

STRATEGI MENGATASI *COMMON MEASURES BIAS* DALAM *BALANCED SCORECARD*

**Sekar Akrom Faradiza¹⁾
Vera Desy Nurmalia²⁾**

¹⁾Fakultas Bisnis dan Teknologi Informasi, Universitas Teknologi Yogyakarta
e-mail: sekar.akrom@gmail.com

²⁾Fakultas Bisnis dan Teknologi Informasi, Universitas Teknologi Yogyakarta
e-mail: vera_dmalia@yahoo.co.id

Abstract

Balanced Scorecard (BSC) is a comprehensive performance measurement. BSC is not only used financial indicators but also non financial indicators such as customers, internal process business, and learning and growth perspective. By using BSC, evaluators have common and unique measures. When evaluate manager performance, evaluator tends to only use common measures and ignore unique measures. This is called common measures bias. This study aims to investigate whether disaggregated and aggregated BSC and management communication can overcome common measures bias and intent to BSC approach. This study also will evaluate whether these approach will affect evaluator decision when allocated compensation. We conduct 2x2x2 experiment of undergraduate accounting students. Participant act as a senior manager and evaluate the performance of two divisions and then allocated the bonus. ANOVA repeated measurement are used to conduct hypothesis test. The results showed that disaggregated BSC and management communication could not overcome common measures bias but effected management decision when allocated compensation.

Keywords: common measures bias, dissaggregated, management communication, compensation allocation

PENDAHULUAN

Balanced Scorecard (BSC) adalah suatu alat pengukuran kinerja yang paling populer dan banyak diimplementasikan oleh organisasi-organisasi baik publik maupun privat dibanding model pengukuran lainnya. Di Indonesia, BSC juga banyak digunakan oleh beberapa perusahaan BUMN dan swasta. BSC pertama kali muncul pada tahun 1992 yang dikenalkan oleh Kaplan dan Norton. BSC memiliki 4 perspektif penilaian kinerja yaitu perspektif finansial, pelanggan, proses bisnis internal, dan pertumbuhan dan pembelajaran. Masing-masing perspektif memiliki hubungan sebab akibat dan memiliki keterkaitan dengan strategi. Dalam perkembangannya BSC tidak lagi sebagai alat pengukuran kinerja semata namun juga merupakan alat penerapan strategi. Dengan BSC maka misi dan strategi perusahaan akan dapat dikomunikasikan dengan baik kepada seluruh anggota organisasi sehingga akan mendorong mereka untuk melaksanakannya dan pada akhirnya tujuan utama organisasi dapat tercapai.

Kaplan dan Norton (1996) menemukan bukti bahwa banyak manajer yang hanya mempertimbangkan ukuran finansial semata dalam mengevaluasi kinerja bawahannya. Mereka cenderung mengabaikan elemen penting dalam misi strategik perusahaan dan

secara tidak sengaja menekankan pada *lag indicator* dibandingkan *lead indicator*. Sehingga Kaplan dan Norton (1996) menciptakan *Balanced Scorecard* sebagai alat pengukur kinerja yang dapat menangkap strategi unik dari perusahaan atau unit bisnis.

Ketika perusahaan menerapkan BSC, maka penilai kinerja atau evaluator harus memberikan bobot penilaian pada seluruh komponen penilaian baik komponen yang bersifat umum (*common*) maupun yang unik (*unique*). Namun demikian, isu terkait penggunaan BSC sebagai alat pengukuran kinerja yang sering muncul adalah *common measures bias* (Lipe dan Salterio, 2000; Malina dan Selto, 2001; Banker *et al.* 2004, Roberts *et al.*, 2004). *Common measures bias* terjadi ketika evaluator lebih menekankan penilaian kinerja berdasarkan ukuran yang bersifat umum dan cenderung mengabaikan ukuran kinerja yang bersifat unik yang merepresentasikan masing-masing unit usaha secara spesifik. Hal ini mengindikasikan bahwa BSC sebagai alat manajemen yang dianggap telah modern yang seharusnya dapat melihat keseluruhan indikator, ternyata masih belum optimal diterapkan dalam perusahaan karena para pengguna BSC dalam menilai kinerja bawahannya masih berfokus pada ukuran-ukuran umum yang dilekatkan pada BSC.

Beberapa penelitian telah dilakukan untuk mencari pendekatan apa yang dapat secara efektif mengurangi atau bahkan menghilangkan *common measures bias* ini. Libby *et al.* (2004) menggunakan proses peningkatan akuntabilitas dan peningkatan persepsi kualitas dari ukuran BSC melalui *assurance* untuk mengurangi bias tersebut. Hasil penelitiannya menunjukkan bahwa kedua pendekatan tersebut mampu mengatasi permasalahan bias yang terjadi. Namun desain eksperimen yang dilakukan kurang baik karena manajer sebagai partisipan diarahkan untuk mengikuti strategi dengan *common measure* yang relatif lebih penting dibandingkan *unique measures*. Desain eksperimennya secara sengaja memilih untuk tidak menekankan pada *unique measure* yang artinya secara langsung masih terdapat *common measure bias* di dalamnya.

Hasil penelitian Dilla dan Steinbart (2002) serta Roberts *et al.* (2002) menemukan bukti bahwa pelatihan dan komunikasi yang jelas dapat meningkatkan penilaian kinerja yang mempertimbangkan seluruh indikator kinerja dalam BSC (*common dan unique measures*). Bentuk komunikasi yang dapat dilakukan oleh evaluator adalah dengan memberikan penekanan dengan memberikan informasi *causal chain* dari masing-masing perspektif (Tayler, 2007) atau dengan penekanan bahwa ukuran kinerja tersebut memiliki keterkaitan dengan strategi yang diterapkan oleh perusahaan (Banker, *et al.* 2004). Namun hasil ini tidak terdukung oleh hasil penelitian Kaplan dan Wisner (2009) bahwa adanya komunikasi manajemen tentang pentingnya suatu indikator kinerja tidak akan memberikan peningkatan pertimbangan evaluator pada indikator tersebut.

Roberts *et al.* (2004) melakukan eksperimen dua tahap untuk menemukan cara mengatasi bias (*debiasing*) yang muncul yaitu dengan pendekatan *disaggregated mechanically aggregated* yang kemudian disebut dengan *Disaggregated BSC*. Dalam eksperimennya tersebut partisipan diminta menilai secara terpisah (*disaggregated*) 16 ukuran kinerja dan kemudian menggabungkan kembali (*mechanically aggregated*) ukuran-ukuran tersebut dengan mengalikannya dengan bobot tertentu. Hasil penelitiannya tersebut bahwa pendekatan yang digunakan mampu mengatasi *common measures bias*.

Ketepatan penilaian kinerja adalah hal yang perlu dijaga di dalam sebuah organisasi. Terlebih apabila penilaian kinerja tersebut akan berdampak pada kompensasi yang diterima oleh seorang karyawan. Meskipun demikian, alokasi bonus atau kompensasi oleh evaluator terhadap unit yang dievaluasinya tergantung pada pertimbangan subjektivitasnya

selama proses pengukuran kinerja. Ketika seluruh indikator kinerja digunakan maka alokasi bonus akan menjadi lebih tepat. Dalam penelitian Roberts *et al.* (2004) bahwa dengan pendekatan *disaggregated* BSC maka seluruh indikator akan digunakan dalam menentukan bonus oleh evaluator. Namun demikian adanya komunikasi yang lebih ekstensif tentang pentingnya suatu ukuran kinerja ternyata tidak memberikan dampak yang signifikan pada perbedaan alokasi bonus (Kaplan dan Wisner, 2009).

Berdasarkan latar belakang yang telah diuraikan sebelumnya, maka penelitian ini akan menguji efektivitas pendekatan *debiasing*. Penelitian ini adalah ekstensi dari penelitian yang dilakukan oleh Roberts *et al.* (2004). Perbedaan penelitian ini dengan penelitian Roberts *et al.* (2004), seperti yang disarankannya, bahwa penelitian ini memisahkan dua langkah dalam BSC (*disaggregated* dan *mechanically aggregated*) apakah masing-masing langkah mampu mengatasi secara efektif *common measures bias* yang terjadi dalam penerapan BSC. Selain itu penelitian ini akan menambahkan variabel komunikasi manajemen yang dianggap mampu memberikan pemahaman bahwa penilaian dengan mempertimbangkan seluruh indikator (*common* dan *unique*) adalah hal penting dalam penerapan BSC. Penelitian ini juga ingin melihat bagaimana penerapan BSC yang harus diikuti oleh skema kompensasi agar menjadikan kompensasi dapat ditentukan dengan lebih tepat, di mana hal ini jarang diteliti.

Common Measures Bias

Penelitian di bidang psikologi menunjukkan bahwa ketika pengambil keputusan dihadapkan pada evaluasi komparatif maka mereka cenderung menggunakan informasi yang umum dan mengabaikan ukuran yang unik atas masing-masing objek (Slovic dan MacPhillany, 1974). Kecenderungan penggunaan ukuran yang umum (*commonmeasure*) muncul ketika objek dinilai secara berpasangan. Namun hal yang sama tidak akan terjadi ketika objek dievaluasi secara individual.

Slovic dan MacPhillany (1974) melakukan eksperimen dengan meminta partisipan untuk memprediksi IPK mahasiswa pada tahun pertama di universitas. Masing-masing partisipan menerima informasi yang terdiri dari dua informasi umum dan satu informasi unik tentang masing-masing mahasiswa. Hasil eksperimen tersebut menunjukkan partisipan lebih menggunakan pertimbangannya pada informasi umum dan hanya sedikit menggunakan informasi unik. *Common measures bias* ini tetap muncul meskipun peneliti memberikan *feedback*, meningkatkan kualitas informasi dan memberikan insentif untuk mendorong partisipan menggunakan seluruh informasi sehingga dapat mengatasi bias yang muncul.

Lipe dan Salterio (2000) melakukan eksperimen dengan menggunakan mahasiswa MBA sebagai partisipan. Mereka diminta untuk mengevaluasi kinerja manajer secara independen. Hasil penelitiannya konsisten dengan Slovic dan MacPhillany (1974) bahwa partisipan cenderung menggunakan ukuran yang umum dan mengabaikan ukuran unik dalam mengevaluasi kinerja manajer dengan menggunakan BSC.

Dari perspektif teori, penilaian kinerja yang memperlihatkan adanya *common measure bias* merupakan penilaian yang berkualitas lebih buruk daripada evaluasi yang mendasarkan pada seluruh ukuran dalam BSC. Holmstrom (1979) menyatakan bahwa ukuran yang informatif adalah ukuran kinerja yang dapat memberikan informasi tentang tindakan manajer selama terdapat ketidakpastian spesifik dari lingkungan divisinya. Adanya ukuran yang unik dalam BSC menunjukkan tingkat perbedaan ketidakpastian dari masing-masing lingkungan operasi divisi sehingga penggunaan ukuran yang umum saja

menjadi tidak sesuai (Libby *et al.* 2004). Dari perspektif praktis, penilaian kinerja akan mendorong manajer untuk menyelaraskan tindakannya berdasarkan pada strategi yang unik dari divisinya. Dalam BSC seluruh ukuran menjadi relevan dengan strategi (Kaplan dan Norton, 1996), sehingga seluruh ukuran (*common dan unique*) harus dipertimbangkan ketika menilai kinerja.

Roberts *et al.* (2004) menyatakan bahwa ukuran umum seringkali mendominasi evaluasi komparasi karena tiga alasan yaitu:

- a. Merupakan bentuk yang lebih kecil dari seluruh informasi yang secara kognitif lebih mudah untuk dijaga dan diproses
- b. Merupakan kategori atau tipe informasi yang lebih kecil untuk diproses
- c. Ukuran umum merupakan informasi yang hanya tersedia secara langsung membandingkan manajer

Adanya kecenderungan penggunaan ukuran yang umum menimbulkan adanya bias pada penilaian kinerja. Hal ini cukup menarik perhatian para peneliti untuk melihat dan menguji model yang paling efektif dalam mengatasi bias ini. Libby *et al.* (2004) melakukan eksperimen dengan menggunakan mahasiswa MBA sebagai responden. Hasil penelitiannya menunjukkan bahwa proses akuntabilitas dengan menyesuaikan evaluasi kinerja dengan atasan atau memberikan laporan penjaminan atas BSC akan meningkatkan penggunaan ukuran unik.

Salah satu metode untuk meningkatkan kualitas pertimbangan adalah dengan menggunakan model bantuan keputusan (Kennedy, 1995). Einhorn (1972) menyatakan bahwa akurasi akan meningkat ketika pertimbangan manusia atas keputusan informasi dikodekan dalam bentuk kuantitatif dan output akan diperoleh dari penggabungan metode. Berdasarkan pernyataan tersebut kemudian Roberts *et al.* (2004) melakukan penelitian untuk melihat apakah pendekatan *disaggregated* BSC mampu mengatasi *common measures bias* yang muncul dalam konteks BSC. Hasil penelitiannya menunjukkan bahwa pendekatan tersebut mampu mengatasi adanya bias yang terjadi pada penggunaan BSC sebagai alat penilaian kinerja.

Slovic dan MacPhillamy (1974) menyatakan bahwa pembuat keputusan gagal untuk menggunakan informasi unik karena hal tersebut lebih sulit digunakan dalam evaluasi dibandingkan ukuran yang umum. Informasi umum dapat langsung digunakan dan tidak mengandung ambiguitas sedangkan informasi unik memiliki *trade-off* antar atribut yang berbeda. Adanya *trade-off* tersebut secara kognitif membuat penggunaan informasi unik menjadi lebih sulit karena seseorang harus menggunakan usaha kognitif tambahan untuk mengintegrasikan informasi unik tersebut (Libby *et al.* 2004). Hal ini juga mendukung temuan Lippe dan Salterio (2000) yang menyatakan bahwa manajer mungkin mengabaikan ukuran unik dari BSC karena untuk memprosesnya diperlukan usaha kognitif yang lebih besar.

Disaggregated dan Aggregated BSC

Einhorn (1972) dan Bowman (1963) menyatakan bahwa pendekatan dalam BSC melibatkan dua langkah proses, yaitu memisahkan (*disaggregate*) keputusan evaluasi menjadi beberapa keputusan yang lebih kecil dan menggabungkan (*aggregate*) keputusan yang lebih kecil menjadi satu skor secara keseluruhan berdasarkan bobot yang telah ditentukan. Pada langkah pertama, dengan memisahkan atau memecahkan keputusan yang kompleks, maka akan mendorong proses yang luas atas masing-masing dimensi secara individual. Ketika perhatian berfokus pada satu dimensi, maka kerja dari memori jangka

pendek seorang evaluator akan bebas dari informasi dimensi lain. Perubahan perhatian dan kemampuan memproses tersebut harus difasilitasi dengan usaha yang lebih besar dan memastikan bahwa usaha tersebut digunakan dalam seluruh ukuran. Langkah ini akan mengatasi *common measures bias* sejauh bias tersebut disebabkan oleh kegagalan untuk memperhatikan secara memadai pada ukuran unik.

Langkah kedua, bobot yang telah ditentukan digunakan untuk menggabungkan evaluasi menjadi skor menyeluruh akan memperkuat pentingnya ukuran umum dan unik dalam organisasi. Sehingga kedua ukuran tersebut akan digunakan untuk mengevaluasi secara keseluruhan karena pengambil keputusan mengeluarkan biaya untuk memproses seluruh dimensi (Roberts *et al.* 2004).

Pertimbangan yang terpisah (*disaggregated*) lebih berguna untuk masalah yang lebih kompleks, misalnya pertimbangan yang melibatkan 9 informasi. Sebagai perbandingan, BSC membutuhkan 4 sampai 7 ukuran kinerja untuk masing-masing perspektif (Kaplan dan Norton, 1996). Pengguna BSC akan memiliki 16 – 28 informasi yang akan digunakan untuk mengevaluasi kinerja manajer. Sehingga pertimbangan menggunakan BSC cukup kompleks dan akan menguntungkan ketika digunakan pertimbangan *disaggregated* dan *aggregated*.

Pertimbangan *disaggregated* dan *mechanically aggregated* akan mengurangi dan meningkatkan proses kognitif. Kognitif yang dibutuhkan berkurang pada suatu waktu karena sejumlah informasi harus dipertimbangkan secara individual. Jumlah waktu dan usaha yang dibutuhkan meningkat karena jumlah evaluasi dan perhitungan meningkat (Roberts *et al.* 2004). Berdasarkan eksperimen yang dilakukan, Roberts *et al.* (2004) menemukan bukti bahwa kedua langkah dalam BSC (*disaggregate* dan *aggregate*) efektif dalam mengatasi *common measures bias*. Dua langkah dalam BSC tersebut masing-masing seharusnya mampu untuk mengatasi *common measure bias* dalam konteks BSC karena adanya penekanan masing-masing langkah dalam menggunakan BSC sebagai alat pengukuran kinerja.

Berdasarkan uraian tersebut maka peneliti mengharapkan bahwa langkah dalam BSC (*disaggregate* dan atau *aggregate*) memiliki tingkat efektivitas yang sama dalam mengatasi *common measures bias* dalam BSC. Sehingga dapat dirumuskan hipotesis sebagai berikut:

H1 : Penyajian BSC dalam format *disaggregated* akan membuat evaluator menggunakan seluruh pengukuran kinerja (tidak terjadi *common measures bias*)

Komunikasi Manajemen

Komunikasi manajemen atas visi dan strategi kepada seluruh karyawan dalam organisasi merupakan hal penting dalam kesuksesan implementasi visi dan strategi tersebut (Kaplan dan Norton, 1992, 1996). Efektivitas implementasi strategi sangat tergantung pada luasnya manajer menengah dan operasional untuk menginformasikan dan berbagi pemahaman manajemen puncak atas tujuan strategik yang akan dicapai dan diprioritaskan dalam perusahaan. Dengan demikian maka seluruh anggota organisasi menjadi paham tentang apa yang harus mereka lakukan agar berkontribusi terhadap kesuksesan perusahaan. Hal ini didukung oleh temuan Malina dan Selto (2001) bahwa komunikasi manajemen terkait strategi perusahaan adalah hal penting dalam efektivitas penerapan BSC.

Ketika strategi terdiri dari beberapa komponen, maka komunikasi digunakan untuk mengindikasikan pentingnya komponen strategi yang spesifik. Komunikasi adalah bentuk persuasi, dimana manajemen menyebarkan informasi yang akan mendorong pada penyelarasan yang lebih baik antara nilai dan norma pada manajemen dan karyawannya (Kaplan dan Wisner, 2009). Dalam hal ini komunikasi yang mendiskusikan komponen strategik tertentu, konsisten dengan perspektif institusional yang melihat komunikasi organisasi sebagai proses di mana norma diadaptasi dan dilakukan oleh anggota organisasi.

Dengan demikian, ketika evaluator mengetahui bahwa ukuran-ukuran kinerja yang digunakan memiliki keterkaitan dengan strategi maka evaluator akan lebih mempertimbangkan seluruh indikator kinerja (*common dan unique*) dibandingkan dengan evaluator yang tidak memahami keterkaitan ukuran kinerja dengan strategi (Banker *et al.* 2004). Hasil penelitian Talyer (2010) menunjukkan bahwa ketika terdapat komunikasi terhadap evaluator bahwa terdapat *causal chain* antar perspektif dalam BSC maka evaluator akan menilai bahwa suatu inisiatif atau kegiatan telah sukses diterapkan atau dilaksanakan dari pada evaluator yang tidak mengetahui adanya *causal chain* tersebut.

Dalam kondisi dimana ukuran yang spesifik atau unik diabaikan oleh evaluator, maka manajemen harus mampu menekankan tentang pentingnya ukuran ini melalui komunikasi dalam organisasi. Literatur bisnis profesional dan akademis menunjukkan bahwa ketika manajer mengkomunikasikan justifikasi yang kuat kepada anggota organisasi, maka hal tersebut telah memberikan sinyal tentang komitmen dan dukungan yang kuat untuk mencapai tujuan strategik perusahaan (Kaplan dan Wisner, 2009).

Berdasarkan uraian *disaggregated* dan *aggregated* dalam BSC serta komunikasi manajemen, maka peneliti mengharapkan bahwa ketika intensitas komunikasi manajemen tinggi, maka seluruh pendekatan dalam BSC (*disaggregated* dan atau *aggregated*) akan mampu mengatasi *common measures bias* secara lebih efektif dibandingkan ketika intensitas komunikasi manajemen rendah. Sehingga dapat dirumuskan hipotesis sebagai berikut:

H2: Penyajian BSC dalam format *disaggregated* akan lebih efektif membuat evaluator mempertimbangkan seluruh indikator kinerja ketika intensitas komunikasi manajemen tinggi dibandingkan ketika intensitas komunikasi rendah.

Balanced Scorecard dan Kompensasi

Secara konseptual, ketika evaluasi kinerja dilaksanakan dengan menggunakan BSC maka harus dihubungkan dengan kompensasi manajer (Kaplan dan Norton, 1996). Meskipun demikian, secara tradisional perusahaan mencoba mengimplementasikan BSC sebagai usaha percobaan dan berusaha untuk menjadi familiar terlebih dahulu sebelum mengubah sistem kompensasi mereka. Dalam hal ini Kaplan dan Norton (1996) tidak memberikan rekomendasi bagaimana penilaian dengan BSC digunakan untuk memutuskan kompensasi.

Supervisor mungkin menjadi enggan untuk menggunakan alat pengukuran kinerja yang formal yang tidak memungkinkan mereka untuk memberikan kompensasi berdasarkan kebijakan mereka. Sehingga menjadi penting untuk menentukan apakah atasan akan mengikuti prosedur formal dari BSC untuk membuat keputusan kompensasi (Roberts *et al.* 2004).

Lippe dan Salterio (2000) tidak menghubungkan kompensasi dengan penilaian berdasar BSC. Hal tersebut kemudian dilakukan oleh Roberts *et al.* (2004) melalui

eksperimen. Hasil penelitiannya menunjukkan bahwa partisipan menggunakan penilaian berdasarkan BSC dalam menentukan bonus atau kompensasi bagi manajer. Namun modelnya hanya menunjukkan 55% dari variasi perbedaan bonus yang artinya bahwa dalam menentukan bonus mereka menyesuaikan dengan faktor-faktor yang tidak termasuk dalam BSC.

Ketika perusahaan mengimplementasikan BSC dan manajemen mampu mengkomunikasikan dengan baik tentang tujuan, ukuran dan target yang akan digunakan dalam menilai kinerja, maka hal tersebut akan berimplikasi pada penentuan bonus. Ketika intensitas komunikasi tinggi maka keputusan kompensasi akan dipengaruhi oleh penilaian secara menyeluruh (*common dan unique measure*) dari BSC. Berdasarkan uraian tersebut maka dapat dirumuskan hipotesis sebagai berikut:

H3: Intensitas komunikasi manajemen yang tinggi dan penggunaan BSC dalam format *disaggregated* akan mempengaruhi keputusan kompensasi untuk manajer.

METODE PENELITIAN

Partisipan

Partisipan dalam penelitian ini adalah mahasiswa semester akhir (semester 7 atau 8) yang telah lulus mata kuliah Akuntansi Manajemen dan Sistem Pengendalian Manajemen dari beberapa universitas di Yogyakarta. Pemilihan sampel tersebut didasari pada pemikiran bahwa mahasiswa yang telah menempuh 2 mata kuliah itu telah memperoleh pengetahuan tentang BSC dan telah memiliki pertimbangan yang lebih dewasa dan bijak dalam melihat atau menilai sesuatu. Selain itu mahasiswa tingkat atas biasanya telah memiliki pengalaman organisasi baik di dalam maupun luar kampus sehingga mampu berpikir secara lebih kritis.

Tugas Partisipan

Partisipan dalam eksperimen ini akan dikenalkan dengan dua divisi dari PT Sandang yang merupakan industri garmen khusus untuk pakaian wanita. Eksperimen dilakukan dalam suatu ruangan. Pada tahap awal partisipan diasumsikan berperan sebagai eksekutif senior dari PT Sandang yang telah mengikuti pelatihan tentang BSC yang diselenggarakan oleh IAMI (Ikatan Akuntan Manajemen Indonesia) di Jakarta beberapa waktu lalu. Kemudian partisipan akan diberitahu visi dari PT Sandang dan dikenalkan dengan 2 manajer divisi yaitu Divisi Feminim dan Divisi Kasual. Kemudian partisipan diberitahu strategi dan *Balanced Scorecard* dari masing-masing divisi.

Masing-masing BSC terdiri dari 16 ukuran yang terpisah, 4 indikator untuk setiap perspektif, 2 ukuran merupakan *common measures* dari kedua divisi dan 2 ukuran adalah *unique measures* untuk masing-masing divisi. Dua *unique measures* dari Divisi Kasual adalah *new store sales* dan *market share to retail space*. Untuk Divisi Feminim adalah *revenue per sales visit* dan *catalog profit*. Berdasarkan 16 ukuran kinerjanya, kedua divisi memiliki kinerja yang lebih baik dibandingkan dengan targetnya. Masing-masing ukuran kinerja memiliki bobot yang secara eksplisit tercantum yang merupakan bentuk dari komunikasi manajemen akan pentingnya seluruh indikator kinerja (*common dan unique measures*).

Kemudian berdasarkan masing-masing BSC partisipan diminta untuk memberikan penilaian dengan skala 1-100 pada kedua divisi. Pada *disaggregated* BSC maka partisipan

akan menerima BSC dengan terdapat satu kolom tambahan berupa bobot penilaian dari masing-masing indikator. Pada *aggregated* BSC partisipan tidak memiliki tentang bobot tersebut. Pada kondisi tingkat intensitas komunikasi manajemen yang tinggi maka partisipan akan menerima informasi bahwa manajer selalu menyampaikan dalam rapat koordinasi bahwa seluruh indikator kinerja (*common* dan *unique measures*) penting untuk digunakan terkait penggunaan BSC sebagai alat pengukuran kinerja. Pada kondisi komunikasi manajemen yang rendah maka partisipan tidak mendapat informasi apa-apa.

Berdasarkan penilaian kinerja yang dibuat kemudian partisipan diminta mengalokasikan bonus sebesar Rp50.000.000,- untuk manajer kedua divisi. Setelah menyelesaikan eksperimennya kemudian partisipan menyelesaikan *debriefing questionnaire* yang menanyakan tentang kasus, menjawab pertanyaan *manipulation check* dan menanyakan tentang kesulitan kasus, realisme, dan tingkat keterpahaman kasusnya.

Desain Eksperimen

Desain eksperimen yang digunakan dalam penelitian ini adalah desain 2 x 2 x 2 dengan 2 *between subject* dan 1 *within subject* di mana 2 faktor yang pertama adalah metode evaluasi kinerja dengan BSC (*disaggregated* dan *aggregated*), 2 faktor yang kedua adalah dua intensitas komunikasi manajemen (rendah dan tinggi) dan 2 faktor yang ketiga adalah *within subject* di mana partisipan diminta untuk menilai kinerja dari manajer pada dua divisi yang berbeda. Jika digambarkan dalam matriks, maka desain eksperimen yang akan dilakukan adalah sebagaimana disajikan dalam Tabel 1.

Tabel 1. Desain Eksperimen

		Format BSC	
		<i>Disaggregated</i>	<i>Aggregated</i>
Komunikasi	<i>Low</i>	Grup 1	Grup 2
Manajemen	<i>High</i>	Grup 3	Grup 4

Prosedur Eksperimen

- Berikut ini adalah tahapan yang dilakukan selama eksperimen berlangsung, meliputi:
1. Partisipan dikenalkan dengan misi dan strategi dari dua divisi yang akan diukur kinerjanya. Kedua divisi tersebut adalah Divisi Kasual yang fokus pada pakaian remaja. Divisi Feminim adalah spesialis pakaian seragam kerja wanita.
 2. Kemudian secara random partisipan dibagi menjadi empat kelompok berdasarkan pada amplop yang mereka ambil sebelum masuk ke ruang eksperimen. Partisipan diberitahu bahwa tidak terdapat jawaban yang benar atau salah dan tidak perlu mencantumkan nama pada lembar kerjanya. Peneliti akan memberikan *souvenir* sebagai tanda terima kasih atas partisipasi dalam eksperimen ini.
 3. Terdapat 4 amplop yang menunjukkan kondisi untuk 4 grup. Pada evaluasi kinerja dengan pendekatan *disaggregated* BSC di mana kinerja manajer akan dievaluasi pada masing-masing 16 ukuran kinerja BSC dengan menggunakan skala 0 (tidak dapat diterima) sampai 100 (sangat baik). Pada pendekatan *aggregated* maka partisipan akan mengalihkan pertimbangan individu dengan bobot yang telah ditentukan sebelumnya.

4. Pada kondisi tingkat intensitas komunikasi manajemen yang tinggi maka partisipan akan menerima informasi bahwa manajer selalu menyampaikan dalam rapat koordinasi bahwa seluruh indikator kinerja (*common* dan *unique measures*) penting untuk digunakan terkait penggunaan BSC sebagai alat pengukuran kinerja. Pada kondisi komunikasi manajemen yang rendah maka partisipan tidak mendapat informasi apa-apa.
5. Kemudian partisipan membuat penilaian kinerja masing-masing manajer secara keseluruhan dengan menggunakan skala dari 0 – 100.
6. Setelah itu partisipan akan mengalokasikan bonus akhir tahun sebesar Rp50.000.000,- kepada dua manajer divisi berdasarkan pengukuran kinerja yang telah dilakukan sebelumnya.
7. Setelah menyelesaikan eksperimennya kemudian partisipan menyelesaikan *debriefing questionnaire* yang menanyakan tentang kasus, menjawab pertanyaan *manipulation check* dan menanyakan tentang kesulitan kasus, realisme dan tingkat keterpahaman kasusnya.

Pilot Test

Pilot test bertujuan untuk menginvestigasi apakah skenario yang diberikan dapat dipahami oleh partisipan dan mengevaluasi kesalahan dalam rancangan dan ketepatan *treatment* yang akan diberikan. *Pilot test* akan melibatkan mahasiswa Program Profesi dan Program Pascasarjana di Universitas Teknologi Yogyakarta. Prosedur ini dilakukan untuk menguji apakah kasus dan perlakuan yang akan diberikan dapat dipahami oleh subjek dan untuk menguji kesalahan yang mungkin terdapat dalam desain.

Dalam *pilot test* tersebut partisipan diminta untuk memberikan saran-saran perbaikan yang berkenaan dengan instrumen yang akan digunakan untuk meningkatkan validitas instrumen. Tujuan *pilot test* ini juga untuk menginvestigasi bagaimana mengendalikan ancaman yang dapat merusak validitas internal selama eksperimen berlangsung misalnya keadaan lingkungan atau peristiwa yang tidak direncanakan, partisipan mengalami kewalahan, lapar, dan efektivitas waktu eksperimen. Pertimbangan akan hal tersebut menjadi perhatian eksperimen agar validitas internal tetap terjaga.

Dalam penelitian ini, *pilot test* dilakukan dengan partisipan yaitu 5 orang dosen Akuntansi di FITB UTY dan 15 orang mahasiswa Pendidikan Profesi Akuntansi FITB UTY. Hasil *pilot test* berupa masukan atas instrumen penelitian yang meliputi tata bahasa, konsep, dan penekanan manipulasi atau *treatment* yang dilakukan.

Analisis Data

Untuk menguji H_1 dan H_2 yang diajukan dalam penelitian ini digunakan uji ANOVA *repeated measures*. Dalam penelitian ini variabel dependen yang akan diuji adalah penilaian kinerja dan alokasi kompensasi yang dilakukan berdasarkan pada format BSC dan intensitas komunikasi manajemen yang berbeda. Variabel independen dalam penelitian ini adalah format BSC dan tingkat komunikasi dalam perusahaan. Jika pendekatan ini dapat mencegah munculnya *common measures bias*, maka akan terdapat interaksi yang signifikan antara format penyajian BSC dan tingkat komunikasi yang artinya hipotesis yang diajukan dalam penelitian dapat diterima.

Untuk menguji H_3 yaitu apakah format penyajian BSC secara *disaggregated* dan tingkat komunikasi yang tinggi dapat mempengaruhi keputusan alokasi bonus maka

dilakukan analisis regresi. Sebagai variabel dependen adalah selisih besarnya bonus yang dialokasikan oleh partisipan kepada kedua divisi. Variabel independen yang digunakan adalah selisih penilaian kinerja dari kedua divisi ketika BSC disajikan secara *disaggregated* dengan tingkat komunikasi di dalam perusahaan yang tinggi.

HASIL DAN PEMBAHASAN

Data dan Pengumpulan Data

Data diperoleh dengan melakukan eksperimen pada 3 PTS di Yogyakarta. Berdasarkan eksperimen yang telah dilakukan diperoleh 183 data yang dapat digunakan untuk analisis berikutnya. Tabel 2 berikut ini adalah tabulasi perolehan data yang dilakukan melalui eksperimen.

Tabel 2. Tabulasi Hasil Eksperimen

PTS	Jumlah Kuisisioner (buah)		
	Dibagi	Tidak lengkap	Lengkap
UTY	70	15	55
UMY	100	18	82
USD	60	14	46
Jumlah	230	47	183

Beberapa kuisisioner yang tidak lengkap disebabkan adanya partisipan yang tidak menyelesaikan pengisian instrumen eksperimennya. Hasil data yang akan digunakan selanjutnya adalah data yang lengkap.

Manipulation Check

Terdapat tiga pertanyaan yang digunakan untuk cek manipulasi. Tiga pernyataan tersebut berisi pertanyaan yang berkaitan dengan desain eksperimen yang dilakukan terkait target pasar dua divisi, ukuran kinerja yang digunakan serta tingkat komunikasi dari dua divisi tersebut. Partisipan dikatakan lolos cek manipulasi jika menjawab dengan benar ketiga pertanyaan tersebut sesuai dengan manipulasi yang diterima masing-masing. Berdasarkan hasil olah data diketahui terdapat 44 partisipan yang tidak menjawab dengan benar ketiga pertanyaan tersebut sehingga dikatakan partisipan tersebut tidak lolos cek manipulasi. Berdasarkan hasil tersebut maka jumlah data yang dapat digunakan untuk analisis berikutnya adalah sebanyak 139 partisipan. Berikut ini adalah rincian dari olah data yang telah dilakukan:

Tabel 3. Hasil Tabulasi Cek Manipulasi

PTS	Kuisisioner		
	Lengkap	Gagal Manipulasi	Digunakan
UTY	55	17	38
UMY	82	14	68
USD	46	13	33
Jumlah	183	44	139

Selain itu terdapat 2 pertanyaan untuk melakukan cek manipulasi secara keseluruhan untuk memastikan apakah desain dari eksperimen ini berhasil atau tidak. Dua pertanyaan tersebut berkaitan dengan kemudahan kasus yang dikerjakan dan tingkat realistis dari kasus yang digunakan. Dua pertanyaan tersebut diukur dengan skala dari 1 (sangat tidak setuju) sampai 5 (sangat setuju sekali). Cek manipulasi dikatakan berhasil jika jawaban partisipan terhadap dua pertanyaan berbeda signifikan dengan 1 (parameter jawaban tidak setuju). Hasil *uji one sample t test* menunjukkan bahwa jawaban partisipan atas kedua pertanyaan manipulasi berbeda secara signifikan ($p < 0,000$). Hal ini menunjukkan bahwa manipulasi atau *treatment* dalam penelitian ini berhasil. Hasil pengujian dapat dilihat pada Tabel 4.

Tabel 4. Hasil Cek Manipulasi

Ket	<i>Test Value = 1</i>		
	t	Sig	Mean Difference
Pertanyaan 1	26,942	0,000	1,647
Pertanyaan 2	41,132	0,000	1,871

Statistik Deskriptif

Statistik deskriptif yang menunjukkan rata-rata penilaian kinerja kedua divisi berdasarkan manipulasi eksperimen yang telah dilakukan dapat dilihat pada Tabel 5.

Tabel 5. Statistik Deskriptif

Komunikasi	Divisi Feminim		Divisi Kasual	
	Format BSC		Format BSC	
	<i>Disaggregated</i>	<i>Aggregated</i>	<i>Disaggregated</i>	<i>Aggregated</i>
Rendah	69,97	72,55	75,24	79,42
Tinggi	70,00	71,79	75,53	74,62

Berdasarkan tabulasi pada Tabel 5 dapat diketahui bahwa terdapat perbedaan penilaian kinerja pada tiap divisi berdasarkan *treatment* penelitian yang ditetapkan. Kinerja tertinggi dari kedua divisi masing-masing terjadi ketika partisipan menerima *treatment* format BSC dengan format *aggregated* saat tingkat komunikasi manajemen rendah.

Pengujian H_1

H_1 menyatakan bahwa penyajian BSC dengan format *disaggregated* akan membuat *evaluator* menggunakan seluruh indikator yaitu *common* dan *unique measures*. Format BSC tersebut diharapkan akan mampu mengatasi *common measures bias*. H_1 diterima jika interaksi divisi dan format BSC pada ANOVA *repeated measures* menunjukkan interaksi yang signifikan. Hasil olah data yang disajikan pada Tabel 6 menunjukkan bahwa pada uji *within subject* interaksi antara divisi dengan format menunjukkan hasil yang tidak signifikan dengan nilai $F = 0,144$ dan $p = 0,705$. Dengan demikian H_{a1} dalam penelitian ini tidak dapat didukung.

Tabel 6. Hasil Uji ANOVA *Repeated Measures*

Variabel	Df	SS	MS	F	p
<i>Between Subject:</i>					
Format	1	252,23	252,23	2,22	0,139
Komunikasi	1	118,80	118,80	1,05	0,308
Format X Komunikasi	1	149,30	149,30	1,31	0,254
<i>Error</i>	135	15341,20	113,64		
<i>Within Subject:</i>					
Divisi	1	1817,02	1817,02	49,694	0,000
Divisi X Format	1	5,258	5,258	0,144	0,705
Divisi X Komunikasi	1	62,477	62,477	1,709	0,193
Divisi X Format X Komunikasi	1	80,490	80,490	2,201	0,140
<i>Error</i>	135	4936,18	36,564		

Hasil penelitian ini menunjukkan bahwa penyajian BSC secara *disaggregated* yaitu dengan menyajikan bobot untuk masing-masing indikator tidak cukup mengatasi adanya *common measures bias* dalam penggunaan BSC. Robert *et al.* (2004) menyatakan bahwa penyajian BSC secara *disaggregated* tanpa menyediakan mekanisme untuk menggabungkan atau mengombinasikan hasil penilaiannya justru akan menurunkan kualitas penilaian. Dengan menyediakan bobot pada masing-masing indikator akan membuat *evaluator* memiliki terlalu banyak informasi yang harus diolah saat menilai kinerja kedua divisi, sehingga untuk meningkatkan kualitas penilaian kinerja diperlukan strategi yang secara kognitif dapat diterapkan secara sederhana.

Pengujian H_2

H_2 dalam penelitian ini menyatakan bahwa penyajian BSC dengan format *disaggregated* akan lebih efektif membuat *evaluator* menggunakan seluruh indikator ketika intensitas komunikasi tinggi. H_2 diterima jika interaksi divisi dengan format BSC serta tingkat komunikasi menunjukkan hasil yang signifikan. Hasil pengujian *within subject* dengan ANOVA *repeated measures* pada Tabel 6 menunjukkan interaksi divisi dengan

format BSC dan tingkat komunikasi tidak signifikan nilai $F = 2,201$ dan $p = 0,140$. Hasil ini menunjukkan bahwa ketika BSC dibuat dalam format *disaggregated* dan komunikasi di dalam perusahaan yang menekankan bahwa seluruh indikator penting untuk digunakan dalam penerapan BSC tidak mampu membuat evaluator menggunakan seluruh indikator. Hal ini berarti *common measures bias* tidak dapat diatasi dengan strategi ini. Sehingga dapat disimpulkan bahwa H_2 dari penelitian ini tidak dapat didukung.

Pengujian H_3

H_3 dari penelitian ini menyatakan bahwa ketika intensitas komunikasi tinggi dan BSC disajikan secara *disaggregated* maka keputusan alokasi bonus yang dilakukan oleh evaluator terhadap manajer kedua divisi menjadi lebih baik. Untuk menguji H_3 maka dilakukan uji regresi. Variabel dependen yang digunakan adalah selisih alokasi bonus dari kedua divisi yang dilakukan oleh partisipan (evaluator). Sedangkan variabel independen yang digunakan adalah selisih penilaian kinerja yang dilakukan oleh partisipan dengan manipulasi komunikasi tinggi dan format penyajian BSC secara *disaggregated*. Variabel kontrol juga digunakan dengan menggunakan selisih penilaian kinerja dua divisi pada 3 *treatment* yang lainnya. Tabel 7 menunjukkan bahwa model keputusan alokasi bonus oleh evaluator dengan tingkat komunikasi tinggi dan BSC *disaggregated* adalah signifikan dengan nilai $F=77,722$ $p=0,000$. Model alokasi bonus dari penelitian ini menunjukkan nilai Adj R^2 69,9%.

Tabel 7. Uji Model

Source	df	Sum of Squares	Mean Square	F	Sig
Model	4	2579,036	644,759	18,568	0,000
Error	28	972,297	34,725		
Corrected Total	32	3551,333			
R ²	0,726				
Adj R ²	0,687				

Hasil uji regresi yang disajikan pada Tabel 8 menunjukkan bahwa penilaian kinerja yang dilakukan oleh partisipan yang menggunakan BSC secara *disaggregated* dan memiliki tingkat komunikasi yang tinggi berpengaruh signifikan terhadap keputusan pembagian bonus dengan nilai $t=8,816$ dan $p=0,000$. Sehingga H_3 dari penelitian ini dapat didukung.

Tabel 8. Uji Regresi

Variable	Parameter Estimates	Standar Error	t-value	Sig
Intercept	-1,866	1,760	-1,061	0,298
Dis_High	1,070	0,130	8,219	0,000
Dis_Low	-0,051	0,143	-0,358	0,723
Agg_Low	-0,081	0,128	-0,637	0,530
Agg_High	-0,035	0,119	-0,293	0,772

Keterangan:

Variabel Dependen: Selisih bonus divisi kasual dengan feminim

Variabel Independen:

- Dis_High: Selisih penilaian kinerja divisi kasual dan feminim saat BSC disajikan secara *disaggregated* dengan tingkat komunikasi tinggi
- Dis_Low: Selisih penilaian kinerja divisi kasual dan feminim saat BSC disajikan secara *disaggregated* dengan tingkat komunikasi rendah
- Agg_Low: Selisih penilaian kinerja divisi kasual dan feminim saat BSC disajikan secara *aggregated* dengan tingkat komunikasi rendah
- Agg_High: Selisih penilaian kinerja divisi kasual dan feminim saat BSC disajikan secara *aggregated* dengan tingkat komunikasi tinggi

Hasil penelitian ini mendukung hasil penelitian yang dilakukan oleh Robert, *et al.* (2004) yang menyatakan bahwa hasil penilaian kinerja evaluator yang menggunakan BSC yang disajikan secara *disaggregated* menggunakannya sebagai bagian dalam membuat pertimbangan dalam menentukan bonus.

Supplement Analysis

Hasil pengujian terhadap H_1 dan H_2 dari penelitian ini menunjukkan hasil yang tidak signifikan. Untuk menganalisis lebih lanjut dari temuan ini, penulis melakukan analisis dengan membandingkan rata-rata penilaian kinerja dari grup kontrol dengan 3 grup lainnya. *Common measures bias* terjadi jika terdapat selisih positif penilaian kinerja antara divisi kasual dan divisi feminim. Tabel 9 menunjukkan hasil olah data.

Tabel 9. Selisih Penilaian Kinerja Divisi Kasual dengan Feminim

Komunikasi	Format BSC	
	<i>Disaggregated</i>	<i>Aggregated</i>
Rendah	5,27	6,88
Tinggi	5,53	2,82

Grup kontrol dari eksperimen ini adalah saat partisipan menerima manipulasi penyajian BSC dengan format *aggregated* dengan tingkat komunikasi yang rendah. Selisih penilaian kinerja pada selain grup kontrol menunjukkan nilai yang semakin kecil. Hal ini menunjukkan adanya penurunan *common measures bias* setelah adanya *treatment* penelitian ini. Hal ini menunjukkan penyajian BSC secara *disaggregated* dan tingkat komunikasi yang tinggi dapat mengatasi munculnya *common measures bias*. Namun demikian penyajian BSC secara *disaggregated* dengan tingkat komunikasi yang tinggi tidak mampu mengatasi bias secara maksimal. Tingkat komunikasi yang tinggi dengan format BSC secara *aggregated* justru mampu mengatasi *common measures bias* secara lebih efektif dibandingkan yang lainnya.

KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini mencoba memberikan pendekatan atau metode untuk mengatasi *common measures bias* yang banyak terjadi dalam penggunaan BSC. *Common measures bias* yang muncul akan menyebabkan kualitas keputusan menjadi tidak tepat. Hasil penelitian ini adalah:

1. Penyajian BSC secara *disaggregated* tidak mampu secara signifikan mengatasi *common measures bias* yang terjadi saat evaluator menggunakan BSC.
2. Adanya komunikasi manajemen yang tinggi ketika BSC disajikan secara *disaggregated* tidak mampu secara signifikan mengatasi *common measures bias*.
3. Intensitas komunikasi yang tinggi dan penyajian BSC secara *disaggregated* berpengaruh signifikan terhadap keputusan alokasi bonus oleh evaluator kepada manajer kedua divisi.

Keterbatasan

Berdasarkan proses penelitian yang telah dilakukan diketahui terdapat beberapa keterbatasan penelitian ini, yaitu:

1. Partisipan dalam eksperimen ini adalah mahasiswa S1 yang kurang berpengalaman dalam penilaian kinerja menggunakan BSC. Walaupun semua partisipan telah mengikuti mata kuliah Akuntansi Manajemen dan Sistem Pengendalian Manajemen dan telah mempelajari BSC namun tingkat pemahaman mereka terkait BSC tidak dapat diketahui dengan pasti.
2. Partisipan dalam penelitian ini tidak terlibat dalam tahap perancangan BSC. Keterlibatan karyawan dalam perancangan dan pengembangan BSC dapat mendorong evaluator untuk menggunakan seluruh indikator.

Saran

Penelitian yang akan datang sebaiknya mempertimbangkan beberapa hal berikut ini:

1. Meningkatkan kualitas partisipan agar dapat menangkap fenomena penerapan BSC, misalnya menggunakan partisipan dari para manajer atau karyawan perusahaan.
2. Meneliti apakah keterlibatan karyawan dalam penyusunan dan pengembangan BSC dapat mengatasi adanya *common measures bias*.

REFERENSI

- Banker, R.D., Hsuihui Chang dan Mina J. Pizzini. 2004. The Balanced Scorecard: Jugmental Effects of Performance Measures Linked to Strategy. *The Accounting Review*. Vol. 79 No. 1. pp 1-24.
- Bowman, E. H. 1963. Consistency and Optimality in Managerial Decision Making. *Management Science* Vol.9 No. 3. pp. 10-321.
- Dilla. W. N. and P. J. Steinbart. 2002. The effects of alternative supplementary information display formats on judgments made using the Balanced Scorecard. *Working paper*. Iowa State University.

- Einhorn, H. I. 1972. Expert Measurement and Mechanical Combination. *Organizational Behavioral and Human Decision Processes*. Feb. Vol. 19. pp. 8-106.
- Holmstrom, B. 1979. Moral Hazard and Observability. *Bell Journal of Economics*. Vol. 10 No. 1. pp. 74-91.
- Kaplan, R. and D. Norton. 1996. *The Balanced Scorecard*. Boston, MA: Harvard Business School Press.
- Kaplan, Steven E. dan Wisner P. S. 2009. The Judgmental Effects of Management Communication and a Fifth Balanced Scorecard Category on Performance Evaluation. *Behavioral Research in Accounting*. Vol. 21 No. 2. pp. 17-56.
- Kennedy J. 1995. Debiasing the Curse of Knowledge in Audit Judgment. *The Accounting Review*. Vol. 70 No. 2. pp: 249-273.
- Libby, T., S.E. Salterio dan A. Webb. 2004. The Balanced Scorecard: The Effect of Assurance and Process Accountability on Managerial Judgment. *The Accounting Review*. Vol. 79 No. 4. pp. 1075-1094.
- Lipe, M., and S. Salterio. 2000. The balanced scorecard: Judgmental effects of common and unique performance Measures. *The Accounting Review*. Vol.75 No. 3. pp: 283- 298.
- Malina, M. A., and F. H. Selto. 2001. Communicating and controlling strategy: An Empirical Study of the Effectiveness of the Balanced Scorecard. *Journal of Management Accounting Research*. Vol. 3. pp: 47-90.
- Roberts, M. L., T. L. Albright. and A. R. Hibberts. 2002. Improving utilization of unique measures in the Balanced Scorecard: The effects of increased awareness and experience. *Working paper*. The University of Alabama.
- Roberts, M. L., T. L. Albright dan A.R. Hibberts. 2004. Debiasing Balanced Scorecard Evaluation. *Behavioral Research in Accounting*. Vol. 16. pp. 75-88.
- Slovic, P. dan D. MacPhillamy. 1974. Dimensional commensurability and cue utilization in comparative Judgment. *Organizational Behavior and Human Performance*. Vol. 11. pp: 172-194.
- Taylor, William B. 2010. The Balanced Scorecard as a Strategy-Evaluation Tool: The Effect of Implementation Involvement and a Causal Chain Focus. *The Accounting Review*. Vol. 85 No. 3. pp. 1095-1177.