

PENGARUH PERSPEKTIF DAN JENIS UKURAN DALAM BALANCED SCORECARD TERHADAP EVALUASI KINERJA

Hariman Bone

hariman.unmul@gmail.com

Universitas Mulawarman

Mahfud Sholihin

mahfud@ugm.ac.id

Universitas Gadjah Mada

ABSTRACT

This study aims to investigate whether an evaluator has propensity to more weigh common financial measurement than unique financial measurement and common nonfinancial measurement when evaluating performance. We conduct two experiments where participants act as senior management to evaluate two divisions. The first experiment examines the existence of propensity of managers to more weigh common financial measurement than unique financial measurement. The ANOVA repeated measurement show that participants more weigh common financial measurement than unique financial measurement. The second experiment examines the existence of propensity of participants to weigh common financial measurement than common nonfinancial measurement. The result shows that participants more weigh common financial measurement than common nonfinancial measurement in their performance evaluation.

Key words: Balanced Scorecard, performance evaluation, performance measurement

ABSTRAK

Penelitian ini bertujuan menginvestigasi apakah evaluator mempunyai kecenderungan untuk memberi bobot lebih pada ukuran kinerja keuangan umum dibanding ukuran kinerja keuangan unik dan ukuran kinerja nonkeuangan umum ketika mengevaluasi kinerja. Kami melakukan dua eksperimen yang mana partisipan bertindak sebagai manajemen senior yang mengevaluasi dua divisi. Eksperimen pertama menguji apakah manajer lebih memperhatikan ukuran kinerja keuangan umum dibanding ukuran kinerja keuangan unik. Hasil pengujian menggunakan ANOVA menunjukkan bahwa partisipan memang lebih mempertimbangkan ukuran keuangan umum dibanding ukuran keuangan unik. Eksperimen kedua menguji apakah partisipan cenderung lebih memperhatikan ukuran keuangan umum dibanding ukuran nonkeuangan umum. Hasil penelitian menunjukkan bahwa partisipan lebih memperhatikan ukuran keuangan umum dibanding ukuran nonkeuangan umum dalam mengevaluasi kinerja.

Kata kunci: *Balanced Scorecard*, evaluasi kinerja, ukuran kinerja

PENDAHULUAN

Sejak diperkenalkan oleh Kaplan dan Norton pada awal tahun 1990an, *Balanced Scorecard* (BSC) telah berhasil menarik perhatian praktisi maupun peneliti. Terkait dengan praktik, Lawson *et al.*, (2008) melaporkan hasil survei internasionalnya bahwa sejak tahun 2004 hingga 2006 sekitar 50% perusahaan telah menggunakan BSC.

Anand *et al.* (2005) melakukan survei di India dan menemukan bahwa 45,28% responden telah menggunakan BSC sebagai alat manajemen kinerja mereka.

BSC juga sudah banyak diteliti. Hasil penelitian terdahulu menunjukkan bahwa BSC ditemukan memiliki pengaruh positif terhadap kinerja organisasi (Hoque dan James, 2000), unit bisnis (Davis dan Albright

2004), serta terhadap motivasi kerja individu (Burney dan Swanson, 2010).

Namun demikian beberapa peneliti juga menemukan adanya dampak negatif penerapan BSC di dalam perusahaan (Malina dan Selto, 2001; Ittner *et al.*, 2003a; Ittner *et al.*, 2003b). Beberapa hal yang dianggap sebagai penghambat keefektifan BSC antara lain adalah adanya komunikasi satu arah (Malina dan Selto, 2001), subjektivitas dalam evaluasi (Ittner *et al.*, 2003b), dan keterbatasan kognitif (Lipe dan Salterio, 2000).

Penelitian ini menindaklanjuti penelitian yang dilakukan oleh Lipe dan Salterio (2000) serta Cardinaels dan Veen-Dirks (2010) untuk mengidentifikasi kemungkinan adanya bias dalam evaluasi kinerja. Lipe dan Salterio (2000) menemukan adanya “bias umum” dalam evaluasi kinerja menggunakan BSC. Artinya, manajer perusahaan cenderung menggunakan ukuran-ukuran umum dan mengabaikan ukuran-ukuran unik dalam mengevaluasi kinerja. Hanya saja, mereka tidak meneliti apakah manajer juga cenderung menggunakan ukuran-ukuran kinerja umum dibanding ukuran-ukuran kinerja unik dalam perspektif yang sama dalam BSC.

Cardinaels dan Veen-Dirks (2010) meneliti apakah manajer lebih cenderung mempertimbangkan perspektif finansial dibanding perspektif nonfinansial. Mereka menemukan bahwa manajer lebih cenderung untuk mempertimbangkan perspektif finansial dibandingkan nonfinansial. Hanya saja, mereka tidak membandingkan apakah para manajer cenderung menggunakan ukuran kinerja umum dan mengabaikan ukuran kinerja unik.

Dari hasil penelitian Lipe dan Salterio (2000) serta Cardinaels dan Veen-Dirks (2010) ada dua pertanyaan yang belum terjawab: (1) apakah manajer cenderung menggunakan ukuran finansial umum (*common financial*) dibanding dengan ukuran finansial unik (*unique financial*)?; dan (2) apakah manajer cenderung menggunakan ukuran finansial umum (*common financial*) diban-

ding ukuran nonfinansial umum (*common nonfinancial*)? Penelitian ini bertujuan untuk menjawab dua pertanyaan tersebut.

Penelitian ini menggunakan dua kali eksperimen. Eksperimen pertama membandingkan penggunaan finansial umum dan finansial unik dalam evaluasi kinerja. Sedangkan eksperimen kedua membandingkan penggunaan finansial umum dan non-finansial umum dalam evaluasi kinerja. Hasil temuan kami menyatakan bahwa ukuran finansial umum lebih diperhatikan daripada ukuran finansial unik dan non-finansial umum.

Menginvestigasi isu ini penting untuk dilakukan karena memiliki implikasi praktis maupun konseptual. Sebagaimana dikatakan oleh Kaplan dan Norton (1992, 1993, 1996a, 1996b, dan 2001), perusahaan dalam membangun ukuran-ukuran BSC tersebut seharusnya menghubungkan ukuran-ukuran tersebut dengan strategi dan tujuan perusahaan. Dengan kata lain, strategi dan tujuan perusahaan seharusnya termanifestasi dalam ukuran-ukuran BSC. Dengan demikian, hanya berfokus dengan menggunakan ukuran-ukuran tertentu dan mengabaikan ukuran-ukuran yang lain dalam BSC ketika melakukan evaluasi kinerja akan menghambat tercapainya tujuan organisasi karena individu yang dievaluasi merasa bahwa kinerjanya tidak dievaluasi secara komprehensif. Hal ini akan mengarahkan bawahan untuk mengalokasikan sumber daya secara tidak tepat sehingga akan mengarahkan pada ketidakefektifan alokasi sumber daya. Lebih lanjut, proses evaluasi yang tidak komprehensif akan menjadikan bawahan tidak memperhatikan aktivitas-aktivitas yang tidak dimasukkan dalam evaluasi kinerja oleh atasan (Holmstrom dan Milgrom, 1991) dan pada akhirnya akan menghambat tercapainya tujuan suatu organisasi.

TINJAUAN TEORETIS

Balanced Scorecard, Implementasi, dan Konsekuensi

Balanced scorecard pertama kali diperkenalkan oleh Kaplan dan Norton (1992) sebagai seperangkat pengukuran kinerja yang memberi manajemen puncak pandangan yang cepat dan komprehensif tentang perusahaan. BSC memasukkan ukuran-ukuran finansial yang memberi hasil yang telah dilakukan dan melengkapinya dengan ukuran-ukuran operasional diantaranya kepuasan pelanggan, proses internal dan inovasi dan perkembangan.

Perkembangan BSC dapat ditelusuri melalui tulisan-tulisan Kaplan dan Norton (1992, 1993, 1996a, 1996b, dan 2001) yang dimulai dari BSC sebagai pengukuran kinerja sampai pada perkembangan BSC sebagai sistem manajemen strategi. Setelah itu, Kaplan (2009) memberi gambaran mengenai perkembangan BSC mulai dari akar pemikiran munculnya BSC sampai menjadi sistem manajemen strategi. Ia merinci bahwa BSC awalnya dibuat sebagai pengukuran kinerja yang komprehensif yang memadukan ukuran finansial dan nonfinansial. Selanjutnya BSC dibuat memiliki hubungan *cause and effect* sehingga melahirkan peta strategi dan pada akhirnya BSC menjadi sistem manajemen strategi yang kemudian dapat digunakan untuk mengevaluasi tidak hanya ukuran yang digunakan namun strategi yang akan digunakan.

BSC memiliki karakteristik-karakteristik tertentu yang harus ada agar dapat dikatakan BSC. Pertama, mengkombinasi ukuran finansial dan nonfinansial (Kaplan dan Norton, 1992). Kedua, menyelaraskan ukuran-ukuran dengan strategi perusahaan (Kaplan dan Norton; 1996a dan 1996b). Ketiga, mempunyai hubungan *cause and effect* antara ukuran-ukuran dalam perspektif-perspektif BSC (Kaplan dan Norton, 2001; Kaufmann dan Becker, 2005).

BSC telah diterapkan oleh banyak perusahaan di dunia. Lawson *et al.* (2008) melaporkan survei internasionalnya bahwa sejak tahun 2004 hingga 2006 sekitar 50% perusahaan telah menggunakan BSC. Anand *et al.* (2005) melakukan survei di India dan menemukan bahwa 45,28%

responden telah menggunakan BSC sebagai alat manajemen kinerja mereka.

Hasil survei Downing (2000) menunjukkan bahwa alasan perusahaan menerapkan BSC adalah:

1. Menyelaraskan organisasi dengan strategi (66%).
2. Mensinergikan organisasi (61%).
3. Membangun sistem manajemen strategi (57%).
4. Menghubungkan strategi dengan perencanaan dan penganggaran (54%).
5. Menetapkan target strategi (51%).
6. Memprioritaskan inisiatif strategi (50%).
7. Menyelaraskan individu dengan strategi (47%).

Malmi (2001) memberi alasan penggunaan BSC di Finlandia yaitu:

1. Karena kemampuan BSC menerjemahkan strategi ke dalam tindakan.
2. Adanya program kualitas dan *reward* kualitas sehingga mendorong perusahaan menerapkan BSC. Untuk mendukung agenda perubahan.
3. *Fads* dan *fashion*.
4. Ketidak-akuratan penganggaran tradisional.

Penerapan BSC juga sering diteliti berkaitan dengan kinerja perusahaan. Davis dan Albright (2004) menemukan bahwa divisi yang menggunakan BSC memiliki kinerja yang meningkat dua tahun setelah penerapan BSC. Braam dan Nijssen (2004) menyatakan bahwa penggunaan BSC tidak secara otomatis akan meningkatkan kinerja perusahaan tetapi juga tergantung pada cara menggunakan, yang mana penggunaan BSC yang selaras dengan strategi akan berpengaruh positif terhadap kinerja dan yang tidak selaras mungkin justru akan menurunkannya. Burney dan Swanson (2010) menemukan bahwa penekanan manajer adalah pada ukuran-ukuran perspektif pelanggan dan perspektif pembelajaran dan pertumbuhan pada saat mengambil keputusan serta keterkaitan ukuran dengan strategi mempengaruhi kepuasan kerja manajer.

Kaufmann dan Becker (2005) mengidentifikasi hambatan-hambatan baik pada tahap inisiasi dan pendesainan BSC maupun pada tahap implementasi BSC dengan menggunakan pendekatan kualitatif. Penelitian mereka dilakukan pada perusahaan multinasional di Brazil. Hambatan-hambatan yang teridentifikasi pada tahap inisiatif dan pendesainan BSC diantaranya: kurangnya komitmen anggota organisasi, dukungan negatif dari konsultan, kurangnya dukungan dari manajemen puncak, kurang adanya kekomprehensifan strategi, kesulitan dalam mengidentifikasi sasaran strategi dan adanya hubungan *cause-and-effect*, ketidakcukupan keselarasan antara tingkat korporat sampai pada unit bisnis, dan kurangnya kelengkapan yaitu kurangnya pendokumentasian setelah tahap penyelesaian proyek BSC. Sedangkan hambatan-hambatan pada tahap implementasi antara lain: lingkungan Brazil yang tidak pasti, hilangnya pelaporan dan *review* BSC, ketidakcukupan komunikasi, perspektif perbedaan waktu antara periode perencanaan dan periode implementasi, dan ketidaklengkapan data pasar dan kinerja.

Isu Kognitif *Balanced Scorecard*

Isu keterbatasan kognitif dalam evaluasi kinerja menggunakan BSC yang banyak mendapat perhatian saat ini adalah penelitian yang dilakukan oleh Lipe dan Salterio (2000). Lipe dan Salterio (2000) secara khusus telah meneliti tentang perbandingan antara ukuran kinerja yang umum (*common measures*) dan ukuran kinerja yang unik (*unique measures*). Penelitian tersebut dilakukan menggunakan metode eksperimen dengan mahasiswa sebagai subjek (desain eksperimen 2 X 2 *between subject* dan 2 level *within subject*). Hasil penelitian menunjukkan bahwa partisipan cenderung lebih mempertimbangkan ukuran kinerja yang umum dibandingkan ukuran kinerja yang unik dalam pengambilan keputusan.

Lipe dan Salterio (2002) juga membandingkan evaluasi kinerja manajer yang

menggunakan format BSC dan yang tidak menggunakan format BSC. Hasil penelitian mereka menyatakan bahwa ketika perbedaan kinerja dua divisi terletak pada satu perspektif saja (dalam format BSC), penilaian yang diberikan lebih moderat di bandingkan dengan penilaian ketika dalam format nonBSC; namun ketika perbedaan kinerja dua divisi disebar pada keempat perspektif maka BSC tidak mempengaruhi evaluasi terhadap manajer. Dengan kata lain tidak ada perbedaan signifikan antara format BSC dan nonBSC pada eksperimen mereka.

Penelitian selanjutnya lebih berfokus pada bagaimana mengurangi *common measure bias* yang telah teridentifikasi sebelumnya oleh Lipe dan Salterio (2000). Beberapa solusi yang diteliti diantaranya: meminta manajemen puncak untuk menilai kualitas strategi dalam menentukan kinerja suatu divisi (Wong-on-wing *et al.*, 2007), pemberian pengetahuan BSC (Dilla dan Steinbart, 2005), pemberian informasi mengenai strategi unit bisnis (Banker *et al.*, 2004), peta strategi (Banker *et al.*, 2009), *General Problem solving* (Hibbets *et al.*, 2006), pendekatan "*disaggregated/mechanically aggregated*" (Robert *et al.*, 2004), dan toleransi terhadap ambiguitas (Liedtka *et al.*, 2008).

Itner *et al.* (2003a) menguji tentang bagaimana tipe pengukuran kinerja yang berbeda dalam BSC digunakan dalam pertimbangan bonus pada perusahaan jasa keuangan. Hipotesis yang diajukan oleh penulis dikembangkan dari dua teori yaitu teori ekonomi dan teori psikologi. Hasil penelitian menemukan bahwa dalam pertimbangan evaluasi kinerja menggunakan BSC, seorang atasan lebih menggunakan informasi finansial dibandingkan dengan informasi non-finansial. Begitupun dalam penentuan bonus, kinerja keuangan menjadi penentu utama bonus. Sebagai tambahan, subjektivitas dalam perencanaan BSC memungkinkan atasan tersebut untuk memasukkan faktor-faktor selain ukuran-ukuran BSC dalam evaluasi kinerja, mengubah kriteria evaluasi kinerja dari kuartal

ke kuartal, mengabaikan ukuran-ukuran kinerja yang memicu kinerja keuangan masa depan dan juga mempertimbangkan ukuran-ukuran yang tidak berkaitan dengan hasil yang kita inginkan.

Penelitian lain tentang isu kognitif dalam evaluasi kinerja yang terbaru dilakukan oleh Cardinaels dan Veen-Dirks (2010). Mereka membandingkan dua *framing* pengukuran kinerja, satu *frame* dibuat dalam bentuk BSC dan *frame* yang lain diformat tidak dalam bentuk BSC. Mereka ingin mengetahui apakah ketika pengukuran kinerja diformat dalam bentuk BSC, manajer akan cenderung untuk lebih menitikberatkan ukuran finansial daripada ukuran non-finansial dibandingkan ketika pengukuran tidak diformat dalam bentuk BSC. Hasilnya, partisipan cenderung menitikberatkan ukuran finansial dibandingkan ukuran non-finansial dalam format BSC.

Pengembangan Hipotesis

Kaplan dan Norton (1992) menyatakan bahwa manajer seharusnya tidak memilih hanya salah satu antara ukuran kinerja keuangan dan ukuran operasional tetapi manajer seharusnya menggunakan kedua ukuran dalam pengukuran kinerja. BSC terdiri atas ukuran yang multiperspektif (finansial, pelanggan, proses bisnis internal, dan pertumbuhan dan pembelajaran) yang menggambarkan visi dan strategi unit bisnis tertentu (Kaplan dan Norton, 1996a). Konsekuensinya, semua ukuran-ukuran tersebut harus dipertimbangkan dalam pengukuran kinerja.

Namun demikian, individu dibatasi oleh ketersediaan informasi dalam memori mereka (Tversky dan Kahneman, 1973 dan 1974). Informasi keuangan telah dikenal dan digunakan sejak lama sehingga ketika individu diminta untuk mengevaluasi kinerja menggunakan informasi keuangan dan nonkeuangan maka mereka akan cenderung untuk menggunakan informasi keuangan daripada nonkeuangan. Kecenderungan untuk menggunakan informasi keuangan dibandingkan nonkeuangan juga

dipengaruhi oleh *priming effect* (Baron dan Byne, 1997 dalam Kliger dan Kudryavtsev, 2010). *Priming effect* mengarahkan para individu untuk berfikir bahwa para pemegang saham adalah tujuan utama sehingga mengarahkan mereka untuk lebih mempertimbangkan informasi keuangan di bandingkan informasi nonkeuangan.

Ittner *et al.* (2003a) menemukan bahwa penggunaan subjektivitas dalam mempertimbangkan ukuran dalam perencanaan bonus dengan BSC memungkinkan *supervisor* mengabaikan banyak ukuran kinerja, dan kinerja keuangan menjadi penentu utama bonus. Cardinaels dan Veen-Dirks (2010) menyatakan bahwa ukuran keuangan (seperti pertumbuhan penjualan, margin penjualan) memberi orientasi yang umum sebagai *outcome* keuangan sedangkan ukuran nonfinansial mengandung kombinasi antara ukuranyang berorientasi *outcome* dan ukuran yang dipandang sebagai pemicu *outcome*. Oleh karena itu ukuran finansial lebih terlihat sebagai *outcome* yang jelas dibandingkan nonfinansial.

DeBusk *et al.* (2003) menyatakan bahwa ada kecenderungan untuk lebih memilih ukuran keuangan karena ukuran tersebut lebih familiar. Cardinaels dan Veen-Dirks (2010) menemukan bahwa ketika pengukuran kinerja diformat dalam bentuk BSC, manajer akan cenderung untuk lebih menitikberatkan ukuran finansial daripada ukuran nonfinansial. Beberapa literatur-literatur sebelumnya mengarahkan pada simpulan sementara bahwa individu dalam mengevaluasi kinerja akan lebih memperhatikan ukuran finansial dibandingkan dengan ukuran nonfinansial.

Bentuk keterbatasan evaluasi kinerja yang lain ditunjukkan oleh Lipe dan Salterio (2000). Penelitian mereka banyak didasarkan dari penelitian klasik yang dilakukan oleh Slovic dan MacPhillamy (1974). Mereka meminta partisipan (mahasiswa) untuk menilai mana di antara dua mahasiswa yang memiliki IPK yang lebih tinggi pada tahun pertama. Para partisipan

mendasarkan pertimbangan mereka pada informasi angka terkait kemampuan bahasa Inggris, bakat kuantitatif, dan kebutuhan akan pencapaian. Partisipan melihat informasi yang umum antara kedua mahasiswa dan informasi yang unik. Mereka menemukan bahwa partisipan tersebut lebih mempertimbangkan ukuran-ukuran yang umum dibandingkan ukuran-ukuran yang unik.

Dalam konteks BSC, Lipe dan Salterio (2000) secara khusus telah meneliti tentang perbandingan antara ukuran kinerja yang umum (*common measurement*) dan ukuran kinerja yang unik (*unique measurement*). Penelitian tersebut menggunakan eksperimen dengan mahasiswa sebagai subjek (desain eksperimen 2 X 2 *between subject* dan 2 level *within subject*). Hasil penelitian menunjukkan bahwa partisipan tersebut cenderung lebih mempertimbangkan ukuran yang umum dibandingkan ukuran yang unik dalam pengambilan keputusan.

Penelitian Slovic dan MacPhillamy (1974) dan Lipe dan Salterio (2000) memperoleh temuan yang sama yaitu individu cenderung untuk lebih memperhatikan ukuran-ukuran yang umum dibandingkan ukuran-ukuran yang unik. Berdasarkan literatur-literatur sebelumnya penelitian ini menyimpulkan bahwa ukuran-ukuran finansial dan ukuran-ukuran umum akan lebih diperhatikan daripada ukuran-ukuran nonfinansial dan ukuran-ukuran unik. Dengan memperhatikan literatur-literatur di atas maka dirumuskan hipotesis sebagai berikut:

- H₁: Ketika melakukan evaluasi kinerja menggunakan BSC, manajer akan lebih memperhatikan ukuran finansial yang umum (*common financial*) dibandingkan dengan ukuran finansial yang unik (*unique financial*)
- H₂: Ketika melakukan evaluasi kinerja menggunakan BSC, manajer akan lebih memperhatikan ukuran finansial yang umum (*common financial*) dibandingkan dengan ukuran nonfinansial yang umum (*common nonfinancial*).

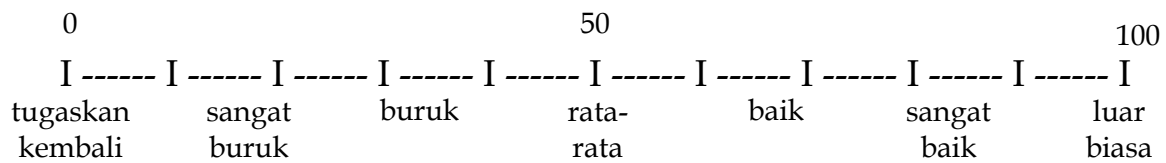
METODE PENELITIAN

Prosedur dan Tugas Partisipan

Untuk menguji hipotesis, penelitian ini menggunakan metode eksperimen. Partisipan diminta untuk membaca kasus yang mana dalam kasus tersebut mereka diminta untuk berperan sebagai direktur pada perusahaan yang bernama Harapan Garmen, yaitu perusahaan yang berspesialisasi pada pakaian wanita. Materi kasus berfokus pada dua divisi Harapan Garmen yang terbesar. Kasus tersebut menceritakan bahwa direktur keuangan Harapan Garmen menghadiri seminar *balanced scorecard* dan memutuskan untuk menerapkannya. Kemudian direktur keuangan tersebut menjelaskan konsep BSC dan mendaftar empat kategori pengukuran. Selanjutnya para partisipan mempelajari bahwa perusahaan akan menerapkan BSC. Selanjutnya kasus tersebut juga memberi gambaran tentang pernyataan misi Harapan Garmen dan memperkenalkan manajer dari dua unit bisnis (divisi). Harapan Garmen juga memiliki dua divisi yaitu divisi busana remaja yang berspesialisasi pada pakaian remaja dan divisi Busana Kerja yang berspesialisasi pada pakaian seragam kerja wanita. Kedua divisi tersebut memiliki tujuan dan strategi yang berbeda. Oleh karena itu, mereka memiliki ukuran-ukuran yang umum dan memiliki ukuran-ukuran yang unik yang hanya dimiliki oleh divisi tertentu untuk mencapai tujuan dan strategi divisi tersebut. Instrumen ini diadopsi dari Lipe dan Salterio (2000).

Tabel 1 menyajikan salah satu BSC dari divisi tersebut. Partisipan dalam penelitian ini bertindak sebagai direktur Harapan Garmen dan membuat keputusan tentang kinerja dari bawahannya (manajer unit/divisi).

Setelah membaca latar belakang kasus, para partisipan diminta untuk menilai kinerja dari kedua divisi dengan memberi tanda silang (X) pada skala berikut:



1. Luar biasa: penilaian kinerja yang jauh di atas perkiraan;
2. Sangat baik: penilaian kinerja yang secara signifikan di atas perkiraan;
3. Baik: penilaian kinerja agak di atas perkiraan;
4. Rata-rata: realisasi kinerja sama dengan perkiraan;
5. Buruk: realisasi kinerja agak di bawah perkiraan, membutuhkan sedikit perbaikan;
6. Sangat buruk: realisasi kinerja secara signifikan jauh di bawah harapan, membutuhkan perbaikan yang signifikan;
7. Tugaskan kembali: perbaikan yang cukup sudah tidak mungkin lagi.

Tugas terakhir partisipan adalah menyelesaikan pertanyaan simpulan umum. Pertanyaan simpulan umum tersebut merupakan pertanyaan terkait kasus yang telah dikerjakan oleh para partisipan. Pertanyaan tersebut meliputi kesulitan tugas dan kemampuan untuk menggambarkan dunia nyata dan keterpahaman.

Partisipan

Partisipan dalam eksperimen ini melibatkan 100 mahasiswa S1 Fakultas Ekonomi Universitas Muhamadiyah Yogyakarta dan 50 mahasiswa Fakultas Ekonomi Universitas Teknologi Yogyakarta yang telah menempuh mata kuliah akuntansi manajemen. Pemilihan subjek tersebut dikarenakan beberapa alasan: 1) Mahasiswa pada tahap tersebut telah banyak belajar tentang BSC dan evaluasi kinerja, hal tersebut menjadi pertimbangan telah mampu melakukan penilaian kinerja; 2) Penggunaan manajer sesungguhnya dari perusahaan yang telah

menerapkan BSC kemungkinan akan bias karena masing-masing manajer mempunyai BSC pada perusahaannya dan kemungkinan manajer mempertimbangkan pengukuran berdasarkan pengalaman mereka pada perusahaan masing-masing dan tidak berdasarkan pada desain kasus yang telah dibangun (Lipe dan Salterio, 2000); 3) Penggunaan partisipan yang telah di-*training* terlebih dahulu juga akan memiliki bias, karena kemungkinan hasil eksperimen yang diperoleh bukan karena manipulasi, akan tetapi karena dampak pembelajaran sebelumnya (Lipe dan Salterio 2000); 4) Clinton (1999) dalam Cheng *et al.* (2003) menyatakan bahwa penelitian-penelitian sebelumnya dalam literatur akuntansi dan manajemen terkait penelitian pertimbangan dan pembuatan keputusan, mahasiswa dibenarkan sebagai pengganti manajer.

Desain Eksperimen

Penelitian ini menggunakan dua kali eksperimen dengan desain 2X2 *between subject*, dilengkapi dengan dua level *within subject* sehingga keseluruhan desain penelitian digambarkan dengan faktor 2X2X2. Eksperimen pertama didesain untuk hipotesis pertama yang membandingkan ukuran-ukuran finansial umum dan ukuran-ukuran finansial unik. Faktor independen yang pertama adalah pola kinerja tertentu dari kedua divisi berdasarkan ukuran finansial umum, dimana Divisi Busana Remaja dapat berkinerja lebih baik pada finansial umum dibandingkan Divisi Busana Kerja (FC_BR) atau Divisi Busana Kerja dapat berkinerja lebih baik pada finansial umum dibandingkan Divisi Busana Remaja (FC_BK).

Tabel 1
Pengukuran Kinerja Divisi Busana Remaja

No.	Pengukuran	Target	Realisasi	% Peningkatan dari Target
Keuangan				
1	<i>Return on sales</i> (laba bersih divisi/ penjualan divisi)	24%	26,5%	8,33
2	Penjualan toko baru (% penjualan toko baru yang dibuka tahun 2012)	30,2%	32,72%	8,34
3	Pertumbuhan penjualan (penjualan 2012- penjualan 2011)/ penjualan 2011)	35%	39,25%	12,14
4	Pangsa pasar relatif terhadap ukuran toko (penjualan/total luas toko pakaian remaja yang ada dalam radius 10 km dari toko)	Rp80 juta	Rp83,5 juta	4,38
Pelanggan				
1	<i>Rating</i> program pembelian misterius (skala 100 point untuk menilai tampilan toko, sikap pegawai, efisiensi, dan efektivitas sistem, <i>chek out</i> dan lain-lain)	85	92	8,24
2	Penjualan kembali (% penjualan pada konsumen yang sama)	30%	34%	13,33
3	% retur dari konsumen dibandingkan dengan penjualan	12%	12,4%	3,33
4	<i>Rating</i> kepuasan konsumen (hasil survei konsumen)	92%	95%	3,26
Proses bisnis internal				
1	% retur pada pembeli (karena masalah kualitas)	6%	7%	16,67
2	Rata-rata jumlah nama merk utama	32	36	12,5
3	Rata-rata diskon yang diberikan pada pembeli	16%	18,5%	15,63
4	% penjualan dari baru yang di tentukan oleh <i>head buyer</i> (<i>market leader</i>)	25%	29,55%	18,2
Pembelajaran dan Pertumbuhan				
1	Rata-rata masa kerja tenaga penjualan (dalam tahun)	1.4	1.62	15,71
2	Jumlah jam pelatihan per karyawan	15	17	13,33
3	% Fungsi utama dari komputerisasi toko yang diidentifikasi oleh manajemen puncak	85%	92%	8,24
4	Jumlah saran dari tiap karyawan	3,3	3,5	6,06

Faktor independen yang kedua adalah pola kinerja tertentu dari kedua divisi berdasarkan ukuran finansial unik, dimana Divisi Busana Remaja didesain berkinerja lebih baik pada finansial unik dibandingkan Busana Kerja (FU_BR) atau Divisi Busana Kerja didesain berkinerja lebih baik pada finansial unik dibandingkan Busana Remaja (FU_BK). Sedangkan dua level terakhir adalah *within subject* yaitu Divisi Busana Remaja dan Divisi Busana Kerja.

Desain eksperimen kedua sama dengan eksperimen pertama. Eksperimen ini dilakukan berkaitan hipotesis kedua yang membandingkan antara ukuran finansial umum dan ukuran nonfinansial umum. Faktor independen yang pertama adalah pola kinerja berdasarkan ukuran finansial umum, dimana Divisi Busana Remaja didesain berkinerja lebih baik daripada Divisi Busana kerja pada ukuran finansial umum (FC_BR) atau Divisi Busana Kerja didesain berkinerja lebih baik pada ukuran finansial umum dibandingkan Divisi Busana Remaja (FC-BK). Faktor independen kedua adalah pola kinerja tertentu dari kedua divisi berdasarkan ukuran nonfinansial umum, dimana Divisi Busana Remaja didesain berkinerja lebih baik pada nonfinansial umum dibandingkan Divisi Busana Kerja (NFC_BR) atau Divisi Busana Kerja didesain berkinerja

lebih baik pada nonfinansial umum dibandingkan Divisi Busana Remaja (NFC_BK). Sedangkan dua level terakhir adalah Divisi Busana Remaja dan Divisi Busana Kerja. Tabel 2 dan 3 menunjukkan desain untuk kedua eksperimen.

Sebagaimana instrumen yang dikembangkan oleh Lipe dan Salterio (2000), penelitian ini juga menggunakan 16 ukuran BSC untuk kedua divisi Harapan Garmen. Pengukuran tersebut memiliki empat ukuran kinerja pada setiap kategori BSC. Setiap kategori terdiri atas dua ukuran yang umum dan dua ukuran yang unik. Untuk setiap pengukuran tersebut, setiap divisi berkinerja lebih baik dari yang ditargetkan, akan tetapi beragam sebagai-mana desain yang dijelaskan sebelumnya.

Variabel Terikat

Penilaian kinerja untuk kedua divisi yang diberikan oleh partisipan merupakan variabel terikat dalam penelitian ini. Para partisipan diminta untuk menilai kinerja baik divisi busana remaja maupun divisi busana kerja dengan menentukan angka kinerja dari skala 0 sampai dengan 100, dimana 0 adalah angka terendah (tugaskan kembali) sedangkan angka 100 adalah kinerja tertinggi (luar biasa).

Tabel 2
Matriks Desain Eksperimen 1

		FU_BR	FU_BK
FC_BR	Divisi Busana Remaja Divisi Busana Kerja	I	II
FC_BK	Divisi Busana Remaja Divisi Busana Kerja	III	IV

Tabel 3
Matriks Desain Eksperimen 2

		NFC_BR	NFC_BK
FC_BR	Divisi Busana Remaja Divisi Busana Kerja	I	II
FC_BK	Divisi Busana Remaja	III	IV

ANALISIS DAN PEMBAHASAN Cek Manipulasi

Terdapat empat pertanyaan yang digunakan untuk cek manipulasi. Empat pertanyaan tersebut diukur dengan skala dari 5 (sangat tidak setuju) sampai 5 (sangat setuju). Dikatakan lolos cek manipulasi jika jawaban partisipan terhadap empat pertanyaan tersebut berbeda signifikan dengan titik 0 (parameter jawaban tidak setuju). Dari 74 partisipan pada desain eksperimen pertama, 28 partisipan tidak lolos cek manipulasi. Dengan demikian, ada 46 respon yang digunakan pengolahan data selanjutnya. Hasil uji *one sample t test* menunjukkan bahwa jawaban partisipan atas keempat pertanyaan manipulasi (manfaat pengkategorian, target pasar yang sama, ukuran kinerja yang berbeda, ketepatan bagi kedua divisi menggunakan ukuran yang berbeda) berbeda signifikan ($p < 0,00$). Hal ini menunjukkan bahwa *treatment* dalam penelitian berhasil.

Pada desain eksperimen kedua, jumlah partisipan yang dilibatkan sebanyak 76 orang, 33 diantaranya tidak lolos cek manipulasi sehingga jumlah yang diolah dalam penelitian ini sebanyak 43 partisipan.

Keempat pertanyaan manipulasi cek berbeda signifikan dari titik 0 ($p < 0,00$).

Hasil Eksperimen Pertama

Uji ANOVA *repeated measured* untuk *between subject* menunjukkan bahwa pengaruh variabel finansial umum (FC) terhadap penilaian kinerja tidak signifikan pada tingkat signifikansi 0,05 dengan nilai $F = 0,057$ dan nilai $p = 0,813$. Pengaruh variabel finansial unik (FU) terhadap penilaian kinerja juga tidak signifikan pada tingkat signifikansi 0,05 dengan nilai $F = 0,297$ dan nilai $p = 0,589$ (disajikan pada tabel 4).

Hipotesis 1 diuji dengan memperhatikan interaksi antara *within* dan *between subject*. Jika interaksi antara divisi dan finansial umum signifikan pada level 0,05 sedangkan interaksi antara divisi dan finansial unik tidak signifikan maka hipotesis 1 didukung. Uji ANOVA *within subject* menunjukkan bahwa interaksi antara divisi dan finansial umum signifikan pada level signifikansi 0,05 dengan nilai $F = 12,843$ dan nilai $P = 0,001$. Sedangkan interaksi antara divisi dan finansial unik tidak signifikan pada level signifikansi 0,05 dengan nilai $F = 1,210$ dan nilai $p = 0,278$. Hal ini menunjukkan bahwa hipotesis 1 didukung. Tabel 5 mengindikasikan bahwa ketika finansial umum lebih baik pada BR, manajer puncak mengevaluasi divisi BR 7,6363 lebih tinggi daripada divisi BK.

Tabel 4
Hasil Eksperimen untuk Evaluasi Kinerja Manajer
pada Eksperimen 1

Variabel	df	SS	MS	F	P
<i>Between subject:</i>					
FC	1	5,400	5,400	0,057	0,813
FU	1	28,185	28,185	0,297	0,589
FC X FU	1	7,361	7,361	0,078	0,782
Error	42	3984,151	94,861		
<i>Within Subject:</i>					
DIVISI	1	14,120	14,120	0,171	0,681
DIVISI X FC	1	1061,650	1061,650	12,843	0,001
DIVISI X FU	1	99,997	99,997	1,210	0,278
DIVISI X FC X FU	1	47,232	47,232	0,571	0,454
Error	42	3471,901	82,664		

Sama halnya, ketika finansial umum lebih baik pada BK, divisi BK dievaluasi 6,1667 lebih tinggi daripada divisi BR. Sedangkan ketika finansial unik lebih baik pada divisi BR (BK), tidak ada perbedaan yang signifikan pada evaluasi manajer.

Desain eksperimen kedua digunakan

untuk menjawab hipotesis kedua yang menyatakan bahwa finansial umum lebih dipertimbangkan daripada nonfinansial umum. Hasil uji ANOVA *repeated measures* untuk *between subject* menunjukkan bahwa pengaruh variabel finansial umum terhadap penilaian kinerja tidak signifikan pada level

Tabel 5
Rata-Rata (Deviasi Standar) Evaluasi Kinerja Kedua Divisi

Ukuran		Ketika BR berkinerja lebih baik	Ketika BK berkinerja lebih baik
FC	BR	73,863 (8,4419)	66,333 (10,9372)
	BK	66,227 (9,7976)	72,500 (7,7234)
	Selisih rata-rata	7,6363	-6,1667
FU	BR	72,000 (7,3824)	68,200 (12,3220)
	BK	68,714 (11,1586)	70,160 (7,4312)
	Selisih rata-rata	3,286	-1,96

Tabel 6
Hasil Eksperimen untuk Evaluasi Kinerja Manajer pada Eksperimen 2

Variabel	df	SS	MS	F	P
<i>Between subject:</i>					
FC	1	56,046	56,046	0,819	0,371
NFC	1	163,442	163,442	2,387	0,130
FC X NFC	1	26,862	26,862	0,392	0,535
Error	39	2669,850	68,458		
<i>Within Subject:</i>					
DIVISI	1	5,936	5,936	0,108	0,744
DIVISI X FC	1	1216,638	1216,638	22,125	0,000
DIVISI X NFC	1	263,152	263,152	4,786	0,035
DIVISI X FC X NFC	1	0,679	0,679	0,012	0,912
Error	39	2144,588	54,989		

Tabel 7
Rata-Rata (Deviasi Standar) Evaluasi Kinerja Kedua Divisi

Ukuran		Ketika BR berkinerja lebih baik	Ketika BK berkinerja lebih baik
FC	Busana Remaja	79,000 (6,5212)	69,478 (8,4416)
	Busana Kerja	70,900 (9,7273)	76,956 (7,0290)
	Selisih rata-rata	8,7	-7,478
NFC	Busana Remaja	77,550 (9,4115)	70,739 (7,2814)
	Busana Kerja	73,500 (8,5992)	74,695 (9,1871)
	Selisih rata-rata	4,05	-3,956

signifikansi 0,05 dengan nilai $F=0,819$ dan nilai $p=0,371$. Pengaruh variabel non-finansial umum terhadap penilaian kinerja juga tidak signifikan pada level 0,05 dengan nilai $F=2,387$ dan nilai $p=0,130$.

Hipotesis 2 diuji dengan memperhatikan interaksi antara *between* dan *within subject*, dalam hal ini adalah interaksi antara finansial umum dan divisi serta interaksi antara nonfinansial umum dengan divisi. Jika interaksi divisi dan finansial umum signifikan pada level 0,05 sedangkan interaksi antara divisi dan nonfinansial umum tidak signifikan pada level 0,05 maka hipotesis kedua diterima. Hasil uji *within subject* menunjukkan bahwa interaksi antara divisi dan finansial umum signifikan pada level signifikansi 0,05 dengan nilai $F=22,125$ dan nilai $p=0,000$. Interaksi kedua yaitu divisi dan nonfinansial umum juga signifikan pada level signifikansi 0,05 dengan nilai $F=4,786$ dan nilai $p=0,035$ (tabel 6). Hal ini menunjukkan bahwa baik finansial umum maupun nonfinansial umum sama-sama digunakan dalam evaluasi kinerja.

Tabel 7 mengindikasikan bahwa ketika finansial umum lebih baik pada BR, manajer puncak mengevaluasi divisi BR 8,7 lebih tinggi daripada divisi BK. Sama halnya, ketika finansial umum lebih baik pada BK, divisi BK dievaluasi 7,478 lebih tinggi dari pada divisi BR. Sedangkan ketika nonfinansial umum lebih baik pada divisi BR, manajer puncak mengevaluasi divisi BR 4,05 lebih tinggi daripada divisi BK. Sama halnya, ketika nonfinansial umum lebih baik pada BK, divisi BK dievaluasi 3,956 lebih tinggi daripada divisi BR. Seluruh perbedaan tersebut signifikan.

Analisis Tambahan

Uji ANOVA sebelumnya belum mampu untuk menjawab hipotesis kedua yang menyatakan bahwa finansial umum lebih dipertimbangkan daripada non-finansial umum. Untuk melihat kecenderungan tersebut dilakukan perbandingan nilai *eta-squared* antara finansial umum dan non-finansial umum. Nilai *eta-squared* dihitung

dengan membagi nilai *sum of squares* untuk tiap interaksi dengan *total sum of squares* pada *within subject effect* (Kirk 1995 dalam Dilla dan Steinbart 2005). Hasil perhitungan menunjukkan bahwa *eta-squared* untuk divisi dengan pengaruh finansial umum (0,34) lebih tinggi daripada *eta-squared* untuk divisi dengan pengaruh nonfinansial umum (0,072). Hal ini menunjukkan bahwa finansial umum lebih digunakan dalam pertimbangan kinerja dibandingkan dengan nonfinansial umum. Hasil ini mendukung hipotesis kedua.

Diskusi dan Pembahasan

Tujuan penelitian ini adalah untuk menguji apakah dalam evaluasi kinerja manajer divisi yang menggunakan BSC, manajer atasan lebih cenderung untuk memperhatikan ukuran finansial umum (*common financial*) dibandingkan ukuran finansial yang unik (*unique financial*) dan ukuran non finansial yang umum (*common nonfinancial*). Penelitian ini merupakan lanjutan dari dua penelitian sebelumnya yaitu penelitian Lipe dan Salterio (2000) dan penelitian Cardinaels dan Veen-Dirks (2010). Pada penelitian Lipe dan Salterio (2000) mereka menemukan bahwa ketika melakukan evaluasi kinerja menggunakan BSC seseorang akan lebih cenderung memperhatikan ukuran-ukuran yang umum (*common measures*) dibandingkan ukuran-ukuran yang unik (*unique measures*). Sedangkan penelitian dari Cardinaels dan Veen-Dirks (2010) menemukan bahwa ketika mengevaluasi menggunakan BSC, seseorang cenderung untuk lebih memperhatikan ukuran finansial dari pada ukuran nonfinansial.

Hipotesis pertama penelitian ini mengeksplorasi *common measures bias* pada perspektif keuangan. Temuan penelitian ini mengindikasikan bahwa finansial umum lebih diperhatikan daripada finansial unik. Temuan ini sejalan dengan Slovic dan MacPhillamy (1974) yang menemukan bahwa ukuran umum lebih diperhatikan daripada ukuran unik ketika membandingkan dua mahasiswa *hipotetical*. Temuan ini juga

mendukung temuan Lipe dan Salterio (2000) bahwa ukuran-ukuran umum akan lebih diperhatikan daripada ukuran-ukuran unik. Penelitian ini juga mengkonfirmasi penelitian yang dilakukan oleh Banker *et al.* (2004) yang juga menemukan adanya kecenderungan untuk lebih memperhatikan ukuran-ukuran umum dibandingkan ukuran-ukuran unik. Mereka juga menemukan bahwa ukuran umum yang berkaitan dengan strategi (*common linked measures*) akan lebih diperhatikan daripada ukuran unik yang berkaitan dengan strategi (*unique linked measures*).

Hasil uji ANOVA *repeated measures* menunjukkan bahwa baik finansial umum maupun nonfinansial umum keduanya signifikan pada level 0,05. Ada beberapa alasan mengapa interaksi divisi dengan finansial umum dan nonfinansial umum keduanya signifikan pada level 0,05.

Pertama, kemungkinan para partisipan telah memahami bahwa penilaian kinerja menggunakan BSC tidak saja dipengaruhi oleh perspektif keuangan saja tetapi juga perspektif non keuangan. Hal ini menjadikan para partisipan menggunakan baik ukuran finansial umum maupun ukuran nonfinansial umum dalam evaluasi kinerja mereka.

Kedua, para partisipan dapat dengan mudah melakukan evaluasi kinerja karena desain kasus eksperimen dibuat lebih sederhana. Untuk finansial umum diwakili dua ukuran umum yang dimanipulasi dan untuk nonfinansial umum hanya diwakili ukuran-ukuran umum pada perspektif pelanggan tanpa memanipulasi ukuran-ukuran umum pada perspektif proses internal bisnis dan perspektif pertumbuhan dan inovasi. Penyederhanaan ini dimaksudkan untuk menghindari informasi yang *overloaded* bagi partisipan. Penyederhanaan seperti ini pernah dilakukan oleh Liedtka *et al.* (2008) dimana informasi kinerja yang ambigu hanya dimanipulasi pada perspektif keuangan saja tanpa memanipulasi pada

perspektif-perspektif yang lain.

Hasil uji *eta-squared* kemudian mengkonfirmasi hipotesis kedua. Temuan ini mendukung penelitian Cardinaels dan Veen-Dirks (2010) yang menemukan bahwa ketika mengevaluasi menggunakan BSC, seseorang cenderung untuk lebih memperhatikan ukuran finansial dari pada ukuran nonfinansial. Temuan ini juga mendukung penelitian dari Ittner *et al.* (2003a) yang menemukan bahwa dalam pertimbangan evaluasi kinerja menggunakan BSC, atasan lebih menggunakan informasi finansial dibandingkan dengan informasi nonfinansial.

SIMPULAN DAN SARAN

Simpulan

Ada dua implikasi penting dari penelitian ini. Pertama, bukti bahwa finansial unik kurang diperhatikan dalam evaluasi kinerja manajer akan memiliki implikasi yang signifikan pada keputusan strategi manajer setelahnya. Holmstrom dan Milgrom (1991) menunjukkan bahwa keputusan agen dipengaruhi oleh evaluasi kinerja dan kompensasi. Mereka juga menunjukkan bahwa item yang tidak dimasukkan dalam evaluasi kinerja dan kompensasi atas agen akan memiliki dampak yang kecil dalam keputusan agen. Hal ini akan mendorong manajer untuk lebih memperhatikan ukuran finansial umum dibandingkan ukuran finansial unik dalam setiap strategi keputusannya kemudian.

Kedua, BSC terdiri atas *lagging* dan *leading indicators*. *Lagging indicators* umumnya termanifestasi pada perspektif keuangan sedangkan *leading indicators* pada perspektif-perspektif yang lain. Fokus pada ukuran-ukuran keuangan akan mengabaikan *leading indicators* yang menjadi penunjang tercapainya ukuran-ukuran keuangan. Ketiga, evaluasi yang tidak komprehensif terhadap manajer divisi akan mengakibatkan misalokasi sumber daya perusahaan sehingga akan menyimpang dari tujuan yang diinginkan oleh perusahaan.

Keterbatasan

Beberapa keterbatasan dalam penelitian ini adalah: 1) Penelitian ini menggunakan mahasiswa S1 yang kurang berpengalaman dalam penilaian kinerja menggunakan BSC. Walaupun semua partisipan telah mengikuti mata kuliah akuntansi manajemen dan telah mempelajari BSC namun tingkat pemahaman mereka terkait BSC tidak dapat diketahui dengan pasti; 2) Partisipan dalam penelitian ini tidak terlibat dalam pengembangan BSC, padahal keterlibatan dalam pengembangan BSC akan mendorong mereka untuk mempertimbangkan seluruh pengukuran BSC termasuk ukuran unik dan ukuran nonfinansial; 3) Partisipan dalam penelitian ini masih sangat baru terkait BSC sehingga mereka tidak mempunyai pengalaman terkait dunia bisnis retail dan pakaian dimana latar belakang kasus ini dikembangkan; 4) Desain kasus eksperimen dalam penelitian ini terutama untuk menjawab hipotesis dua didesain secara sederhana sehingga kurang mampu menangkap realitas penilaian kinerja menggunakan BSC.

Saran

Penelitian lebih lanjut dapat diarahkan pada *debiasing* dari evaluasi kinerja. Sebagai mana dibahas dalam keterbatasan penelitian, partisipan tidak terlibat dalam pengembangan BSC. Penelitian selanjutnya dapat melihat apakah ketika seseorang dilibatkan dalam penyusunan BSC akan menggunakan seluruh ukuran-ukuran BSC baik finansial unik maupun ukuran nonfinansial.

Penelitian lanjutan dapat menguji perbandingan antara finansial umum dan nonfinansial umum dengan desain kasus yang lebih kompleks yaitu memanipulasi tidak hanya pada perspektif pelanggan namun juga pada perspektif proses internal bisnis maupun perspektif pembelajaran dan inovasi. Hal itu dilakukan untuk melihat apakah terdapat informasi *overloaded* atau tidak ketika mengevaluasi menggunakan BSC.

DAFTAR PUSTAKA

- Anand, M., B. S. Sahay, dan S. Saha. 2005. Balanced scorecard in Indian companies. *Vikalpa* 30(3): 11-23.
- Banker, R. D., H. Chang, dan M. Pizzini. 2004. The Balanced Scorecard: judgmental effects of performance measures linked to strategy. *The Accounting Review* 79(1): 1-23.
- . 2009. The judgmental effects of strategy maps in Balanced Score-card performance evaluations. *Working Paper*.
- Braam, G. J. M. dan E. J. Nijssen. 2004. Performance effect of using the Balanced Scorecard: a note on Dutch experience. *Long Range Planning* 37: 335-349.
- Burney, L. L. dan N. J. Swanson. 2010. The relationship between Balanced Scorecard characteristics and managers' job satisfaction. *Journal of Managerial Issues* 22(2): 166-181.
- Cardinaels, E. dan P. M. G. V. Veen-Dirks. 2010. Financial versus non-financial information: the impact of information organization and presentation in a Balanced Score-card, *Accounting, Organizations and Society* 35: 565-578.
- Cheng, M. M., A. K. D. Schulz, P. F. Lockett, dan P. Booth. 2003. The Effects of Hurdle Rates on the Level of Escalation of Commitment in Capital Budgeting. *Behavioral Research in Accounting* 15: 63-85.
- Davis, S. dan T. Albright. 2004. An investigation of the effect of Balanced Scorecard implementation on financial performance. *Management Accounting Research* 15: 135-153.
- DeBusk, G. K., R. M. Brown, dan L. N. Killough. 2003. Components and relative weights in utilization of performance measurement systems like the Balanced Scorecard. *British Accounting Review* 35(3): 215-231.
- Dilla, W. N., dan P. J. Steinbart. 2005. Relative weighting of common and unique Balanced Scorecard measures by knowledgeable decision makers.

- Behavioral Research in Accounting* 17: 43-53.
- Downing, L. 2000. Progress report on the Balanced Scorecard: a global users' survey. *Balanced Scorecard Report*.
- Hibbets, A. R., M. L. Roberts, dan T. L. Albright. 2006. Common-measure bias in the Balanced Scorecard: cognitive effort and general problem-solving ability. *Working Paper*.
- Holmstrom, B., dan P. Milgrom. 1991. Multi task principal-agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization* 7: 24-52.
- Hoque, Z dan W. James. 2000. Linking balanced scorecard measures to size and market factors: Impact on organizational performance. *Journal of Management Accounting Research*, 12: 1-17.
- Ittner, C. D., D. F. Larcker dan T. Randall. 2003a. Performance implications of strategic performance measurement in financial services firms. *Accounting, Organizations and Society* 28(7-8): 715-741.
- Ittner, C. D., D. F. Larcker, dan M. W. Meyer. 2003b. Subjectivity and the weighting of performance measures: evidence from a Balanced Scorecard. *The Accounting Review* 78(3): 725-758.
- Kaplan, R. S. 2009. Conceptual foundations of the Balanced Scorecard. *Handbook of Management Accounting Research* 3: 1253-1269.
- Kaplan, R. S. dan D. P. Norton. 1992. The Balanced Scorecard measures that drive performance. *Harvard Business Review* 70(1).
- . 1993. Putting the Balanced Scorecard to work. *Harvard Business Review* 71(5).
- . 1996a. Using the Balanced Scorecard as a strategic management system. *Harvard Business Review* 74(1): 75-85.
- . 1996b. *The Balanced Scorecard: Translating strategy into action*. Harvard Business School Press. Boston, MA
- . 2001. *The Strategy-Focused Organization: how Balanced Scorecard companies thrive in the new business environment*. Harvard Business School Press. Boston, MA.
- Kaufmann, L. dan A. Becker. 2005. Overcoming the barriers during implementation and use of the Balanced Scorecard by multinational companies in Brazil. *Latin America Business Review* 6(3): 39-62.
- Kliger, D. dan A. Kudryavtsev. 2010. The Availability Heuristic and Investors' Reaction to Company-Specific Events. *The Journal of Behavioral Finance* 11: 50-65.
- Lawson, R., T. Hatch, dan D. Desroches. 2008. *Scorecard best practices: design, implementation and evaluation*. John Wiley and Sons. New Jersey.
- Liedtka, S. L., B. K. Church, dan M. R. Ray. 2008. Performance variability, ambiguity intolerance, and Balanced Scorecard-based performance assessments. *Behavioral Research in Accounting* 20(2): 73-88.
- Lipe, M. G. dan S. Salterio. 2000. The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review* 75(3): 283-298.
- . 2002. A note on the judgmental effects of the Balanced Scorecard's information organization. *Accounting, Organizations and Society* 27: 531-540.
- Malina, M. A. dan F. H. Selto. 2001. Communicating and controlling strategy: an empirical study of the effectiveness of the balanced scorecard. *Journal of Management Accounting Research*, 13: 47-75.
- Malmi, T. 2001. Balanced Scorecards in Finnish companies: a research note. *Management Accounting Research* 12: 207-220.
- Roberts, M., T. Albright, dan A. Hibbets. 2004. Debiasing Balanced Scorecard evaluations. *Behavioral Research in Accounting* 16: 75-88.

- Slovic, P., dan D. MacPhillamy. 1974. Dimensional commensurability and cue utilization in comparative judgment. *Organizational Behavior and Human Performance* 11: 172-194.
- Tversky, A., dan D. Kahneman. 1973. Availability: a heuristic for Judging frequency and Probability. *Cognitive Psychology* 4: 207-232.
- . 1974. Judgment under uncertainty: heuristics and biases. *Science* 185: 1124-1131.
- Wong-On-Wing, B., L. Guo, W. Li, dan D. Yang. 2007. Reducing conflict in Balanced Scorecard evaluations. *Accounting, Organizations and Society* 32: 363-377.