

**Studi Eksperimental: Mengurangi Bias Pengukuran Umum Balanced
Scorecard dalam Penilaian Kinerja pada Mahasiswa S1 Program Akuntansi
Pajak**

Hellen Josephine Cardiana Mitan dan Juniarti
Akuntansi Bisnis Universitas Kristen Petra
Email: yunie@peter.petra.ac.id

ABSTRAK

Lipe & Salterio (2000) menemukan adanya bias pengukuran umum. Bias pengukuran umum ini mengurangi manfaat dari BSC dan berpengaruh pada keputusan manajer atas alokasi kompensasi. Penelitian selanjutnya mencoba untuk mengatasi bias namun belum optimal, yaitu Roberts, Albright & Hibbets (2004) yang menggunakan pendekatan *disaggregated/ mechanically aggregated* dan Dilla & Steinbart (2005) menggunakan pendekatan pengetahuan. Untuk itu, peneliti melakukan penelitian eksperimen dengan menggabungkan kedua pendekatan tersebut. Metode penelitian yang dilakukan adalah eksperimen laboratorium dengan menggunakan *repeated measures ANOVA, Independent Sample T Test*, dan analisis regresi linear berganda untuk menguji ketiga hipotesis. Hasil penelitian menunjukkan bahwa pendekatan *disaggregated/ mechanically aggregated* dapat mengurangi bias pengukuran umum dalam penilaian kinerja menggunakan BSC. Ditemukan pula bahwa pendekatan pengetahuan terbukti dapat mengurangi bias pengukuran umum dan penilaian kinerja menggunakan BSC terbukti mempengaruhi alokasi kompensasi.

Kata kunci: *Balanced Scorecard*, pengukuran umum, pengukuran unik, *debiasing*, bias pengukuran umum, *disaggregated/ mechanically aggregated*, pengetahuan, alokasi kompensasi.

ABSTRACT

Lipe & Salterio (2000) found a common measure bias. This common measure bias reduces the benefits of BSC and affected the manager's decision in allocating compensation. Several subsequent studies had tried to overcome the bias but not optimal, Roberts, Albright & Hibbets (2004) uses disaggregated / mechanically aggregated approach and Dilla & Steinbart (2005) use knowledge approaches. For that reason, the researcher conducted an experimental study to combine these two approaches. The methodology of the research was laboratory experiment by using repeated measures ANOVA, Independent Sample T Test, and multiple linear regression analysis to test the three hypotheses. The results showed that the disaggregated/ mechanically aggregated approach could reduce the common measure bias in the performance evaluation by using BSC. It was also found that the knowledge approach proven to reduce the common measure bias and performance evaluation by using BSC proved to affect the allocation of compensation.

Keywords: Balanced Scorecard, common measures, unique measures, debiasing, common measure bias, disaggregated/ mechanically aggregated, knowledge, allocation of compensation

PENDAHULUAN

Penggunaan *Balance Scorecard* (BSC) harus meningkatkan pengambilan keputusan manajerial dengan menyelaraskan ukuran kinerja dengan tujuan dan strategi perusahaan dan unit bisnis perusahaan (Lipe & Salterio, 2000). BSC terdiri dari pengukuran umum, yang merupakan ukuran yang sesuai dengan strategi perusahaan, dan ukuran yang unik, yang merupakan ukuran yang dibuat sesuai dengan strategi divisi (Grevinga, 2013). Hasil penelitian Lipe dan Salterio (2000) menemukan bias ukuran umum dimana terdapat bobot yang tidak proporsional pada pengukuran umum dan unik dalam penelitiannya. Partisipan eksperimen mereka, mayoritas mengandalkan ukuran umum dan mengabaikan ukuran yang unik dikarenakan partisipan sendiri kurang memahami dengan baik BSC tersebut. Bias ukuran umum ini dapat menggagalkan tujuan menerapkan BSC, yaitu untuk memperluas serangkaian pengukuran pada BSC yang digunakan manajer dalam pengambilan keputusan (Lipe & Salterio, 2000). Jika ukuran yang unik pada BSC tidak mempengaruhi keputusan manajer, perusahaan tidak akan menuai manfaat yang diharapkan dari penerapan BSC.

Pada penelitian Roberts, Albright & Hibbets (2004), menggunakan pendekatan *disaggregated/mechanically aggregated* yang merupakan format pengukuran kinerja BSC secara terpisah, sehingga mengurangi adanya bias ukuran umum, namun belum optimal.

Selanjutnya Dilla & Steinbart (2005) meneruskan penelitian sebelumnya oleh Lipe & Salterio (2000), namun menambahkan pelatihan mengenai BSC terhadap pesertanya karena pelatihan merupakan suatu cara untuk memperoleh pengetahuan. Dengan pengetahuan yang lebih banyak akan menghasilkan penilaian yang lebih baik (Bonner 1990; Bonner and Lewis 1990). Penelitian ini cukup berhasil mengurangi bias ukuran umum, meskipun juga belum optimal.

Dengan hasil dua penelitian sebelumnya yang belum efektif mengurangi bias ukuran umum, peneliti kali ini mencoba mengkombinasikan pendekatan yang tidak dilakukan oleh kedua peneliti sebelumnya yaitu pendekatan *disaggregated/mechanically aggregated* dan pendekatan pengetahuan pada partisipan. Diharapkan dengan penggabungan ini, hasilnya menjadi lebih baik, bobot antara ukuran umum dan unik dapat lebih seimbang

sehingga BSC dapat bermanfaat dengan optimal bagi perusahaan.

Konsep Balanced Scorecard

Balanced Scorecard (BSC) adalah alat untuk mengukur kinerja perusahaan secara komprehensif yang menggabungkan perspektif keuangan dan non keuangan, yang meliputi perspektif pelanggan, proses bisnis internal, dan pembelajaran dan pertumbuhan (Kaplan & Norton, 1992).

Pengukuran Umum dan Unik

BSC digunakan untuk mengevaluasi kinerja dalam hal ini termasuk ukuran baik umum dan unik yang diselenggarakan oleh atribut keuangan dan non keuangan (Dilla & Steinbart, 2005). Menurut Kaskey (2008) pengukuran umum adalah pengukuran kinerja baik finansial maupun non finansial yang berlaku untuk seluruh lapisan perusahaan dan pada umumnya merupakan ukuran yang diterapkan di seluruh unit dalam perusahaan. Dilla dan Steinbart (2005) juga menyatakan setiap unit bisnis biasanya harus mencakup ukuran-ukuran yang umum di seluruh unit usaha dan ukuran umum ini merupakan ukuran yang sesuai dengan strategi organisasi (Grevinga, 2013).

Kaplan & Norton (1996) berpendapat bahwa semua BSC cenderung menggunakan ukuran umum, namun setiap unit bisnis dalam organisasi akan memiliki *scorecard* sendiri yang khusus dirancang untuk unit tersebut dan juga turut menjadi kekuatan penting dari BSC. Ukuran-ukuran spesifik yang dipilih untuk masing masing unit bisnis ini akan berbeda dari unit bisnis lainnya, karena ukurannya harus sesuai dengan strategi dan tujuan spesifik masing-masing unit (Lipe & Salterio, 2000). Ukuran pada *scorecard* unit bisnis menurut Kaplan & Norton (1993) seharusnya dirancang agar sesuai dengan “misi, strategi, teknologi, dan budaya” masing-masing unit tersebut. Ukuran spesifik ini yang disebut sebagai ukuran yang unik.

Bias Pengukuran Umum

Bias pengukuran umum sering dijelaskan sebagai ketidakmampuan pengambil keputusan untuk menyertakan informasi unik dalam evaluasi kinerja, karena informasi ini membutuhkan usaha kognitif yang lebih untuk bisa diproses (Lipe & Salterio, 2000; Grevinga, 2013). Hal ini dikarenakan individu memiliki keterbatasan memori dalam menyerap atau memproses informasi (Tversky & Kahneman, 1973, 1974)

yang juga merupakan asumsi *bounded rationality* yaitu individu memiliki keterbatasan informasi, waktu, kapasitas memori, pengetahuan terbatas dan sebagainya (Wahyuni & Hartono, 2014; Miqdad, 2012).

Dalam penelitiannya, Lipe & Salterio (2000) menemukan bahwa partisipannya mengabaikan ukuran unik karena merasa ukuran unik lebih kompleks dan sulit dipahami dan bias ukuran umum timbul akibat evaluator menggunakan sebagian besar ukuran umum BSC pada kedua unit bisnis dan mengabaikan ukuran unik di masing-masing unit bisnis. Terjadinya bias dapat mempengaruhi keputusan para *decision maker* pada saat evaluasi untuk menentukan kompensasi/bonus bagi para manajer divisi (Ittner, Larcker & Meyer, 2003) sehingga menghasilkan penilaian yang tidak adil bagi karyawan. Hal ini akan berdampak terhadap rasa keadilan karyawan karena hasil kinerja mereka tidak dinilai secara komprehensif (Liedtka, Church & Ray, 2008). Dan menurut Gibbs, Merchant, Van der Stede & Vargus (2004), kurangnya sistem *reward* yang adil dan transparan juga dapat mengurangi semangat karyawan untuk memberikan usaha lebih pada perusahaan serta menciptakan ketidakpuasan pada BSC (Dilla & Steinbart, 2005), yang akhirnya akan membuat pengukuran BSC menjadi sia-sia.

Mengurangi Bias Pengukuran Umum (Debiasing)

Dengan mengetahui seputar mengenai bias ukuran umum dan dampak yang dihasilkan atas terjadinya bias tersebut yang kemudian mengarahkan peneliti untuk melakukan teknik *debiasing*. Bazerman (2006) mendefinisikan bahwa *debiasing* adalah prosedur untuk mengurangi atau menghilangkan bias dari strategi kognitif pembuat keputusan, sehingga untuk membuatnya berhasil memerlukan banyak waktu dan usaha (Serfas, 2011). Dan menurut Kennedy (1995) untuk meningkatkan kualitas penilaian ketika usaha yang diberikan tidak cukup adalah dengan menggunakan alat bantu keputusan (dalam Roberts, Albright & Hibbets, 2004). Dan pada penelitian eksperimen ini, peneliti menggunakan alat *disaggregated/mechanically aggregated* dan pengetahuan untuk mengurangi bias.

Disaggregated/mechanically aggregated sendiri merupakan alat untuk mengurangi bias dengan menyederhanakan informasi yang perlu diproses. Contoh dari penerapan *disaggregated/mechanically aggregated* ini

terdapat pada penelitian Roberts, Albright & Hibbets (2004) dimana para partisipan melakukan *disaggregated* dengan mengevaluasi kinerja secara terpisah atas 16 ukuran kinerja BSC dan kemudian *mechanically aggregated*, yaitu mengumpulkan secara mekanis penilaian yang terpisah tersebut dengan menggunakan bobot yang telah ditetapkan sebelumnya untuk masing-masing ukuran. Setelah keduanya dilaksanakan, partisipan kemudian membuat evaluasi keseluruhan. Karena melakukan penilaian yang terpisah dan diukur satu persatu tiap ukuran, penilai menjadi lebih focus dalam melakukan penelitian kinerja, karena memperhatikan baik pengukuran umum maupun pengukuran unik, sehingga dapat mengurangi bias pengukuran umum.

Selain menggunakan *disaggregated/mechanically aggregated*, pengetahuan juga merupakan teknik untuk mengurangi bias. Penelitian penilaian audit menunjukkan bahwa pengetahuan yang lebih luas menghasilkan penilaian yang lebih baik di berbagai macam tugas audit (Bonner 1990; Bonner & Lewis 1990). Tubbs (1992) juga menyatakan bahwa pengetahuan dapat membantu dalam penilaian yang lebih efektif dari auditor untuk tugas-tugas yang mereka miliki sesuai dengan tingkat pengetahuan. Umpan balik kognitif seperti pengetahuan ini dapat secara efektif meningkatkan kualitas penilaian dan pengambilan keputusan (Bonner, 2008).

Pengukuran BSC dan Alokasi Kompensasi

Kaplan dan Norton (1996, 2001) menyarankan untuk menghubungkan BSC dengan kompensasi untuk meningkatkan manfaat BSC tersebut. Manfaat dari mengadopsi BSC ini tergantung pada seberapa besar BSC memberikan kemajuan terhadap pengambilan keputusan manajer, yang salah satunya yaitu keputusan kompensasi (Lipe & Salterio, 2000).

Jika dikaitkan dengan alokasi kompensasi, timbulnya bias ukuran umum akibat pemberian bobot yang berlebihan pada ukuran umum dibanding dengan ukuran unik (Lipe & Salterio, 2000; Banker, Chang & Pizzini, 2004; Ittner, Larcker & Meyer, 2003) dapat melemahkan tujuan BSC, yang secara jelas dirancang untuk menggabungkan ukuran-ukuran tersebut ke dalam pemikiran manajerial dan pengambilan keputusan (Kaplan & Norton, 1996), yang salah satunya adalah keputusan kompensasi.

BSC digunakan sebagai alat untuk mengevaluasi kinerja dan membantu mengalokasikan jumlah kompensasi (Banker, Chang & Pizzini, 2004; Roberts, Albright & Hibbets, 2004; Dilla & Steinbart, 2005). Karena BSC merupakan alat untuk menilai kinerja, jadi penilaian kinerja perlu dikaitkan dengan keputusan kompensasi. Jika tidak dikaitkan, motivasi karyawan akan berkurang dalam bekerja.

Holmstrom dan Milgrom (1991) mengungkapkan bahwa keputusan agen dalam hal ini manajemen dipengaruhi oleh hal-hal yang dimasukkan ke dalam evaluasi kinerja dan kompensasi, sehingga mengabaikan ukuran unik tiap unit bisnis dalam pengukuran kinerja dan akan mengakibatkan *decision maker* mengambil keputusan yang tidak tepat, yang salah satunya adalah keputusan alokasi kompensasi. Berkenaan dengan hal ini, dibutuhkan penilaian yang tepat dan adil atas kinerja karyawan. Karena jika penilaian kinerja tidak tepat, maka keputusan kompensasi juga tidak sesuai dengan performa manajer yang sebenarnya dan ini akan berdampak terhadap rasa keadilan karyawan karena hasil kinerja mereka tidak dinilai secara komprehensif (Liedtka, Church & Ray, 2008).

Evaluasi kinerja yang tidak konsisten juga akan berpengaruh pada kinerja karyawan di perusahaan, dimana karyawan merasa tidak adil dan menyebabkan berkurangnya semangat untuk memberikan usaha lebih pada perusahaan (Gibbs, Merchant, Van der Stede, Vargus, 2004) karena aktivitas yang mereka lakukan dalam pekerjaan dianggap tidak akan mempengaruhi hasil evaluasi kinerjanya dan kompensasi yang akan di terima (Holmstrom & Milgrom, 1991; Malina & Selto, 2001).

2.7 Hipotesis Penelitian

Untuk mendapatkan hasil empirik dari penelitian ini, hipotesis yang ada perlu ditetapkan dan diuji. Hipotesis disimpulkan berdasarkan landasan teori dan dari kajian penelitian terdahulu.

H₁: Penilaian kinerja menggunakan pendekatan *dissaggregated/mechanically aggregated* akan mengurangi bias ukuran umum.

H₂: Pendekatan pengetahuan yang memadai tentang BSC akan mengurangi bias ukuran umum.

H₃: Evaluasi kinerja menggunakan BSC berpengaruh terhadap alokasi kompensasi

METODE PENELITIAN

Model Eksperimen

Model eksperimen yang digunakan oleh peneliti adalah eksperimen laboratorium. Eksperimen laboratorium merupakan penelitian yang menguji hubungan sebab akibat pada lingkungan yang artifisial/buatan dalam eksperimen laboratorium dan keterlibatan dari peneliti sangat tinggi. Peneliti terlibat dalam pembuatan *setting* buatan dan melakukan manipulasi terhadap variable tertentu (Indriantoro & Supomo, 1999).

Model eksperimen ini digunakan karena peneliti terlibat langsung pada saat penelitian dilakukan sehingga memiliki kontrol pada eksperimen tersebut.

Partisipan

Partisipan pada eksperimen ini adalah 40 mahasiswa S1 Fakultas Ekonomi Universitas Kristen Petra yang telah mengambil mata kuliah Akuntansi Manajemen, sehingga telah memiliki pengetahuan awal mengenai struktur dan konsep BSC.

Desain Eksperimen

Peneliti mengadaptasi kasus dari peneliti sebelumnya yaitu dari Lipe dan Salterio (2000), Roberts, Albright & Hibbets (2004) dan Dilla & Steinbart (2005). Kasus ini melibatkan dua divisi WCS Incorporated yaitu divisi RadWear dan divisi WorkWear. WCS Incorporated merupakan perusahaan spesialisasi busana wanita. Pada penelitian eksperimen ini menggunakan desain 2 x 2 x 2, dengan dua *between-subject* (umum dan unik), dan satu *within-subject* (divisi).

Pertama-tama yang dilakukan adalah mengumpulkan partisipan di dalam sebuah ruangan yang menjadi lokasi eksperimen. Lalu partisipan dibagi menjadi dua kelompok. Dimana 50 persen dari jumlah partisipan diminta untuk mengikuti penjelasan mengenai BSC. Setelah selesai diberikan penjelasan, 50 persen partisipan lainnya dipersilahkan masuk kedalam ruangan. Dan setelah seluruh partisipan berkumpul, partisipan diminta untuk berperan sebagai senior eksekutif yang akan menilai kinerja dua divisi dari WCS Incorporation, yaitu divisi RadWear yang merupakan divisi pakaian

remaja perempuan dan WorkWear yang merupakan divisi seragam kerja wanita. Setelah itu partisipan diberikan kasus yang berisi misi perusahaan, visi dan strategi dari divisi RadWear dan WorkWear serta instruksi untuk mengerjakan kasus. Peneliti kemudian akan memberikan penjelasan tentang latar belakang perusahaan dan memberikan waktu 15-20 menit kepada partisipan untuk membaca instruksi dan kasus yang diberikan.

Langkah selanjutnya, partisipan diberi *scorecard* divisi RadWear yang berisi 16 ukuran kinerja BSC yang terbagi pada empat perspektif. Di masing-masing perspektif terdapat dua ukuran umum dan 2 ukuran unik. Total berat untuk masing-masing dari empat kategori yang ditetapkan sebesar 25 persen, dan dalam setiap kategori *pre-assigned weight* ditetapkan dengan variasi yaitu antara 4,0 dan 9,0 persen. Ukuran unik ditetapkan sebesar 64 persen dari total beban.

Berikutnya partisipan mulai menilai kinerja manajer divisi RadWear dengan menyelesaikan dua langkah dari *disaggregated* BSC,

1. Menilai kinerja manajer pada 16 item individual BSC dengan menggunakan skala dari 0 (*Unacceptable*) sampai 100 (*Excellent*).
2. Mengalikan hasil penilaian kinerja manajer RadWear dengan *pre-determined weight*.

Setelah itu, partisipan menyelesaikan langkah *mechanically aggregated* yaitu dengan menjumlahkan hasil perkalian tersebut untuk memperoleh nilai total agregat untuk divisi RadWear.

Berbeda dengan Roberts et al (2004), peneliti tidak menduplikasi *separate overall judgment*, yang merupakan penilaian keseluruhan terpisah yang dilakukan oleh partisipan ketika mereka merasa tidak puas dengan skor *mechanically aggregated* dengan alasan apapun. Hal ini karena penilaian yang dilakukan secara terpisah dengan tidak berdasarkan hasil perhitungan melainkan melalui perasaan ketidakpuasan pribadi masing-masing individu dalam penilaian kinerja berpotensi menimbulkan bias.

Selanjutnya, partisipan mengumpulkan penilaian kinerja divisi RadWear kepada peneliti, dan peneliti kemudian memberikan *scorecard* divisi WorkWear kepada partisipan. Prosedur yang dilakukan untuk menilai kinerja WorkWear

sama seperti ketika menilai kinerja divisi RadWear sebelumnya.

Seusai menilai kinerja dari kedua divisi, peneliti mengembalikan *scorecard* RadWear kepada partisipan. Partisipan kemudian diminta untuk mengalokasikan bonus akhir tahun sebesar Rp 100.000.000 untuk masing-masing manajer divisi berdasarkan skor total agregat penilaian kinerja kedua manajer tersebut. Kemudian mereka menyelesaikan pertanyaan lanjutan tentang kasus ini dengan memberikan informasi demografis (Roberts, Albright & Hibbets, 2004).

Teknik Analisis

Penelitian eksperimen ini memiliki 3 hipotesis yang diuji dengan 3 teknik analisis berbeda. Untuk menguji apakah pendekatan *Disaggregated/Mechanically Aggregated* berpengaruh dalam mengurangi bias pengukuran umum (hipotesis 1) digunakan uji *repeated measures ANOVA (Analysis of Variance)* dengan melihat interaksi antara *between-subject factor* (pengukuran umum dan unik) dan *within-subject factor* (divisi). Uji *repeated measures ANOVA* dilakukan sebanyak tiga kali yaitu pada kelompok partisipan yang mendapat pengetahuan mengenai BSC dan menggunakan pendekatan *disaggregated/mechanically aggregated*; pada kelompok partisipan yang tidak mendapat pengetahuan mengenai BSC dan menggunakan pendekatan *disaggregated/mechanically aggregated*; dan pada kedua kelompok partisipan secara total dan menggunakan pendekatan *disaggregated/mechanically aggregated*.

Untuk menguji apakah pendekatan pengetahuan berpengaruh dalam mengurangi bias pengukuran umum digunakan uji *Independent Sample T-Test* yang membandingkan antara hasil evaluasi dari kelompok pengetahuan dan kelompok non pengetahuan.

Dan untuk menguji hipotesis ketiga menggunakan teknik analisis regresi linear berganda. Penggunaan regresi linear berganda dikarenakan teknik analisis data ini bertujuan untuk menguji pengaruh dua atau lebih variabel independen (*explanatory*) terhadap satu variabel dependen. Umumnya regresi linear berganda dinyatakan dengan persamaan:

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_nX_n + \varepsilon$$

..... (3.1)

dimana:

- Y *dependent variable*
- X₁ - X_n *independent variable*
- β₀ koefisien konstanta
- ε *Error term*
- β₁ - β_n koefisien parameter

Berikut merupakan tahapan teknik analisis data dalam penelitian ini.

- a. Menyiapkan data-data yang dibutuhkan agar menjadi data yang siap diproses. Tahapan ini merupakan tahapan menilai kelengkapan data, membersihkan data, dan menyajikan asumsi deskriptif.
- b. Menjalankan pengujian untuk ketiga hipotesis penelitian.
- c. Uji Asumsi Klasik
 Pengujian asumsi klasik ini bertujuan untuk mengetahui dan menguji kelayakan atas model regresi yang digunakan dalam penelitian ini.
 - Uji Normalitas
 - Uji Heteroskedastisitas
 - Uji Multikolinieritas

- d. Koefisien Determinasi (R²)
 Untuk mengukur seberapa jauh kemampuan model dalam menerangkan variasi variabel dependen, maka diperlukannya koefisien determinasi (R²). Nilai koefisien determinasi adalah antara nol dan satu.

Analisis Deskriptif

Berikut ini disajikan nilai rata-rata dan standar deviasi dari data penelitian yaitu data pengukuran kinerja manajer pada Divisi Rad Wear dan Work Wear dengan menggunakan pengukuran umum dan unik:

Tabel 1. Statistik Deskriptif

Descriptive Statistics					
	N	Minimum	Maximum	Mean	Std. Deviation
Rc	32	10.20	27.10	21.4281	4.13064
Ru	32	22.36	48.00	38.5753	6.34469
Wc	32	9.06	24.40	15.9834	3.65182
Wu	32	14.96	51.40	41.9044	7.99483
Valid N (listwise)	32				

Sumber: Data Diolah (2014)

Berdasarkan tabel di atas dapat diketahui bahwa *minimum* merupakan nilai terendah dari keseluruhan data dimana berarti dari data Rc, Ru, Wc, dan Wu masing-masing terdapat nilai minimum sebesar 10.20, 22.36, 9.06, dan 14.96. Sedangkan untuk *maximum* merupakan nilai tertinggi dari keseluruhan data dimana berarti dari data Rc, Ru, Wc, dan Wu masing-masing terdapat nilai maksimum 27.1, 48.0, 24.4, dan 51.4.

Mean merupakan nilai rata-rata dari keseluruhan data, yang berarti rata-rata nilai pengukuran umum pada divisi Rad Wear adalah sebesar 21.428, sedangkan dengan pengukuran unik pada divisi ini didapatkan nilai rata-rata sebesar 38.575. Demikian pula halnya dengan divisi yang lain yaitu Work Wear, nilai rata-rata pengukuran umumnya yaitu sebesar 15.98 sedangkan dengan pengukuran unik didapatkan rata-rata sebesar 41.90. Nilai standart deviasi menunjukkan tingkat keragaman data penelitian. Berdasarkan data pada tabel di atas dapat diketahui bahwa secara umum nilai standar deviasi nilai Rc, Ru, Wc dan Wu relatif lebih kecil jika dibandingkan dengan rata-ratanya, hal ini menunjukkan kecilnya variasi dari data penelitian atau dapat disebut dengan data homogen.

Hasil Repeated measures ANOVA

Tabel 4.8. Repeated Measure ANOVA - Pengetahuan

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	28590.383	1	28590.383	1595.558	.000
Common	64.298	1	64.298	3.588	.063
Unique	201.001	1	201.001	11.217	.001
Common * Unique	3568.435	1	3568.435	199.145	.000
Error	1075.124	60	17.919		

a. Kode = Pengetahuan

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	division	Type III Sum of Squares	df	Mean Square	F	Sig.
division	Linear	201.001	1	201.001	11.217	.001
division * Common	Linear	3568.435	1	3568.435	199.145	.000
division * Unique	Linear	28590.383	1	28590.383	1595.558	.000
division * Common * Unique	Linear	64.298	1	64.298	3.588	.063
Error(division)	Linear	1075.124	60	17.919		

a. Kode = Pengetahuan

Sumber: Data Olahan SPSS (2014)

Berdasarkan tabel *between-subjects effect* dapat diketahui bahwa interaksi dari pengukuran umum dan unik menunjukkan signifikansi dibawah 0.05

dengan F hitung 199.145. Dengan pengukuran umum menggunakan asumsi signifikansi dibawah 0.10 yaitu sebesar 0.063, dan pengukuran unik dengan nilai signifikansi dibawah .05 yaitu 0.01. Lebih lanjut pada tabel *within-subjects effect* dapat diketahui bahwa terdapat F hitung 199.145 dengan nilai signifikansi dibawah 0.05 pada interaksi pengukuran umum dan divisi, demikian pula halnya pada interaksi antara pengukuran unik dan divisi terdapat F hitung 1595.558 dengan signifikansi dibawah 0.05. Selanjutnya, ketika diinteraksikan bersamaan antara divisi dengan umum dan unik, nilai signifikansi sebesar 0.063 dengan asumsi syarat signifikansi untuk interaksi dibawah 0.10.

Tabel 4.9. Hasil Repeated Measure ANOVA – Non Pengetahuan

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	27014.177	1	27014.177	2289.220	.000
Common	282.150	1	282.150	23.910	.000
Unique	67.135	1	67.135	5.689	.020
Common * Unique	3853.762	1	3853.762	326.573	.000
Error	708.036	60	11.801		

a. Kode = Non Pengetahuan

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	division	Type III Sum of Squares	df	Mean Square	F	Sig.
division	Linear	67.135	1	67.135	5.689	.020
division * Common	Linear	3853.762	1	3853.762	326.573	.000
division * Unique	Linear	27014.177	1	27014.177	2289.220	.000
division * Common * Unique	Linear	282.150	1	282.150	23.910	.000
Error(division)	Linear	708.036	60	11.801		

a. Kode = Non Pengetahuan

Sumber: Data Olahan SPSS (2014)

Berdasarkan tabel *between-subjects effect* dapat diketahui bahwa interaksi dari pengukuran umum dan unik menunjukkan nilai signifikansi dibawah 0.05 dengan F hitung masing-masing 23.910 dan 5.689. Lebih lanjut pada tabel *within-subjects effect* dapat diketahui bahwa terdapat nilai signifikansi antara interaksi divisi dengan pengukuran umum, divisi dengan pengukuran unik dan interaksi ketiganya secara bersamaan yang masing-masing dibawah 0.05.

Tabel 4.10. Hasil Repeated Measures ANOVA Keseluruhan

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	55216.832	1	55216.832	4025.997	.000
Common	335.405	1	335.405	24.455	.000
Unique	7.711	1	7.711	.562	.455
Common * Unique	7444.142	1	7444.142	542.771	.000
Error	1645.808	120	13.715		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	division	Type III Sum of Squares	df	Mean Square	F	Sig.
division	Linear	7.711	1	7.711	.562	.455
division * Common	Linear	7444.142	1	7444.142	542.771	.000
division * Unique	Linear	55216.832	1	55216.832	4025.997	.000
division * Common * Unique	Linear	335.405	1	335.405	24.455	.000
Error(division)	Linear	1645.808	120	13.715		

Sumber: Data Olahan SPSS (2014)

Berdasarkan tabel *between-subjects effect* bahwa interaksi dari pengukuran umum dan unik menunjukkan nilai signifikansi dibawah 0.05 dengan F hitung 542.771. Namun pengukuran uniknya sendiri menunjukkan nilai signifikansi sebesar 0.455 dengan F hitung 0.562. Lebih lanjut pada tabel *within-subjects effect* dapat diketahui bahwa terdapat nilai signifikansi antara interaksi divisi dengan pengukuran umum, divisi dengan secara bersamaan yang masing-masing dibawah 0.05.

Hasil Independent Sampel T-Test

Jika *independent sample t test* menghasilkan nilai signifikansi < 0.05 ($\alpha=5\%$), maka dapat disimpulkan bahwa terdapat perbedaan yang signifikan antara evaluasi pengukuran dari partisipan dengan pengetahuan BSC dan tanpa pengetahuan BSC. Berikut adalah hasil *independent sample t test with equal variances assumed* data evaluasi pengukuran umum dan unik pada kelompok partisipan dengan pengetahuan BSC dan tanpa pengetahuan BSC:

Tabel 4.13. Hasil Independent sample T-Test

Group Statistics

Kode	N	Mean	Std. Deviation	Std. Error Mean	
Common	Pengetahuan	15	39.4867	5.93581	1.53262
	Non Pengetahuan	16	36.1619	6.06163	1.51541
Unique	Pengetahuan	15	83.2460	8.35682	2.15772
	Non Pengetahuan	16	80.0581	8.85954	2.21488

Independent Samples Test										
		Levene's Test for Equality of Variances		t-Test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Common	Equal variances assumed	.104	.750	1.542	29	.134	3.32479	2.15682	-1.08641	7.73599
	Equal variances not assumed			1.543	28.939	.134	3.32479	2.15531	-1.08372	7.73331
Unique	Equal variances assumed	.000	.988	1.029	29	.312	3.18788	3.09819	-3.14864	9.52439
	Equal variances not assumed			1.031	28.998	.311	3.18788	3.09216	-3.13633	9.51208

Sumber: Data Olahan SPSS (2014)

Berdasarkan Tabel di atas diketahui bahwa rata-rata evaluasi kinerja manajer dengan pengukuran umum oleh kelompok partisipan pengetahuan adalah sebesar 39.487 sedangkan oleh partisipan tanpa pengetahuan BSC sebesar 36.162. Rata-rata nilai evaluasi kinerja dengan pengukuran unik sebesar 83.246 oleh kelompok partisipan pengetahuan dan 80.058 oleh partisipan non pengetahuan.

Uji beda nilai evaluasi dengan pengukuran umum dan unik menghasilkan nilai signifikansi masing-masing sebesar 0.134 dan 0.312 dimana nilai ini lebih besar dari 0.05 ($\alpha = 5\%$).

4.5.4 Hasil Analisis Regresi

Analisis regresi dilakukan dengan variabel bebas yaitu *Aggregated Score* (AggScDiff) serta variabel terikat yaitu alokasi bonus (kompensasi). Berikut ini adalah hasil analisis regresi selengkapannya:

Tabel 4.17 Hasil Uji F Regresi ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.	
1	Regr ession	2.011	4	.503	14.636	.000 ^b
	Residual	.927	27	.034		
	Total	2.938	31			

- a. Dependent Variable: Y_rasio
 - b. Predictors: (Constant), Wu, Ru, Wc, Rc
- Sumber: Data Olahan SPSS (2014)

Dari tabel diatas diketahui bahwa uji F menghasilkan nilai signifikansi < 0.05, maka disimpulkan bahwa pengukuran umum dan unik di kedua divisi Rad Wear dan Work Wear secara simultan berpengaruh signifikan terhadap alokasi kompensasi.

Berdasarkan hasil analisis didapatkan nilai R Square sebesar 0.684 yang berarti

bahwa pengaruh evaluasi BSC melalui pendekatan *Aggregated Score* (AggScDiff) dapat mempengaruhi besarnya alokasi kompensasi sebesar 68,4%, sedangkan sisanya yaitu 31,6% dipengaruhi oleh faktor lain.

Tabel 4.18 Nilai R Square

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.827 ^a	.684	.638	.18532

- a. Predictors: (Constant), Wu, Ru, Wc, Rc
- b. Dependent Variable: Y_rasio

Tabel 4.19 Hasil Analisis Regresi

Coefficients ^a								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	.578	.283		2.046	.051		
	Rc	.001	.021	.013	.048	.962	.145	6.896
	Ru	.038	.014	.793	2.761	.010	.142	7.063
	Wc	-.038	.012	-.448	-3.200	.004	.597	1.676
	Wu	-.008	.006	-.216	-1.451	.158	.529	1.889

a. Dependent Variable: Y_rasio

Sumber: Data Olahan SPSS (2014)

Berdasarkan tabel di atas dapat diketahui bahwa pengaruh Rc, Ru, Wc, dan Wu masing-masing menghasilkan nilai signifikansi 0.962, 0.1, 0.004, dan 0.158.

Kesimpulan

Berdasarkan hasil pengujian yang telah dilakukan, maka beberapa kesimpulan yang dapat diambil adalah:

1. Hipotesis 1 pada penelitian ini diterima, hal ini berarti pendekatan *disaggregated/mechanically aggregated* akan mengurangi bias pengukuran umum.
2. Hipotesis 2 pada penelitian ini diterima, hal ini berarti pendekatan pengetahuan akan mengurangi bias pengukuran umum.
3. Hipotesis 3 pada penelitian ini diterima, hal ini berarti penilaian

kinerja menggunakan BSC mempengaruhi alokasi kompensasi.

Keterbatasan dan Saran

Keterbatasan dalam penelitian ini adalah sebagai berikut:

1. Partisipan eksperimen merupakan mahasiswa Jurusan Akuntansi Pajak Fakultas Ekonomi Universitas Kristen Petra Surabaya. Partisipan dalam eksperimen ini yaitu partisipan yang minimal mengambil mata kuliah akuntansi manajemen karena terdapat materi tentang BSC didalamnya dan partisipan belum memiliki pengalaman membuat dan menggunakan BSC, sehingga pengetahuan partisipan hanya terbatas pada teori dan penerapan secara teoritis.
2. Penelitian ini tidak dapat mengukur seberapa besar pengetahuan tentang BSC yang dimiliki oleh kelompok partisipan yang diberikan pengetahuan mengenai BSC, sehingga tidak dapat memastikan partisipan benar-benar memahami penerapan BSC atau tidak.
3. Penelitian ini belum menggunakan pendekatan-pendekatan lain yang dapat mengurangi bias pengukuran umum dalam penilaian kinerja menggunakan BSC.

Saran yang dapat diberikan terkait penelitian ini adalah sebagai berikut:

1. Dalam penelitian selanjutnya, diharapkan peneliti dapat menggunakan partisipan yang sudah berpengalaman menggunakan BSC dalam penilaian kinerja perusahaan, Peneliti selanjutnya diharapkan dapat mengukur pengetahuan partisipan dengan memberikan pertanyaan-pertanyaan terkait teori dan penerapan BSC.
2. Untuk penelitian yang akan datang, sebaiknya menggunakan pendekatan-pendekatan lain, selain pendekatan *disaggregated/mechanically aggregated* dan pendekatan pengetahuan

DAFTAR REFERENSI

Bazerman, M. (2006). *Judgment in Managerial Decision Making* (6th.ed.). New Jersey: Wiley.
Banker, D. R., Chang, H., & Pizzini, M. (2004). The Balanced Scorecard: Judgmental Effects of

Performance Measures Linked to Strategy. *The Accounting Review*, 79(1), 1-23.

Bonner, S.E. (1990). Experience Effects in Auditing: The Role of Tasks Specific Knowledge. *The Accounting Review*, 65 (1), 72-92.

Bonner, S. E., & Lewis, B.L. (1990). Determinants of auditor expertise. *Journal of Accounting research* (Supplement), 1-20

Bowman, E. H. (1963). Consistency and Optimality in Managerial Decision Making. *Management Science* 9 (1), 310-321.

Debusk, G.K., Killough, L.N., & Brown, R.M. (2005). Intolerance of Ambiguity with the Balanced Scorecard. *The Innovation Congress: Issues in Innovation*, vol. 3, issues 1.

Dilla, W. N., & Steinbart, P. J. (2005a). Relative Weighting of Common and Unique Measures by Knowledgeable Decision Makers. *Behavioral Research in Accounting*, 17(1), 43-53.

Gagne, M. L., Hollister, J., & Tully, G. J. (2006). Using the Balanced Scorecard: Both Common and Unique Measures are Informative. *Journal of Applied Business Research*, 22(1), 147-160.

Gibbs, M. J, Merchant, K. A., Wim A Van der Stede, & Vargus, M. E. (2005). The Benefits of Evaluating Performance Subjectively. *Performance Improvement*, 44 (5), 26-32

Grevinga, K. H. M. (2013). Common Measure Bias in the Balanced Scorecard: an Experiment with Undergraduate Students. *Master's Degree Thesis, University of Twente, Enschede*.

Indriantoro, N., & Supomo, B. (1999, Oktober). Metodologi Penelitian Bisnis untuk Akuntansi dan Manajemen. BPFE-Yogyakarta, Edisi Pertama.

Ittner, C. D., Larcker, D. F., & Meyer, M. W. (2003). Subjectivity and the Weighting of Performance Measures: Evidence from a Balanced Scorecard. *The Accounting Review*, 78(3), 725-758.

Kaplan, R. S., & Norton, D. P. (1992). The Balanced Scorecard - Measures that Drive Performance. *Harvard Business Review*, 70(1), 71-79.

Kaplan, R. S., & Norton, D. P. (1996). Linking the Balanced Scorecard to Strategy. *California Management Review*, 39(1), 53-79.

Kaplan, R. S., & Norton, D. P. (2001a). The Strategy - Focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment. *Harvard Business School Press*.

- Kaplan, R.S., & Norton, D.P. (2001a). Transforming the Balanced Scorecard from Performance Measurement to Strategic Management (Part I). *Accounting Horizons*, 15(1), 87-104.
- Kaplan, R.S., & Norton, D.P. (2001c). Transforming the Balanced Scorecard from Performance Measurement to Strategic Management (Part II). *Accounting Horizons*, 15 (2), 147-160
- Kaskey, V. L. (2008). The Balanced Scorecard: A Comparative Study of Accounting Education and Experience on Common Measure Bias and Trust in a Balanced Scorecard. *Dissertation. School of Business and Technology, Capella University.*
- Lawrie, G., & Cobbold, I. (2004). Third-Generation Balanced Scorecard Evolution of an effective Strategic Control Tool. *International Journal of Productivity and Performance Management*, 53 (7), 611-623.
- Libby, T., Salterio, S., & Webb, A. (2002). The Balanced Scorecard: The Effect of Assurance and Process Accountability on Managerial Judgment.
- Libby, T., Salterio, S. E., & Webb, A. (2004). The Balanced Scorecard: The Effects of Assurance and Process Accountability on Managerial Judgment. *The Accounting Review*, 79(4), 1075-1094.
- Liedtka, S. L., Church, B. K., & Ray, M. R. (2008). Performance Variability, Ambiguity Intolerance, and Balanced Scorecard-Based Performance Assessments. *Behavioral Research in Accounting*, 20(2), 73-88.
- Lipe, M. G., & Salterio, S. E. (2000). The Balanced Scorecard: Judgmental Effects of Common and Unique Performance Measures. *The Accounting Review*, 75(3), 283-298.
- Malina, M. A., & F. H. Selto. (2001). Communicating and controlling strategy. An empirical study for the effectiveness of the balanced scorecard. *Journal Management Accounting Research* 13: 47-90.
- Marr, B., & Neely, A. (2003). Automating the Balanced Scorecard – selection criteria to identify appropriate software application. *Measuring Business Excellence*, 7(3), 29-36.
- Miqdad, M. (2012, September). Praktik Tata Kelola perusahaan (*Corporate Governance*) dan *Usefulness* Informasi Akuntansi (Telaah Teoritis dan Empiris). *Jurnal Manajemen dan Kewirausahaan*, 14(2), 145-153.
- Mulyadi. (1998). Total Quality Management, Edisi I, Aditya Media, Yogyakarta.
- Nelson, R. R. (1982). The Role of Knowledge in R&D Efficiency. *Oxford University Press*, 97(3), 453-470.
- Niven, P. R. (2006). *Balanced Scorecard Step-by-Step Maximizing Performance And Maintaining Results (2th. Ed.)* New Jersey: John Wiley & Sons.
- Roberts, M. L., Albright, T. L., & Hibbets, A. R. (2004). Debiassing Balanced Scorecard Evaluations. *Behavioral Research in Accounting*, 16(1), 75-88.
- Serfas, S. (2011). Cognitive Biases in the Capital Investment Context: Theoretical Consideration and Empirical Experiments on Violations of Normative Rationality. Retrieved May 17, 2014, from <http://books.google.co.id/books?id=i7OJWje1JgQC&printsec=frontcover#v=onepage&q&f=false>
- Tubbs, R. M. (1992). The Effect of Experience on The Auditor's Organization and Amount of Knowledge. *The Accounting Review*, 67(4), 783-801.
- Tversky, A., & Kahneman D. (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology*, 5, 207-232.
- Tversky, A., & Kahneman D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185, 1124-1131.
- Wahyuni, S., & Hartono, J. (2014). Reaksi Investor terhadap Pengumuman Laba: Pengujian atas Pengungkapan Informasi *Management Guidance*.