasil penelitian. Selain itu, penelusuran yang dilakukan oleh Giebe dan Gürtler (2012) menemukan bahwa terjadi perbedaan yang sistematis dalam penilaian kinerja ketika seorang karyawan dievaluasi oleh supervisor yang berbeda. Beberapa supervisor cenderung memberikan penilaian kinerja lebih tinggi dari penilaian kinerja yang diberikan oleh supervisor lain, terutama pada supervisor yang memiliki hubungan yang dekat dengan bawahan yang dinilainya.

Studi ini menggunakan material eksperimen yang dikembangkan oleh Bol dan Smith (2011) dengan sedikit modifikasi yakni menambahkan beberapa kriteria kinerja subjektif yang digunakan oleh Bellavance et al. (2013) untuk membantu penilai melakukan penilaian kinerja. Tugas penilaian kinerja dikerjakan oleh 69 orang mahasiswa tingkat sarjana program studi akuntansi Universitas Gadjah Mada. Partisipan diminta untuk mengasumsikan dirinya sebagai manajer regional yang melakukan penilaian kinerja terhadap seorang manajer distrik yang berada pada *region* yang dipimpinnya.

Temuan penelitian ini menunjukkan bahwa kepribadian altruistik penilai berpengaruh terhadap bias kemurahan hati. Hasil penelitian ini dapat berkontribusi baik bagi pengembangan literatur maupun praktik akuntansi manajemen melalui beberapa cara. Pertama, penelitian ini dapat memberikan kontribusi bagi literatur mengenai bias kemurahan hati dalam penilaian kinerja dengan menguji secara empiris dampak perilaku altruistik terhadap penilaian kinerja yang diberikan. Pada beberapa literatur (Giebe dan Gürtler 2012; Golman dan Bhatia 2012; Grund dan Przemeck 2012), perilaku altruistik manajer digunakan sebagai asumsi dalam mendesain kontrak yang optimal dan memitigasi bias kemurahan hati.

Kedua, hasil penelitian ini memberikan kontribusi praktis bagi upaya mengurangi bias kemurahan hati dalam penilaian kinerja. Berbeda dengan Chen (2012) yang mengusulkan pendekatan terkomputerisasi dengan menggunakan *Multi-Criteria Data*

Analysis (MCDA) berbasis *Interval-Value Fuzzy Sets* (IVFS) untuk mengurangi bias kemurahan hati, penelitian ini mengusulkan pendekatan yang lebih aplikatif bagi organisasi bisnis, melalui usulan untuk menyediakan kriteria penilaian kinerja subjektif yang lebih jelas dan penggunaan grup penilai untuk melakukan tugas penilaian kinerja dalam upaya untuk mengurangi bias kemurahan hati. Chen (2012) mengakui bahwa pendekatan MCDA merupakan sebuah pendekatan yang ideal secara model, namun sangat sulit untuk diaplikasikan dalam perusahaan. Dengan demikian, penelitian ini menjadi salah satu upaya untuk menjawab tantangan bagi para peneliti akuntansi manajemen untuk menghasilkan pendekatan yang lebih aplikatif bagi masalah-masalah yang dialami pada praktiknya (Birnberg 2009).

Penulisan artikel ini diorganisasi sebagai berikut: Bagian 1 menjelaskan pendahuluan. Bagian 2 menjelaskan teori yang digunakan dalam pengembangan hipotesis. Bagian 3 menguraikan tentang metodologi eksperimen yang digunakan. Bagian 4 memaparkan hasil penelitian, dan Bagian 5 menguraikan kesimpulan, keterbatasan, dan saran bagi penelitian selanjutnya.

TELAAH LITERATUR DAN PENGEMBANGAN HIPOTESIS

Teori Penilaian Kinerja (*Theory of Rating*)

Wherry dan Bartlett (1982) mengemukakan teori penilaian kinerja yang menjelaskan faktor-faktor selain kinerja aktual pihak yang dinilai yang memengaruhi penilaian kinerja dan mengidentifikasi metode yang dapat digunakan untuk mengeliminasi bias yang dialami oleh penilai. Keakuratan penilaian kinerja yang diberikan bergantung pada tiga langkah yakni: (1) kinerja pihak yang dinilai; (2) pengamatan penilai terhadap kinerja; serta (3) ingatan penilai mengenai pengamatan yang dilakukan.

Kinerja pihak yang dinilai adalah kombinasi dari kemampuan/keahlian sesungguhnya dengan pengaruh lingkungan. Apa yang diamati oleh penilai adalah fungsi dari

kinerja pihak yang dinilai dan bias amatan, sedangkan apa yang diingat oleh penilai adalah hasil dari semua amatan dikombinasikan dengan bias ingatan. Dengan demikian, keakuratan penilaian dipengaruhi oleh bias baik bias amatan maupun bias ingatan. Bias bersumber dari kontak yang tidak relevan sehingga komponen sesungguhnya dari bias (bias persepsi maupun bias ingatan) hanya akan terjadi bila pengalaman yang tidak relevan lebih dominan daripada amatan sesungguhnya.

Dari beberapa usulan yang dikemukakan oleh Wherry dan Bartlett (1982) untuk mengurangi bias dalam penilaian kinerja, terdapat dua usulan yang relevan dalam upaya mengurangi bias pada penilaian kinerja yakni menggunakan skala yang lebih jelas atau tidak ambigu dan menggunakan penilai yang plural. Penggunaan kriteria yang tidak ambigu akan membuat penilai dapat lebih memusatkan perhatian pada informasi-informasi yang sesuai dengan kriteria yang telah ditetapkan. Bahkan, ketika penggunaan *judgement* lebih diandalkan dalam penilaian kinerja subjektif, ketersediaan kriteria yang jelas diyakini akan mengarahkan *judgement* penilai pada informasi-informasi nonkuantitatif yang sesuai dengan kriteria yang telah ditetapkan.

Selain kriteria kinerja yang jelas, penggunaan grup yang terdiri dari beberapa penilai diyakini akan membuat bias umum yang dimiliki secara individual akan mendekati nol. Hal ini disebabkan ketika beberapa individu bertemu dalam suatu diskusi atau aktivitas bersama, terjadi proses yang menyebabkan bias-bias individu terkonfirmasi dan menjadi minimal.

Bias Kemurahan Hati

Salah satu bias penilaian kinerja yang umum dikenal dalam literatur ekonomi dan psikologi adalah *leniency*. Hartono (2013) menggunakan bias kemurahan hati sebagai padanan kata dari *leniency*. Guilford (1954) dalam Hartono (2013) mendefinisikan bias kemurahan hati sebagai “suatu tendensi dari penilai-penilai untuk menilai lebih tinggi dari seharusnya pada orang-orang yang mereka sudah mengenalnya dengan baik atau orang-

orang yang mereka inginkan terlibat”. Definisi *leniency* juga dikemukakan oleh Taylor dan Hastman (1956) dalam Bol (2011) sebagai memberikan penilaian kinerja rata-rata di atas titik tengah. Dengan demikian, bias kemurahan hati dapat dimaknai sebagai kecenderungan manajer untuk memberikan penilaian kinerja yang lebih tinggi kepada bawahan dibanding dengan kinerja yang mampu mereka berikan.

Temuan Bol dan Smith (2011) menunjukkan bahwa penilai mengalami bias kognisi akibat adanya informasi mengenai kinerja objektif yang disediakan. Adanya efek aura (*halo-effect*) yang dihasilkan oleh informasi kinerja penjualan menyebabkan penilaian kinerja subjektif (kinerja administrasi kantor) ikut terpengaruh. Terdapat empat kondisi yang digunakan untuk memanipulasi variabel keterkendalian dan tingkat informasi objektif. Menariknya, ketika partisipan diberi informasi skor kinerja objektif rendah dan keterkendalian rendah, partisipan justru meningkatkan penilaian kinerja administrasi kantor. Padahal, selaras dengan temuan Bol dan Smith (2011) mengenai distorsi kognitif yang dialami oleh evaluator, skor kinerja administrasi yang diberikan oleh partisipan seharusnya tidak melewati titik tengah skala yang ditetapkan karena terdistorsi oleh informasi kinerja objektif yang tersedia (tingkat penjualan individual rendah). Penilaian kinerja yang dielevasi ini menunjukkan terjadinya bias kemurahan hati.

Terdapat beberapa alasan yang mendasari terjadinya bias dalam penilaian kinerja, antara lain pengumpulan informasi, penghargaan, konsekuensi negatif, favoritisme, dan keterbatasan kognitif (Bol 2008). Beberapa temuan empiris juga mengidentifikasi alasan dan prediktor terjadinya bias kemurahan hati, antara lain ketidakpastian dan ketidaksimetrisan antara sinyal informasi kinerja yang dihasilkan oleh bawahan dengan skor kinerja yang diberikan oleh penilai sebagai akibat dari upaya penilai untuk menghindari ketidakadilan ketika menilai kinerja (Golman dan Bhatia 2012); biaya mengumpulkan informasi dan hubungan antara manajer dengan karyawan (Bol 2011); banyaknya ukuran kinerja yang digunakan

(Moers 2005); dan watak atau karakter penilai (Bernardin et al. 2000).

Kriteria Kinerja Subjektif

Untuk membantu penilai dalam melakukan penilaian kinerja subjektif, Van der Stede et al. (2006) mengusulkan beberapa kriteria untuk menilai kinerja subjektif yang diperoleh dari studi yang melibatkan para manajer dan direktur pada perusahaan manufaktur. Para manajer dan direktur ini diminta untuk menunjukkan kriteria-kriteria yang kemungkinan digunakan oleh atasan untuk menilai kinerja mereka. Dari hasil studi ini, diperoleh beberapa kriteria kinerja subjektif, yaitu: (1) perspektif jangka panjang mengenai bisnis; (2) kemampuan untuk secara efektif mendapatkan pengetahuan atau keahlian yang baru; (3) kemampuan untuk berbagi pengetahuan dalam organisasi; (4) kerja sama dengan departemen lain dalam organisasi; (5) semangat/moral karyawan dalam departemen; (6) manajemen/kemampuan kepemimpinan; dan (7) kesetiaan terhadap perusahaan.

Ketujuh kriteria yang dikemukakan oleh Van der Stede et al. (2006) tersebut kemudian dilengkapi oleh Bellavance et al. (2013) dengan menambahkan tiga kriteria yang diperoleh dari hasil penelusuran terhadap kriteria kinerja yang digunakan oleh industri-industri lain. Kriteria itu terdiri dari: (1) kemampuan untuk secara efektif mengembangkan kompetensi utama; (2) kemampuan komunikasi; serta (3) dukungan dan perwujudan nilai-nilai dalam organisasi.

Dalam penelitian ini, kesepuluh kriteria kinerja subjektif yang digunakan oleh Van der Stede et al. (2006) dan Bellavance et al. (2013) digunakan sebagai kriteria untuk menilai kinerja manajer. Beberapa penyesuaian dilakukan untuk mengubah perspektif penilai. Bila dalam penelitian Van der Stede et al. (2006) dan Bellavance et al. (2013), partisipan merupakan pihak yang dinilai, sebaliknya partisipan dalam penelitian ini menjadi penilai.

Berdasarkan argumen teori penilaian kinerja, dapat dinyatakan bahwa ketersediaan kriteria subjektif dalam membantu penilai melakukan evaluasi kinerja kemungkinan dapat

mengarahkan kognisi penilai kepada informasi-informasi yang sesuai dengan kriteria yang akan dinilai. Hal ini diyakini dapat mengurangi dampak kontaminasi yang ditimbulkan oleh informasi-informasi yang sebenarnya tidak relevan kala melakukan evaluasi kinerja subjektif sehingga meningkatkan keakuratan penilaian kinerja. Individu yang diberi informasi mengenai kriteria kinerja yang dapat digunakan ketika mengevaluasi kinerja individu lain kemungkinan tidak akan menaikkan penilaian kinerja. Berdasarkan uraian yang dikemukakan, hipotesis dalam penelitian ini adalah:

H₁: Sebelum aktivitas diskusi grup, partisipan yang mendapatkan informasi mengenai kriteria kinerja subjektif akan memberikan skor kinerja subjektif lebih rendah daripada partisipan yang tidak mendapatkan informasi mengenai kriteria kinerja subjektif.

Grup Penilai (*Group Rater*)

Grup telah digunakan dalam berbagai *setting* penelitian, misalnya desain insentif (Chen et al. 2012), pengambilan keputusan (Galini 2013; Kocher dan Sutter 2007; Bosman et al. 2006; Cappel dan Windsor 2000), maupun dalam upaya mengurangi bias penilaian kinerja (Martell dan Leavitt 2002; Martell dan Borg 1993). Dalam *setting* pengambilan keputusan, grup ditemukan lebih menghasilkan keputusan yang efektif bahkan dalam kondisi ketidakpastian (*risky-choice*) (Charness et al. 2007).

Penggunaan aktivitas grup untuk meningkatkan keakuratan penilaian kinerja dan mengurangi bias dilakukan oleh Martell dan Leavitt (2002). Mereka menemukan bahwa penggunaan grup penilai lebih mampu mengurangi *performance-cue bias* dibanding penilai individual. Grup juga dapat mengurangi bias korespondensi yang dialami oleh penilai individual (Wright dan Wells 1985), serta mengurangi penggunaan pemikiran heuristik sehingga individu yang terlibat dalam grup dapat lebih memproses informasi secara sistematis dan dalam jika dibanding dilakukan secara individual.

Gibson dan Earley (2007) menjelaskan bahwa informasi dan pengetahuan digunakan dalam seluruh proses kognitif kolektif. Aktivitas grup yang memanfaatkan seluruh pengetahuan dan informasi yang diperoleh grup akan meningkatkan proses kognitif kolektif. Proses kognitif kolektif ini diyakini akan meningkatkan kemanjuran diri (*self-efficacy*) baik secara individual maupun kolektif (Lin et al. 2014). Kemanjuran diri (*self-efficacy*) yang diperoleh individu maupun grup kemungkinan akan memengaruhi keakuratan tugas penilaian kinerja yang diberikan.

Beberapa penelitian yang menggunakan grup telah menunjukkan pergeseran dari interaksi grup secara tatap muka (*face-to-face*) menjadi interaksi grup yang memanfaatkan teknologi komputer (Chen et al. 2012; Chen dan Linn 2012; Bosman et al. 2006; Cappel dan Windsor 2000). Penelitian ini memanfaatkan teknologi komputer untuk memfasilitasi aktivitas grup. Pemanfaatan teknologi komputer untuk mendukung keputusan grup ini sangat bermanfaat dalam memfasilitasi usaha kolaboratif dalam pengambilan keputusan strategis (Chen dan Linn 2012) karena memiliki berbagai keunggulan, antara lain meningkatkan kemampuan anggota grup untuk memberikan kontribusi ide atau komentar secara simultan tanpa diinterupsi oleh pihak lain, menyediakan dokumentasi mengenai pertemuan grup, serta mendukung interaksi pada grup yang terdistribusi misalnya grup yang memiliki anggota yang terpisah secara geografis (Cappel dan Windsor 2000). Diskusi grup dengan memanfaatkan teknologi komputer juga bermanfaat dalam menjamin anonimitas, mengatasi kendala ketidakmampuan melakukan diskusi oral dan efek tekanan superioritas anggota tertentu terhadap anggota lain dalam grup (Huang dan Li 2007).

Pengaruh aktivitas grup terhadap penilaian kinerja yang diberikan dapat dilihat baik secara individual maupun kolektif. Dampak individual ditunjukkan melalui pemberian penilaian kinerja oleh penilai individual sebelum maupun sesudah aktivitas grup, sedangkan secara kolektif ditunjukkan melalui

penilaian kinerja konsensus yang dihasilkan. Bagi individu, aktivitas diskusi grup diharapkan dapat membuat bias individual yang dialami terminimalkan. Individu akan mendapatkan pengetahuan mengenai penilaian yang lebih baik. Proses diskusi ini diyakini juga dapat mengubah penilaian/keyakinan awal yang dimiliki oleh individu. Dalam melakukan penilaian subjektif, diharapkan setelah proses berdiskusi individu akan memberikan penilaian kinerja yang lebih rendah. Dari uraian yang dikemukakan sebelumnya, maka hipotesis penelitian dirumuskan sebagai berikut:

H₂: Pada kondisi tidak tersedia informasi kinerja subjektif, skor kinerja subjektif yang diberikan oleh partisipan sesudah diskusi lebih rendah daripada skor yang diberikan oleh partisipan sebelum diskusi.

Upaya memaksimalkan pemanfaatan aktivitas grup untuk mengurangi bias dan meningkatkan keakuratan penilaian kinerja dapat dilakukan dengan menyediakan informasi mengenai kriteria kinerja subjektif yang relevan dengan tugas penilaian. Penyediaan informasi ini diharapkan akan meningkatkan kehati-hatian dalam memberikan penilaian kinerja. Hal ini disebabkan kriteria yang relevan dapat menjadi sumber berbagai informasi dan saling mengkritisi di antara para anggota grup penilai. Melalui proses itu, para penilai diharapkan akan mampu memisahkan antara informasi kinerja yang relevan dengan tugas penilaian dengan yang tidak relevan. Mereka diharapkan akan mengabaikan informasi yang tidak relevan dengan tugas sehingga ketika informasi yang relevan dengan tugas menunjukkan bahwa kinerja subjektif manajer tidak cukup memadai, para penilai diharapkan akan memberikan penilaian kinerja yang semakin rendah. Dengan demikian, hipotesis dirumuskan sebagai berikut:

H_{3a}: Setelah aktivitas diskusi grup, skor kinerja subjektif yang diberikan oleh grup yang memperoleh informasi mengenai kriteria kinerja subjektif lebih rendah daripada skor kinerja yang diberikan oleh grup yang tidak

memperoleh informasi mengenai kriteria kinerja.

H_{3b}: Setelah aktivitas diskusi grup, skor kinerja subjektif yang diberikan oleh partisipan yang memperoleh informasi mengenai kriteria kinerja subjektif lebih rendah daripada skor setelah aktivitas diskusi grup yang diberikan oleh partisipan yang tidak memperoleh informasi mengenai kriteria kinerja.

H_{3c}: Pada kondisi tersedia kriteria kinerja subjektif, skor kinerja subjektif yang diberikan oleh partisipan setelah aktivitas diskusi grup lebih rendah daripada sebelum aktivitas diskusi grup.

Kepribadian Altruistis dan Bias Kemurahan Hati

Kepribadian altruistis sering didefinisikan secara berbeda oleh beberapa penulis. Namun, Simmons (1991) meyakini bahwa meskipun terdapat perbedaan dalam definisi yang dikemukakan, para penulis sepakat bahwa altruistis adalah kepribadian individu yang: (1) berupaya untuk meningkatkan kesejahteraan/memberi manfaat bagi orang lain, bukan bagi diri sendiri; (2) dilakukan secara sukarela; (3) bertujuan, artinya perilaku ini dimaksudkan untuk menolong orang lain; (4) tidak mengharapkan penghargaan pihak eksternal.

Manusia diasumsikan memiliki keterbatasan dalam memperoleh pengetahuan dan membuat kalkulasi yang dibutuhkan untuk menghasilkan pilihan yang optimal, bahkan meskipun mereka memiliki pengetahuan yang sempurna. Hal ini disebabkan informasi yang dimiliki mengenai variabel-variabel lingkungan, masa kini dan masa akan datang, tidak lengkap dan tidak pasti sehingga akan memengaruhi konsekuensi dari tindakan/pilihan yang diambil. Akibatnya, mereka cenderung menyandarkan pilihan atau keputusan yang akan diambil pada saran, rekomendasi, persuasi, dan informasi yang diperoleh melalui saluran sosial yang tersedia (Simmons 1993).

Penilaian kinerja subjektif merupakan sebuah proses yang kompleks karena membutuhkan pertimbangan individu terhadap berbagai informasi baik yang relevan maupun tidak relevan. Dalam proses ini, keterbatasan rasionalitas individu akan membawa konsekuensi yang luas terutama terhadap penilaian kinerja yang diberikan. Keterbatasan ini kemungkinan akan menyebabkan penilaian kinerja terdistorsi oleh adanya informasi yang tidak relevan juga ketidakmampuan untuk memproses informasi secara sempurna.

Literatur mengenai penilaian kinerja subjektif menghubungkan kepribadian altruistis manajer dengan penilaian kinerja yang tinggi (Giebe dan Gürtler 2012; Golman dan Bhatia 2012; Grund dan Przemeczek 2012). Kepribadian altruistis kemungkinan akan memperparah distorsi kognitif yang dialami kala memberikan penilaian kinerja. Manajer yang altruistis akan berupaya untuk membantu bawahannya dengan memberikan penilaian kinerja yang tinggi. Penilaian kinerja ini berkaitan erat dengan kesejahteraan yang akan diperoleh. Sebaliknya, manajer itu akan menanggung dampak yang timbul dari distorsi penilaian kinerja yang diberikan. Berdasarkan uraian di atas, maka hipotesis dirumuskan sebagai berikut:

H₄: Sifat altruistis penilai berpengaruh positif terhadap skor kinerja subjektif yang diberikan oleh penilai.

METODE PENELITIAN

Partisipan

Penelitian ini berbentuk eksperimen dengan menggunakan 78 orang mahasiswa yang mengambil kelas metodologi penelitian pada program studi akuntansi di Universitas Gadjah Mada sebagai partisipan. Mahasiswa diminta secara sukarela untuk berpartisipasi dengan diberi stimuli bahwa eksperimen ini dapat dimanfaatkan sebagai sarana untuk mempraktikkan pengetahuan mereka mengenai eksperimen yang diperoleh di kelas. Dari hasil pemeriksaan ditemukan satu orang partisipan tidak menyelesaikan tugas eksperimen dan delapan orang partisipan tidak lolos cek

manipulasi. Total partisipan yang digunakan dalam pengujian berjumlah 69 orang.

Desain Eksperimen

Penelitian ini menggunakan desain eksperimen yang dikembangkan oleh Bol dan Smith (2011) yang dimodifikasi. Eksperimen dilakukan melalui dua fase utama yang terdiri dari fase pendahuluan dan fase untuk menguji perlakuan terhadap partisipan. Pada fase pendahuluan, partisipan diminta untuk menjawab dua puluh pertanyaan mengenai kepribadian altruistik yang diadaptasi dari Rushton et al. (1981). Instrumen ini mengukur tingkat kecenderungan altruistik individu. Program komputer kemudian menghitung skor rata-rata tiap partisipan dan secara acak memasukkan mereka ke dalam salah satu grup berisi tiga orang. Grup terdiri dari partisipan yang memiliki tingkat altruistik tinggi maupun rendah. Setelah terbentuk grup, eksperimen memasuki fase kedua.

Fase kedua terdiri dari dua sesi. Sesi pertama adalah sesi penilaian individual. Untuk memulai sesi ini, grup-grup yang telah terbentuk dimasukkan ke dalam dua kondisi yang berbeda yaitu grup yang menerima informasi kriteria kinerja subjektif dan grup yang tidak menerima kriteria kinerja subjektif. Material eksperimen ditampilkan pada layar komputer. Partisipan diminta untuk mengasumsikan bahwa dirinya bekerja sebagai direktur regional pada sebuah perusahaan yang bergerak dalam bidang penjualan pipa dan perangkat pendukungnya. Direktur regional membawahi 10 manajer distrik yang bertanggung jawab terhadap staf penjualan dan kantor. Salah satu tugas dari direktur regional adalah melakukan evaluasi terhadap kinerja manajer distrik. Partisipan diberitahu bahwa tugas mereka adalah melakukan evaluasi kinerja terhadap seorang manajer distrik.

Untuk evaluasi subjektif, partisipan diberitahu bahwa ada catatan pribadi dan hasil wawancara dengan beberapa anggota staf mengenai kinerja sehari-hari manajer distrik itu. Partisipan yang berada pada kondisi tersedia kriteria kinerja subjektif mendapatkan informasi tambahan berupa kriteria kinerja

subjektif yang terdiri dari beberapa kriteria yang diadaptasi dari Bellavance et al. (2013), sedangkan partisipan yang berada pada kondisi tidak tersedia kriteria kinerja subjektif tidak mendapat informasi mengenai kriteria kinerja subjektif. Kemudian, partisipan diminta untuk mengevaluasi dan memberikan penilaian kinerja berdasarkan informasi subjektif itu. Pada sesi pertama ini, partisipan tidak diizinkan untuk berinteraksi baik dengan sesama anggota grup maupun dengan anggota grup lain.

Pada sesi kedua, partisipan diarahkan untuk terlibat dalam aktivitas diskusi bersama anggota grup mengenai penilaian kinerja yang seharusnya diberikan kepada manajer distrik yang mereka evaluasi. Interaksi grup sepenuhnya dikendalikan oleh komputer. Hal ini dimaksudkan untuk memberikan kesempatan yang sama kepada semua anggota untuk mengemukakan pendapat mereka, juga mencegah interupsi yang kemungkinan terjadi bila dilakukan secara tatap muka maupun kecenderungan sikap mendominasi yang dimiliki oleh individu tertentu.

Untuk menjamin anonimitas dan mendukung proses diskusi, identitas partisipan diganti dengan menggunakan kombinasi empat buah huruf, misalnya ABCD, AABB, AABC. Setelah partisipan menyelesaikan tugas eksperimen, mereka diminta untuk menjawab beberapa pertanyaan untuk menguji manipulasi. Eksperimen diakhiri dengan penjelasan yang diberikan kepada partisipan mengenai manipulasi dalam tugas eksperimen yang telah mereka kerjakan. Langkah ini merupakan *debriefing* yang umum dilakukan dalam eksperimen dengan tujuan untuk memastikan bahwa partisipan tidak mengalami masalah setelah mengikuti eksperimen. Selain itu, hal ini dilakukan untuk meningkatkan pemahaman mahasiswa yang menjadi partisipan dalam penelitian ini tentang metode eksperimen.

Pengukuran Variabel

Dalam eksperimen ini, terdapat dua variabel independen yang dimanipulasi yaitu kriteria kinerja subjektif dan penilai. Kriteria kinerja subjektif dimanipulasi melalui tersedia

atau tidaknya kriteria bagi penilai. Terdapat 10 kriteria kinerja subjektif yang diadaptasi dari penelitian Bellavance et al. (2013).

Penilai (*rater*) adalah pihak yang melakukan evaluasi atas kinerja manajer distrik. Variabel penilai dimanipulasi menggunakan penilai individual dan penilai grup. Penilai grup terdiri dari tiga orang partisipan.

Kepribadian altruistis diukur dengan menggunakan 20 buah pertanyaan yang dikembangkan oleh Rushton et al. (1981) dengan menggunakan skala satu (tidak pernah) hingga lima (sangat sering). Terdapat satu buah pernyataan yang disesuaikan dalam instrumen ini, yakni pernyataan "*I have helped push a stranger's car out of the snow*" diterjemahkan menjadi "Saya telah membantu mendorong mobil orang tidak dikenal". Penyesuaian pada item pertanyaan ini lebih didasarkan pada pertimbangan konteks Indonesia. Nilai *loading factor* berkisar antara 0,536 hingga 0,772 dan nilai *Cronbach's Alpha* sebesar 0,828. Kedua nilai ini menunjukkan bahwa instrumen cukup layak untuk digunakan (Hair et al. 2010).

Variabel dependen dalam penelitian ini adalah evaluasi subjektif terhadap kinerja manajer distrik. Setelah mempelajari informasi mengenai catatan pribadi dan hasil wawancara, partisipan diminta untuk memberikan nilai terhadap kinerja manager dengan skala 0-10. Bagi penilai grup, mereka akan diberitahu bahwa penilaian kinerja yang diberikan adalah penilaian kinerja yang disepakati oleh para anggota grup.

HASIL PENELITIAN DAN PEMBAHASAN

Hasil Pengecekan Manipulasi

Pengecekan manipulasi dilakukan untuk mengetahui apakah partisipan memahami dan mampu menginterpretasikan manipulasi yang mereka terima. Pengecekan manipulasi ini dilakukan melalui dua tahap yaitu tahap pertama meminta partisipan merespons pertanyaan apakah David Sutton (manajer yang dinilai) adalah orang yang suka tersenyum dan bersikap terbuka terhadap bawahannya, sedangkan tahap kedua dilakukan dengan:

(1) meminta partisipan untuk memberikan penilaian mengenai kinerja penjualan relatif manajer yang mereka nilai (David Sutton) dibanding manajer distrik yang lain (menggunakan skala 7 poin); (2) meminta persetujuan (menggunakan skala 7 poin) partisipan terhadap pernyataan bahwa catatan pribadi dan respons wawancara mengenai manajer target memberikan pertimbangan yang memadai mengenai kinerja administrasi kantor manajer bersangkutan; (3) meminta partisipan menunjukkan tingkat keyakinan (menggunakan skala 7 poin) terhadap penilaian yang telah mereka lakukan.

Pengecekan tahap pertama dilakukan secara manual. Partisipan yang melakukan kesalahan dalam menjawab dinyatakan tidak lolos ke pengecekan selanjutnya. Hasil pengecekan manipulasi tahap kedua yang ditampilkan pada Tabel 1 menunjukkan nilai *mean* penilaian kinerja penjualan individual sebesar 2,38 (s.d. = 1,318) secara signifikan lebih rendah dari titik tengah skala yang digunakan yaitu 4 ($p < 0,01$). Hal ini menunjukkan bahwa partisipan memahami bahwa kinerja objektif (penjualan individual) yang dihasilkan oleh manajer distrik yang dinilai adalah rendah. Hasil pengecekan manipulasi kedua nilai *mean* sebesar 5,00 (s.d. = 1,029) secara signifikan lebih tinggi dari nilai tengah skala yang digunakan yaitu 4 ($p < 0,01$). Hasil ini menunjukkan bahwa partisipan yakin informasi mengenai kinerja administrasi kantor yang disediakan cukup memadai untuk memberikan pertimbangan mengenai kinerja manajer yang dinilai. Tingkat keyakinan atas skor yang diberikan sebelum diskusi menunjukkan nilai *mean* sebesar 5,333 (s.d. = 0,9341) dan sesudah diskusi sebesar 5,667 (s.d. = 1,133) secara signifikan lebih tinggi dari nilai tengah skala yang digunakan yaitu 4 ($p < 0,01$). Hasil ini menunjukkan bahwa partisipan yakin dengan skor yang diberikan.

Pengecekan manipulasi tahap kedua dilakukan pada partisipan yang berada pada kondisi tersedia kriteria kinerja, dengan cara meminta mereka untuk memberikan persetujuan terhadap beberapa pernyataan mengenai pentingnya kriteria kinerja subjektif

Tabel 1
Hasil Pengecekan Manipulasi

Cek Manipulasi	Mean	Std. Deviation	Sig.
TkKeySblm	5.333	0.9341	0.000
TkKeySsdh	5.667	1.1333	0.000
Cm1	2.38	1.318	0.000
Cm2	5.00	1.029	0.000
Cm3	6.50	0.609	0.000
Cm4	6.69	0.467	0.000
Cm5	6.42	0.649	0.000
Cm6	6.58	0.604	0.000
Cm7	6.50	0.655	0.000

Keterangan:

TkkeySblm : tingkat keyakinan terhadap skor kinerja yang diberikan sebelum diskusi

TkkeySsdh : tingkat keyakinan terhadap skor kinerja yang diberikan sesudah diskusi

Cm 1-7 : pengecekan manipulasi 1-7

yang ditampilkan. Pengecekan manipulasi ini dimaksudkan untuk mengetahui apakah partisipan memahami informasi mengenai kriteria kinerja subjektif yang mereka terima dan mempertimbangkannya ketika melakukan tugas penilaian. Hasil pengecekan menunjukkan nilai *mean* sebesar 6,50 (s.d. = 0,609), 6,69 (s.d. = 0,467), 6,42 (s.d. = 0,649), 6,58 (s.d. = 0,604), dan 6,50 (s.d. = 0,655) secara signifikan lebih tinggi dari nilai tengah skala yang digunakan yaitu 4 ($p < 0,01$). Hasil ini menunjukkan bahwa partisipan mempertimbangkan kriteria kinerja yang disediakan ketika melakukan penilaian.

Statistik Deskriptif

Hasil statistik deskriptif (Tabel 2) menunjukkan bahwa komposisi partisipan terdiri dari 26 orang laki-laki dan 43 orang perempuan. Rata-rata partisipan berusia di bawah 22 tahun. Hasil pengujian rata-rata skor altruistis antar kondisi menunjukkan bahwa tidak terdapat perbedaan yang signifikan dalam tingkat altruistis pada tiap kondisi ($p\text{-value} = 0,680$). Perbandingan rata-rata skor kinerja administrasi kantor antara laki-laki dan perempuan juga menunjukkan tidak terdapat perbedaan yang signifikan dalam skor yang diberikan oleh partisipan laki-laki maupun perempuan ($p\text{-value} = 0,120$). Meskipun jumlah partisipan laki-laki dan perempuan berbeda, namun hasil pengujian menunjukkan

kesamaan dengan temuan empiris Maas dan Torrez-González (2011). Mereka menemukan bahwa skor kinerja subjektif yang diberikan tidak dipengaruhi oleh jenis kelamin evaluator. Hasil pengujian juga menunjukkan tidak terdapat perbedaan skor kinerja objektif (penjualan individual) antar kondisi baik sebelum maupun sesudah diskusi. Nilai *p-value* sebelum diskusi sebesar 0,541 dan sesudah diskusi sebesar 0,268.

Dari hasil pengujian yang ditampilkan, dapat disimpulkan bahwa semua partisipan yang ditempatkan pada setiap kondisi perlakuan memiliki kesamaan ciri (ekuivalen). Hal ini dilakukan untuk memastikan bahwa ancaman terhadap validitas internal yaitu seleksi telah dapat dikendalikan sehingga dapat diyakini bahwa perbedaan hasil yang diperoleh benar-benar berasal dari perlakuan yang diterima bukan disebabkan oleh ketidaksepadanan subjek (Nahartyo 2012).

Pengujian Hipotesis

Untuk menguji hipotesis penelitian digunakan *Analysis of Variance* (ANOVA). Pengujian ini bertujuan untuk menguji perbedaan nilai rata-rata antara grup yang mendapatkan perlakuan yang berbeda. Hipotesis 1 menguji dampak ketersediaan informasi mengenai kriteria kinerja subjektif terhadap penilaian kinerja yang diberikan oleh penilai individual. Hasil pengujian *between subject-*

Tabel 2
Tabel Statistik Deskriptif

Keterangan	Jumlah	Persentase	Kondisi		Hasil Uji F	
			Kriteria Kinerja	Tidak Ada Kriteria	F	Nilai P
Jenis kelamin:						
Laki-laki	26 orang	38%				
Perempuan	43 orang	62%				
Usia:						
19-21 tahun	69 orang	100%				
Mean Skor Adm. Kantor:						
Laki-laki	6.654				2.480	0.120
Perempuan	6.818					
Rata-rata skor altruistik:			2.7681	2.7212	0.171	0.680
Rata-rata skor kinerja objektif:						
Sebelum diskusi			4.636	4.262	0.377	0.541
Sesudah diskusi			3.11	2.67	1.246	0.268

Tabel 3
Hasil Pengujian *between subject*-ANOVA

Kondisi Kriteria Subjektif	Levene's Test	Mean	Std. Deviation	F	Sig.
Tersedia	0.498	6.633	1.1710	3.955	0.051
Tidak tersedia		6.061	1.2166		

Analysis of Variance (Tabel 3) menunjukkan perbedaan skor administrasi kantor sebelum diskusi yang diberikan oleh partisipan yang memperoleh informasi mengenai kriteria kinerja subjektif dengan partisipan yang tidak memperoleh kriteria kinerja subjektif signifikan ($p\text{-value} = 0,051$). Namun, nilai *mean* skor administrasi kantor yang diberikan oleh partisipan yang memperoleh kriteria kinerja subjektif (6,633) lebih tinggi dibanding skor administrasi kantor yang diberikan oleh partisipan yang tidak memperoleh informasi mengenai kriteria kinerja subjektif (6,061). Hasil ini menunjukkan bahwa Hipotesis 1 tidak terdukung. Dengan demikian, dapat dinyatakan bahwa informasi mengenai kriteria kinerja subjektif tidak efektif untuk mengurangi bias kemurahan hati yang dialami oleh penilai individual.

Hipotesis 2 menguji dampak aktivitas grup terhadap penilaian kinerja subjektif yang diberikan oleh penilai individual. Hasil

pengujian *within subject-analysis of variance* (Tabel 4) menunjukkan *mean* skor kinerja administrasi yang diberikan sebelum diskusi sebesar 6,061 dan *mean* sesudah diskusi sebesar 5,485. Kedua skor administrasi kantor ini berbeda secara signifikan ($p\text{-value} < 0,05$). Hasil ini menunjukkan bahwa Hipotesis 2 terdukung. Terdukungnya hipotesis ini menunjukkan bahwa aktivitas diskusi grup efektif dalam mengurangi bias kemurahan hati yang dialami oleh para penilai individual.

Untuk menguji dampak ketersediaan kriteria kinerja subjektif dan aktifitas grup terhadap penilaian kinerja yang diberikan oleh penilai baik secara individual maupun grup, terdapat tiga hipotesis yang diuji. Hasil pengujian *between subject-analysis of variance* (Tabel 5) menunjukkan *mean* skor administrasi kantor yang diberikan oleh grup yang memperoleh informasi mengenai kriteria kinerja subjektif sebesar 5,962 dan grup yang

Tabel 4
Hasil Pengujian *within subject*-ANOVA

Kondisi Diskusi	Mean	Std. Deviation	F	Sig.
Sebelum	6.061	1.1710	9.197	0.005
Sesudah	5.485	1.2530		

Tabel 5
Hasil Pengujian *between subject*-ANOVA

Kondisi Kriteria Kinerja	Levene's Test	Mean	Std. Deviation	F	Sig.
Grup/tidak tersedia	0.138	5.692	0.7511	1.991	0.171
Grup/tersedia		6.269	1.2685		
Individu/tidak tersedia	0.988	5.485	1.2530	3.172	0.079
Individu/tersedia		6.028	1.2758		

tidak memperoleh informasi mengenai kriteria kinerja subjektif sebesar 6,269. Kedua mean skor administrasi kantor ini tidak berbeda secara signifikan ($p\text{-value} > 0,05$). Hasil ini tidak mendukung Hipotesis 3a.

Hasil pengujian terhadap *mean* skor administrasi kantor setelah diskusi yang diberikan oleh penilai yang memperoleh kriteria kinerja dengan penilai yang tidak memperoleh kriteria kinerja (Tabel 5) menunjukkan nilai sebesar 5,485 dan 6,028 pada tingkat signifikansi sebesar 0,079 ($p\text{-value} < 0,1$). Hasil ini menunjukkan dukungan marjinal terhadap Hipotesis 3b.

Hipotesis 3c menguji dampak proses diskusi dalam grup dan penyediaan informasi mengenai kriteria kinerja subjektif terhadap *rating* individual yang dihasilkan. Hasil pengujian menunjukkan (Tabel 6) *mean* sebelum diskusi sebesar 6,633 menurun menjadi 6,028 sesudah aktivitas diskusi grup. Penurunan ini ditemukan secara signifikan ($p\text{-value} < 0,05$). Dengan demikian, Hipotesis 3c terdukung. Berdasarkan hasil ini, maka dapat dinyatakan bahwa aktivitas diskusi grup yang memanfaatkan informasi mengenai kriteria kinerja subjektif efektif untuk mengurangi bias kemurahan hati yang dialami oleh penilai individual.

Hipotesis 4 memprediksikan bahwa

perilaku altruistis penilai berpengaruh terhadap skor administrasi kantor yang diberikan oleh partisipan. Untuk menguji Hipotesis 4, dilakukan pengujian regresi antara tingkat altruistis dengan skor administrasi kantor dan pengujian perbedaan antara tingkat altruistis dengan skor administrasi kantor. Hasil pengujian regresi (Tabel 7) menunjukkan bahwa sikap altruistis secara signifikan ($p\text{-value} = 0,000$) memengaruhi skor kinerja administrasi kantor yang diberikan oleh penilai. Hasil pengujian terhadap *mean* skor administrasi kantor antara partisipan yang memiliki sifat altruistis tinggi dengan altruistis rendah (Tabel 8) juga berbeda signifikan ($p\text{-value} < 0,01$). Sesuai hipotesis, untuk penilai yang mempunyai sifat altruistis tinggi skor penilaiannya lebih tinggi (6,66) dibanding yang sifat altruistisnya rendah (5,08). Dengan demikian, hasil ini mendukung Hipotesis 4.

SIMPULAN

Penelitian ini bertujuan untuk menguji dampak penyediaan informasi mengenai kriteria kinerja subjektif dan grup penilai dalam upaya mengurangi bias kemurahan hati dalam penilaian kinerja subjektif. Penelitian ini juga memberikan bukti empiris mengenai pengaruh

Tabel 6
Hasil Pengujian *within-subject*-ANOVA

Kondisi Diskusi	Mean	Std. Deviation	F	Sig.
Sebelum	6.633	1.2166	4.831	0.035
Sesudah	6.028	1.2758		

Tabel 7
Hasil Pengujian Regresi

Variabel Independen	B	R ²	t	F-test	p-value
Altruistis	1.023	0.279	5.093	25.937	0.000

Variabel dependen : Skor Penilaian Kinerja Subjektif

Tabel 8
Hasil Pengujian Beda Skor Penilaian Kinerja

Tingkat Altruistis	Levene's Test	Mean	Std. Deviation	F	Sig.
Tinggi	0.607	6.66	1.045	23.542	0.000
Rendah		5.08	1.115		

sifat altruistis penilai terhadap bias kemurahan hati. Sebelumnya, perilaku altruistis digunakan sebagai asumsi yang mendasari model matematis dalam mendesain kontrak yang optimal untuk memitigasi bias kemurahan hati (Giebe dan Gürtler 2012; Golman dan Bhatia 2012; Grund dan Przemec 2012). Dalam penelitian ini, kami menguji secara eksplisit asumsi ini.

Hasil penelitian menunjukkan bahwa sifat altruistis individu memengaruhi bias kemurahan hati. Temuan ini mengonfirmasikan asumsi yang umum digunakan bahwa kepribadian altruistik menyebabkan bias kemurahan hati. Temuan ini selaras dengan temuan pada riset psikologi yang mengindikasikan bahwa karakteristik tertentu individu memiliki kecenderungan terhadap bias penilaian kinerja baik dalam bentuk kemurahan hati maupun kompresi (Bernardin et al. 2000; Moers 2005).

Hasil penelitian ini mendukung usulan penggunaan grup penilai sebagai salah satu cara untuk mengurangi bias kemurahan hati. Hal ini ditunjukkan oleh skor kinerja administrasi yang diberikan oleh penilai yang tidak memperoleh informasi mengenai kriteria kinerja yang relevan dengan penilaian kinerja

yang sedang dilakukan pada saat sebelum dan sesudah diskusi. Skor kinerja administrasi yang diberikan oleh penilai setelah diskusi mengalami penurunan. Hal ini menunjukkan bahwa melalui proses diskusi yang terjadi, bias-bias individual yang dimiliki penilai terminimalkan. Hasil ini sejalan dengan argumen teori disonansi kognitif (Festinger 1957) yang mengemukakan bahwa individu mengalami disonansi antara dua atau lebih kognisi yang dimiliki. Kognisi dapat berupa keyakinan, sikap, emosi, persepsi, dan perilaku individu. Disonansi ini akan memaksa individu untuk mencapai konsonansi dengan beberapa cara. Salah satu cara yang digunakan adalah dengan merasionalisasi/mengubah persepsi mengenai tindakan yang diambil.

Dalam aktivitas penilaian kinerja yang melibatkan diskusi grup, individu penilai mengalami disonansi antara kognisi/persepsinya mengenai kinerja pihak yang dinilai dengan berbagai informasi yang diperoleh melalui diskusi. Informasi yang tersedia dalam diskusi berupa informasi yang positif maupun negatif mengenai pihak yang dinilai. Lowe et al. (2002) mengemukakan bahwa dalam membuat keputusan individu cenderung

lebih memerhatikan informasi negatif daripada informasi positif sehingga untuk mengatasi disonansi yang dialami, individu penilai cenderung akan merasionalisasikan persepsinya dengan menerima informasi negatif yang tersedia dan bermuara pada pemberian skor yang lebih rendah.

Meskipun aktivitas diskusi grup berdampak positif dalam mengurangi bias, penilaian kinerja yang diberikan oleh grup berupa penilaian kinerja konsensus ditemukan tidak efektif untuk mengurangi bias. Hal ini kemungkinan disebabkan grup terbentuk dari beberapa individu yang memiliki karakteristik yang berbeda. Misalnya, individu yang memiliki kepribadian altruistik tinggi cenderung menginginkan penilaian kinerja yang tinggi, sedangkan individu yang kepribadian altruistiknya rendah cenderung mengutamakan keakuratan (Golman dan Bhatia 2012). Menghasilkan penilaian kinerja konsensus merupakan sebuah proses yang sulit karena penilaian kinerja yang dihasilkan harus mampu memenuhi ekspektasi semua anggota (Roch 2006). Upaya untuk menghasilkan penilaian kinerja yang dapat diterima oleh semua anggota grup ini akan mengurangi keakuratan penilaian kinerja. Kesimpulan ini diambil berdasarkan hasil pengujian terhadap Hipotesis 3a. Unit analisis untuk pengujian ini adalah grup yang terdiri dari tiga orang. Untuk menjamin kesetaraan antar grup, partisipan dimasukkan secara acak ke dalam setiap grup. Tiap grup terdiri dari partisipan dengan tingkat altruistik tinggi maupun rendah. Dengan demikian, kontrol dilakukan pada level grup.

Temuan lain yang menarik dalam penelitian ini adalah ketidakefektifan kriteria kinerja subjektif untuk meningkatkan keakuratan penilaian kinerja seperti yang diusulkan oleh Wherry dan Bartlett (1982) serta Golman dan Bhatia (2012) tentang penggunaan kriteria yang lebih jelas sebagai salah satu upaya pengurangan bias kemurahan hati. Dari hasil pengujian, terlihat bahwa ketersediaan kriteria kinerja subjektif menghasilkan penilaian kinerja yang secara signifikan lebih tinggi daripada kondisi tidak tersedia. Teori Pembetulan Diri (*Self*

Justification Theory) (Staw 1975) dapat digunakan untuk menjelaskan temuan ini. Teori Pembetulan Diri berakar dari Teori Disonansi Kognitif (Festinger 1957). Individu diasumsikan bukanlah penerima informasi yang pasif sehingga informasi yang diperoleh itu akan ditransformasikan ke dalam keputusan/pilihan yang diambil. Pengambilan keputusan/penilaian kinerja adalah sebuah proses yang kompleks yang mengakibatkan terjadinya disonansi kognitif pada individu yang terlibat. Upaya untuk meredakan disonansi itu dilakukan dengan menetapkan pilihan yang dianggap positif dibanding pilihan lain yang dianggap negatif. Setelah individu menetapkan pilihan/keputusan, ia akan berupaya untuk mencari alasan untuk membenarkan pilihan/keputusan itu, alih-alih membuktikan ketepatannya. Dalam konteks penilaian kinerja subjektif, kriteria kinerja subjektif yang diperoleh digunakan sebagai dasar untuk membenarkan penilaian yang diberikan, bukan sebaliknya menggunakan kriteria itu untuk menilai ketepatan penilaian kinerja yang diberikan.

Hasil penelitian ini memiliki beberapa implikasi praktik yang penting untuk dipertimbangkan oleh perusahaan dalam mengembangkan sistem penilaian kinerja, antara lain: (1) penting bagi perusahaan untuk mempertimbangkan sifat individual tertentu terutama altruistik ketika menetapkan individu tertentu untuk menjadi penilai kinerja; (2) menggunakan beberapa penilai untuk melakukan aktivitas penilaian kinerja akan mengurangi kendala penilaian individual dan mengoptimalkan penilaian kinerja.

Beberapa keterbatasan yang mengemuka dalam penelitian ini memberi peluang bagi penelitian selanjutnya. Salah satu keterbatasan dalam penelitian ini adalah penggunaan kriteria kinerja subjektif yang diadaptasi dari Bellavance et al. (2013) kemungkinan kurang memadai dalam memberikan pengarahannya bagi partisipan. Kriteria kinerja ini kemungkinan dipersepsikan secara berbeda sehingga menimbulkan ambiguitas bagi penilai. Penelitian selanjutnya dapat dilakukan dengan merumuskan deskripsi mengenai kriteria

kinerja subjektif yang lebih lengkap untuk membantu penilaian kinerja subjektif.

Selain itu, tugas eksperimen cenderung menyederhanakan aktivitas penilaian sehingga kemungkinan menyebabkan penilai tidak mempertimbangkan dengan baik skor yang diberikan. Penelitian selanjutnya dapat diperdalam dengan meminta partisipan menunjukkan alasan penetapan skor kinerja yang diberikan, misalnya meminta skor/bobot yang diberikan oleh penilai untuk tiap kriteria yang diusulkan.

Ketiadaan insentif/hukuman terhadap ketidakakuratan kemungkinan juga menyebabkan para penilai individual tidak terdorong untuk memroses informasi dengan baik. Memberikan insentif/hukuman atas ketidakakuratan menggunakan contoh yang dijelaskan oleh Giebe dan Gürtler (2012) dalam kasus Arrows Electronics kemungkinan dapat menjadi salah satu usulan bagi penelitian mendatang. Misalnya, peneliti dapat menginformasikan kepada penilai bahwa hasil penilaian mereka akan dinilai kembali oleh pihak tertentu dan terdapat insentif/hukuman terhadap ketidakakuratan.

Keterbatasan lain dalam penelitian ini adalah penggunaan mahasiswa yang belum berpengalaman untuk mengerjakan tugas penilaian kinerja kemungkinan memengaruhi hasil. Hal ini disebabkan kondisi yang dikembangkan dalam eksperimen kemungkinan tidak terinternalisasi dengan baik pada partisipan. Penelitian selanjutnya dapat dilakukan dengan mengembangkan kondisi yang membuat partisipan dapat memahami dan memainkan perannya dengan baik, atau dapat juga dilakukan dengan menggunakan penilai sesungguhnya sebagai partisipan.

TERIMA KASIH

Kami berterima kasih kepada Lembaga Pengelola Dana Pendidikan (LPDP) Kementerian Keuangan Republik Indonesia yang telah mendanai penelitian ini.

DAFTAR PUSTAKA

- Bellavance, F., S. Landry, and E. Schiehl. 2013. Procedural Justice in Managerial Performance Evaluation: Effects of Subjectivity, Relationship Quality, and Voice Opportunity. *British Accounting Review*, 45, 149-166.
- Berger, J., C. Harbring, and D. Sliwka. 2010. *Performance Appraisals and the Impact of Forced Distribution: An Experimental Investigation*. Working Paper, IZA No. 5020: 1-44.
- Bernardin, H. J., D. K. Cooke, and P. Villanova. 2000. Conscientiousness and Agreeableness as Predictor Rating Leniency Bias. *Journal of Applied Psychology*, 85 (2), 232-234.
- Birnberg, J. G. 2009. The Case for Post-Modern Management Accounting: Thinking Outside the Box. *Journal of Management Accounting Research*, 21, 3-18.
- Bol, J. C. 2008. Subjectivity in Compensation Contracting. *Journal of Accounting Literature*, 27, 1-32.
- Bol, J. C. 2011. The Determinants and Performance Effects of Managers' Performance Evaluation Biases. *The Accounting Review*, 86 (5), 1549-1575.
- Bol, J. C. and S. D. Smith. 2011. Spillover Effects in Subjective Performance Evaluation: Bias and the Asymmetric Influence of Controllability. *The Accounting Review*, 86 (4), 1213-1230.
- Bosman, R., H. Hennig-Schmidt, and F. Van Winden. 2006. Exploring Decision Making in a Power-to-take Experiment. *Experiment Economic*, 9, 35-51.
- Breuer, K., P. Nieken, and D. Sliwka. 2013. Social Ties and Subjective Performance Evaluations: An Empirical Investigation. *Review Management Science*, 7, 141-157.
- Cappel, J. J. and J. C. Windsor. 2000. Ethical Decision Making: A Comparison of Comparison of Computer-supported and Face-to-face Group. *Journal of Business Ethics*, 28, 95-107.

- Charness, G., E. Karni, and D. Levin. 2007. Individual and Group Decision Making under Risk: An Experimental Study of Bayesian Updating and Violation of First-order Stochastic Dominance. *Journal of Risk Uncertainty*, 35, 129-148.
- Chen, T. Y. 2012. Approaches to Leniency Reduction in Multi-Criteria Decision Making with Interval-Valued Fuzzy Sets and an Experimental Analysis. *International Journal of Information Technology & Decision Making*, 11 (3), 579-605.
- Chen, C. X., M. G. Williamson, and F. H. Zhou. 2012. Reward System Design and Group Creativity: An Experimental Investigation. *The Accounting Review*, 87 (6), 1885-1911.
- Chen, J. V. and K. Linn. 2012. User Satisfaction with Group Decision Making Process and Outcome. *Journal of Computer Information Systems*, 52 (4), 30-40.
- Festinger, L. 1957. *A Theory of Cognitive Dissonance*. Evanston, IL: Row & Peterson.
- Galini, A. 2013. Endowment Effect in Negotiations: Group versus Individual Decision Making. *Theory Decision*, 75, 389-401.
- Gibbs, M., K. A. Merchant., W. A. Van der Stede, and M. E. Vargus. 2004. Determinants and Effects of Subjectivity in Incentives. *The Accounting Review*, 79 (2), 409-436.
- Gibson, L. B. and P. C. Earley. 2007. Collective Cognition in Action: Accumulation, Interaction, and Accommodation in the Development and Operation of Group Efficacy? *Academy of Management Review*, 32 (2), 438-458.
- Giebe, T. and O. Gürtler. 2012. Optimal Contract for Lenient Supervisors. *Journal of Economic Behavior and Organization*, 81, 403-420.
- Golman, R. and S. Bhatia. 2012. Performance Evaluation Inflation and Compression. *Accounting, Organizations and Society*, 37, 534-543.
- Grund, C. and J. Przemeczek. 2012. Subjective Performance Appraisal and Inequality Aversion. *Applied Economics*, 44, 2149-2155.
- Hair, J. F. Jr., W. C. Black., B. J. Babin, and R. E. Anderson. 2010. *Multivariate Data Analysis. A Global Perspective*. Upper Saddle River, NJ: Pearson Education.
- Huang, W. and D. Li. 2007. Opening up the Black Box in GSS Research: Explaining Group Decision Outcome with Group Process. *Computers in Human Behavior*, 23, 58-78.
- Hartono, J. 2013. *Pedoman Survei Kuisisioner: Mengembangkan Kuisisioner, Mengatasi Bias dan Meningkatkan Respons*. Yogyakarta: BPFE UGM.
- Kingstrom, P. O. and L. E. Mainstone. 1985. An Investigation of the Rater-ratee Acquaintance and Rater Bias. *Academy of Management Journal*, 85 (3), 641-653.
- Kocher, M. G. and M. Sutter. 2007. Individual versus Group Behavior and the Role of Decision Procedure in Gift-Exchange Experiment. *Empirica*, 34, 63-88.
- Liedtka, S. L., B. K. Church, and M. R. Ray. 2008. Performance Variability, Ambiguity Intolerance, and Balanced Scorecard-based Performance Assessments. *Behavioral Research in Accounting*, 20 (2), 73-88.
- Lin, C. W., C. S. Lin., P. C. Huang, and Y. L. Wang. 2014. How Group Efficacy Mediates the Relationship between Group Affect and Identification. *Journal of Business Research*, 67, 1388-1394.
- Lowe, D. J., P. M. Reckers, and S. M. Whitecotton. 2002. The Effects of Decision-Aid Use and Reliability on Jurors' Evaluation of Auditor Liability. *The Accounting Review*, 77 (1), 185-202.
- Maas, V. S. and R. Torres-González. 2011. Subjective Performance Evaluation and Gender Discrimination. *Journal of Business Ethics*, 101, 667-681.

- Martell, R. F. and M. R. Borg. 1993. A Comparison of the Behavioral Rating Accuracy of Groups and Individuals. *Journal of Applied Psychology*, 78, 43-50.
- Martell, R. F. and K. N. Leavitt. 2002. Reducing the Performance-Cue Bias in Work Behavior Ratings: Can Groups Help? *Journal of Applied Psychology*, 87 (6), 1032-1041.
- Moers, F. 2005. Discretion and Bias in Performance Evaluation: The Impact of Diversity and Subjectivity. *Accounting, Organizations and Society*, 30, 67-80.
- Nahartyo, E. 2012. *Desain dan Implementasi Riset Eksperimen, Edisi 1*. Yogyakarta: UPP STIM YKPN.
- Prendergast, C. and R. H. Topel. 1996. Favoritism in Organization. *Journal of Political Economy*, 104 (5), 958-978.
- Roch, S. G. 2006. Discussion and Consensus in Rater Groups: Implications for Behavioral and Rating Accuracy. *Human Performance*, 19 (2), 91-115.
- Rushton, J. P., R. D. Chrisjohn, and G. C. Fekken. 1981. The Altruistic Personality and the Self-report Altruism Scale. *Personality and Individual Differences*, 2, 293-302.
- Simmons, H. A. 1993. Altruism and Economics. *The American Economic Review*, 83 (2), 156-181.
- Simmons, R. G. 1991. Altruism and Sociology. *Sociological Quarterly*, 32, 1-22.
- Slaughter, J. E. and J. G. Greguras. 2008. Bias in Performance Ratings: Clarifying the Role of Positive versus Negative Escalation. *Human Performance*, 21, 414-426.
- Staw, B. M. 1975. Attribution of the "Causes" of Performance: A General Alternative Interpretation of Cross-sectional Research on Organizations. *Organizational Behavior and Human Performance*, 13, 414-432.
- Van der Stede, C. W., C. W. Chow, and T. W. Lin. 2006. Strategy, Choice of Performance Measures, and Performance. *Behavioral Research in Accounting*, 18, 185-205.
- Varma, A. and S. Pichler. 2007. Interpersonal Affect: Does It Really Bias Performance Appraisal. *Journal of Labour Research*, 28 (2), 207-222.
- Wherry, R. J. and C. J. Bartlett. 1982. The Control of Bias Ratings: A Theory of Rating. *Personnel Psychology*, 35, 521-552.
- Wright, E. F. and G. L. Wells. 1985. Does Group Discussion Attenuate the Dispositional Bias? *Journal of Applied Social Psychology*, 15, 531-546.