

Studi Eksperimental: Mengurangi Bias Pengukuran Umum Balanced Scorecard Dalam Penilaian Kinerja Pada Mahasiswa S1 Angkatan 2010 dan 2012

Evelyne Tri Kurniawati dan Juniarti
Akuntansi Bisnis Universitas Kristen Petra
Email: yunie.peter.petra.ac.id

ABSTRAK

Dalam penelitian Lipe dan Salterio (2000) menemukan adanya bias pengukuran umum. Bias pengukuran umum ini mengurangi manfaat dari BSC serta berpengaruh pada keputusan manajer terutama dalam mengalokasikan kompensasi. Dengan memakai pendekatan *disaggregated/mechanically aggregated* dan pengetahuan, bias pengukuran umum dapat berkurang, tetapi kedua pendekatan ini belum sepenuhnya dapat mengurangi bias pengukuran umum. Oleh karena itu, dilakukanlah penelitian ini dengan menggabungkan kedua pendekatan tersebut serta melihat apakah evaluasi kinerja menggunakan BSC mempengaruhi alokasi kompensasi. Metodologi penelitian yang dilakukan adalah eksperimen laboratorium. Uji hipotesis penelitian ini menggunakan *repeated measure ANOVA*, *independent sample t test* serta analisis regresi. Hasil penelitian menunjukkan bahwa pendekatan *disaggregated/mechanically aggregated* memiliki pengaruh signifikan mengurangi bias pengukuran umum dalam mengevaluasi kinerja menggunakan BSC. Ditemukan juga bahwa pendekatan pengetahuan tidak memiliki pengaruh signifikan mengurangi bias pengukuran umum. Selain itu, evaluasi kinerja menggunakan BSC memiliki pengaruh signifikan terhadap alokasi kompensasi.

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

ABSTRACT

In the research of Lipe & Salterio (2000), it was found a common measurement bias. This common measure bias reduces the benefits of the BSC and the effect on the manager's decision, especially in allocating compensation. By using the disaggregated/mechanically aggregated and knowledge approach, common measure bias can be reduced, but these two approaches is not yet fully reduced the bias. Therefore, this study was conducted by combining both approaches and seen if the performance evaluation by using the BSC affected the allocation of compensation. The methodology of research was a laboratory experiment. The hypothesis test of this study used a repeated measure ANOVA, independent sample t test and regression analysis. The results showed that the disaggregated / aggregated mechanically approach had significant influence in reducing general bias in evaluating the performance measurement by using BSC. It was also found that the knowledge approach didn't have a significant influence in reducing the common measure bias. Furthermore, the evaluation of the performance by using BSC had a significant influence on the allocation of compensation.

Keywords: BSC, common measurement, unique measurement, *debiasing*, common measure bias, *disaggregated /mechanically aggregated*, knowledge, allocation of compensation.

PENDAHULUAN

BSC merupakan sistem pengukuran yang rumit dan memiliki biaya yang mahal (Lipe & Salterio, 2000; Malina & Selto, 2001; Lindberg & Schonfeldt, 2008; Sawalqa, Holloway, & Alam, 2011; Islam, 2010). Sehingga dalam mengadopsi BSC, perlu digunakan secara optimal agar biaya yang telah dikeluarkan untuk menerapkan BSC dapat dimanfaatkan dengan baik. Tetapi hasil penelitian Lipe & Salterio (2000) menemukan adanya bias pengukuran umum pada BSC, dikarenakan manajer lebih cenderung menggunakan pengukuran umum dibandingkan pengukuran unik. Dampak adanya bias ini adalah mengurangi manfaat BSC sebagai alat manajemen strategi perusahaan (Dilla & Steinbart, 2005; Libby, Salterio, & Webb, 2002), dan membatasi keefektifan BSC sebagai alat pengukuran kinerja (Dilla & Steinbart, 2005). Selain itu, bias pengukuran umum juga mempengaruhi manajer untuk mengambil keputusan yang kurang tepat terutama dalam menentukan kompensasi (Malina & Selto, 2001).

Agar penggunaan BSC dapat optimal, ada beberapa penelitian berusaha untuk mengurangi bias pengukuran umum. Seperti penelitian Roberts, Albright, & Hibbets (2004) memakai pendekatan *disaggregated/mechanically aggregated* serta Dilla & Steinbart (2005) memakai pendekatan pengetahuan. Hasil dari kedua pendekatan itu telah berhasil untuk mengurangi bias, namun bias tersebut masih ada. Walaupun kedua pendekatan tersebut dapat mengurangi bias, tetapi hasil dari kedua pendekatan itu belum dapat mengoptimalkan BSC karena kedua pendekatan tersebut dilakukan secara parsial.

Penelitian ini akan mencoba untuk menggabungkan kedua pendekatan tersebut agar dapat mengoptimalkan BSC. Tujuan penelitian ini menguji pengaruh pendekatan *disaggregated/mechanically aggregated* dan pengetahuan terhadap bias pengukuran umum serta menguji pengaruh evaluasi kinerja menggunakan BSC terhadap alokasi kompensasi. Sampel yang digunakan pada penelitian ini adalah mahasiswa S1 angkatan 2010 dan 2012 Fakultas Ekonomi Universitas Kristen Petra. Berdasarkan latar belakang diatas, ada dua rumusan masalah dalam penelitian ini, pertama adalah Apakah penggabungan pendekatan *disaggregated/mechanically aggregated* dan pengetahuan

kepada partisipan dapat mengurangi bias pengukuran umum pada BSC? Serta yang kedua adalah Apakah penilaian kinerja berbasis BSC berpengaruh pada alokasi kompensasi?

Pengertian BSC

BSC adalah alat untuk mengukur kinerja perusahaan secara komprehensif (Kaplan & Norton, 1992), dimana di dalam BSC terdapat ukuran keuangan dan non keuangan yang meliputi perspektif pembelajaran dan pertumbuhan, perspektif proses bisnis internal, perspektif pelanggan (Kaplan & Norton, 1996; Dilla & Steinbart, 2005; Grevinga, 2013). Dalam BSC juga meliputi dua macam pengukuran yaitu pengukuran umum dan unik (Lipe & Salterio, 1998, 2000; Dilla & Steinbart, 2005; Roberts, Albright, & Hibbets, 2004; Grevinga, 2013).

Pengertian Pengukuran Umum dan Unik

Pengukuran umum adalah pengukuran kinerja baik keuangan ataupun non keuangan yang berlaku untuk seluruh organisasi perusahaan (Kaskey, 2008). Pengukuran umum sesuai dengan strategi organisasi (Grevinga, 2013). Manfaatnya adalah dapat membantu menyederhanakan tugas pengambil keputusan pada saat melakukan penilaian kinerja bawahan.

Pengukuran unik adalah pengukuran kinerja baik keuangan ataupun non keuangan yang berlaku untuk masing-masing tiap unit bisnis. Pengukuran unik merupakan ukuran yang sesuai dengan strategi tiap divisi (Grevinga, 2013). Saat mengevaluasi kinerja, pengambil keputusan yang menggunakan pengukuran unik akan dapat berhasil mengimplementasi strategi.

Pengertian Bias Pengukuran Umum

Adanya bias pengukuran umum karena manajer lebih cenderung menggunakan ukuran umum daripada ukuran unik (Lipe & Salterio, 2000; Humphreys & Trotman, 2011; Banker, Chang, & Pizzini, 2004; Bone & Solihin, 2012). Dampak dari bias pengukuran umum adalah mengurangi manfaat BSC sebagai alat manajemen strategi perusahaan (Dilla & Steinbart, 2005; Libby, Salterio, & Webb, 2002), membatasi keefektifan BSC sebagai alat pengukuran

kinerja (Dilla & Steinbart, 2005), mempengaruhi keputusan manajer yang kurang tepat dalam alokasi kompensasi (Roberts, Albright, & Hibbets, 2004).

Mengurangi Bias Pengukuran Umum (*Debiasing*)

Debiasing adalah prosedur untuk mengurangi atau menghilangkan bias dari strategi kognitif pembuat keputusan (Bazerman & Moore, 2009), sehingga memerlukan waktu dan usaha agar dapat berhasil menghilangkannya (Serfas, 2011). Temuan Einhorn (1972) pada BSC telah melibatkan dua langkah yaitu pertama, memisahkan keputusan evaluasi menjadi beberapa keputusan yang lebih kecil serta yang kedua, mengagregatkan keputusan-keputusan kecil kedalam nilai total berdasarkan *pre-determined weights* (dalam Roberts, Albright, & Hibbets, 2004). Dilakukan pemilahan, sebab manusia memiliki kemampuan kognitif yang terbatas (Tversky & Kahneman, 1973, 1974).

Penilaian dalam pendekatan *disaggregated/mechanically aggregated*, dapat mengurangi bias serta dapat menurunkan tuntutan kognitif. Pada pendekatan ini, BSC dibagi menjadi empat sampai tujuh pengukuran untuk setiap empat kategori dalam BSC (Kaplan & Norton, 1996). Pendekatan lain untuk mengurangi bias yaitu pengetahuan. Pengambil keputusan yang memiliki pengetahuan dapat lebih memanfaatkan semua informasi saat mengevaluasi kinerja ataupun mengalokasikan kompensasi.

Pengukuran BSC dan Alokasi Kompensasi

Dalam mengadopsi BSC seharusnya dikaitkan dengan keputusan kompensasi (Kaplan & Norton, 1996). Jika perusahaan tidak menggunakan kompensasi, motivasi karyawan akan menurun dan memberikan dampak pada penggunaan BSC yang tidak efektif. Namun dengan adanya bias pengukuran umum, mempengaruhi keputusan manajer yang kurang tepat terutama dalam menentukan kompensasi (Holmstrom & Milgrom, 1991) serta mempengaruhi motivasi kinerja karyawan dimana karyawan akan mengurangi usaha mereka pada aktivitas yang dianggap tidak utama (Malina & Selto, 2001; Holmstrom & Milgrom, 1991)

Pengaruh pendekatan *Disaggregated/Mechanically Aggregated* terhadap bias pengukuran umum

Lipe & Salterio (2000) menemukan adanya bias pengukuran umum pada saat evaluasi kinerja. Bias ini telah mengurangi manfaat BSC sebagai alat manajemen strategi perusahaan (Dilla & Steinbart, 2005; Libby, Salterio, & Webb, 2002).

Untuk mengurangi bias pengukuran umum, penelitian ini akan memakai pendekatan *disaggregated/mechanically aggregated*. Sebab jika dilakukan pemisahan pada BSC, akan membantu manajer untuk lebih memberikan perhatian pada ukuran unik daripada ukuran umum (Grevinga, 2013). Berdasarkan kajian diatas, hipotesis yang dapat disimpulkan adalah:

H1: *Disaggregated/Mechanically Aggregated* akan mengurangi bias pengukuran umum dalam mengevaluasi kinerja menggunakan *Balanced Scorecard*.

Pengaruh pendekatan pengetahuan terhadap bias pengukuran umum

Salah satu penyebab munculnya bias pengukuran umum yang ditemukan oleh Lipe & Salterio (2000) adalah kurangnya pengetahuan tentang BSC yang dimiliki oleh partisipan dalam eksperimen mereka. Sehingga untuk dapat mengurangi bias pengukuran umum, penelitian ini juga menggunakan pendekatan pengetahuan.

Pengetahuan dapat memungkinkan pengambil keputusan untuk lebih memanfaatkan semua informasi serta digunakan untuk mengevaluasi kinerja dan alokasi kompensasi (Dilla & Steinbart, 2005). Berdasarkan kajian diatas, hipotesis yang dapat disimpulkan adalah:

H2: Pengetahuan akan mengurangi bias pengukuran umum dalam evaluasi kinerja menggunakan BSC.

Pengaruh penilaian kinerja terhadap alokasi kompensasi

BSC digunakan sebagai alat untuk mengevaluasi kinerja dan alat manajemen strategis, karena dapat meningkatkan konsistensi evaluasi kinerja dan keputusan kompensasi (Dilla & Steinbart, 2005). Dengan adanya BSC dapat membantu pengambil keputusan untuk mengevaluasi kinerja tiap

karyawan (Grevinga, 2013; Roberts, Albright, & Hibbets, 2004), sehingga pengambil keputusan dapat menentukan alokasi kompensasi yang tepat. Alokasi kompensasi yang tepat dapat meningkatkan motivasi kinerja karyawan menjadi lebih baik untuk dapat mencapai tujuan perusahaan (Malina & Selto, 2001). Berdasarkan kajian diatas, hipotesis yang dapat disimpulkan adalah:

H3: Evaluasi kinerja menggunakan BSC mempengaruhi alokasi kompensasi.

METODE PENELITIAN

Penelitian ini menggunakan model pendekatan eksperimen laboratorium. Tujuan menggunakan pendekatan ini adalah untuk melihat hubungan sebab akibat (kausal) antara variabel yang dikontrol dan variabel kontrol (Fauzi, 2002; Azwar, 2005). Alasan penelitian ini menggunakan eksperimen laboratorium, karena secara langsung peneliti dapat terlibat dalam penelitian ini (Isnawijayani, 2011). Partisipan dalam penelitian ini adalah 40 mahasiswa S1 angkatan 2010 dan 2012 Fakultas Ekonomi Universitas Kristen Petra baik yang sedang dan sudah mengambil mata kuliah Akuntansi Manajemen sehingga mempunyai pengetahuan tentang struktur dan konsep BSC.

Desain Eksperimen

Penelitian ini diadopsi dari penelitian Lipe & Salterio (2000); Roberts, Albright, & Hibbets (2004); serta Dilla & Steinbart (2005). Penelitian ini tidak menyertakan *reward* untuk partisipan, dan partisipan tidak mencantumkan identitas dalam hasil pekerjaan mereka. Variabel yang tidak dikontrol adalah pengetahuan partisipan. Partisipan akan diberi kasus yang dari dua divisi perusahaan WCS yaitu divisi RadWear (menjual pakaian wanita) dan WorkWear (menjual seragam kerja). Penelitian ini mereplikasi desain pada penelitian Lipe & Salterio (2000) yang menggunakan desain 2 x 2 x 2, diantara dua faktor *between-subjects* (umum dan unik) dan satu faktor *within-subjects* (divisi). Faktor pertama *between-subjects* (umum), menunjukkan apakah RadWear atau WorkWear akan memiliki kinerja lebih baik jika diukur dengan pengukuran umum. Faktor kedua *between-subjects* (unik), menunjukkan apakah RadWear atau WorkWear akan memiliki

kinerja lebih baik jika diukur dengan pengukuran unik. Tiap partisipan menilai manajer dari dua divisi, sehingga divisi merupakan faktor *within-subjects*.

Partisipan diminta berperan sebagai seorang eksekutif senior dari perusahaan WCS untuk menilai kinerja kedua manager divisi. Partisipan akan diberikan informasi mengenai visi dan strategi masing-masing divisi serta instruksi untuk mengerjakan kasus yang diberikan. Setelah itu, partisipan akan diberi informasi tentang latar belakang perusahaan WCS dan partisipan akan diberikan kesempatan untuk membaca serta memahami kasus selama 15-20 menit. Lalu partisipan akan diberikan 16 ukuran Balanced Scorecard yang meliputi 2 ukuran umum dan 2 ukuran unik untuk tiap perspektif. Bobot masing-masing unit pada 16 pengukuran umum dan unik, menunjukkan 6,25% untuk tiap ukuran (100/16). Total bobot untuk tiap perspektif sebesar 25%, sedangkan bobot *pre-determined* untuk setiap ukuran sebesar 4% dan 9%. Serta *pre-determined weight* untuk ukuran unik sebesar 64% dari total keseluruhan *pre-determined weight*.

Partisipan diberikan tugas untuk menyelesaikan dua langkah *disaggregated* BSC, yaitu: partisipan memberikan penilaian kinerja pada tiap manajer, dan partisipan diminta untuk mengalikan penilaian individu dengan bobot yang telah ditentukan (*pre-determined weight*), menjumlahkan bobot nilai untuk membuat total, dan mengagregatkan nilai pada tiap divisi. untuk pengukuran *unique*, bobot yang telah ditentukan adalah 64% dari total *pre-determined weight*.

Setelah selesai melakukan penilaian *dissaggregated* divisi RadWear, partisipan diminta untuk menjumlahkan hasil perkalian nilai tertimbang agar dapat menghitung skor *mechanically aggregated* divisi RadWear. Setelah itu, partisipan menyelesaikan kasus dari divisi WorkWear dengan tahapan pengerjaan yang sama dengan divisi RadWear. Penelitian ini tidak menduplikasi *separate overall judgment*. Setelah melakukan evaluasi kinerja tiap manajer divisi, partisipan mengalokasikan dana pada akhir tahun sebesar Rp 100.000.000,00 di antara dua manajer divisi. Partisipan akan menyediakan informasi demografis mereka, serta penelitian ini tidak menguji manipulasi cek karena diasumsikan sudah diuji oleh penelitian sebelumnya.

Variabel dependen adalah perbedaan penilaian kinerja dari kedua manager divisi

dan alokasi kompensasi. Alokasi kompensasi dalam penelitian ini tidak ada dasar perhitungan yang ditentukan. Teknik yang digunakan pada penelitian ini untuk menguji hipotesis yaitu *repeated measures ANOVA*, *independent sample t test*, dan analisis regresi linier berganda. Penelitian ini tidak melakukan uji autokorelasi karena data penelitian tidak dilakukan secara *time series*.

HASIL PENELITIAN DAN PEMBAHASAN

Eksperimen ini dilakukan dua periode, periode pertama kelompok partisipan berpengetahuan dan periode kedua kelompok partisipan non-pengetahuan. Data dalam penelitian ini dilakukan uji normalitas dan uji hipotesis. Hasil dari pengujian hipotesis yang dilakukan, data sudah berdistribusi normal. Pada hipotesis 3 telah bebas dari heterokedastisitas dan multikolinearitas.

Tabel 1. Profil Partisipan Berdasarkan Angkatan

Angkatan Berdasarkan Partisipan Dengan Pengetahuan dan Tanpa Pengetahuan

Pengetahuan		Frequency	Percent	Valid Percent	Cumulative Percent
Dengan Pengetahuan	Valid 2010	7	46,7	46,7	46,7
	2012	8	53,3	53,3	100,0
	Total	15	100,0	100,0	
Tanpa Pengetahuan	Valid 2010	5	33,3	33,3	33,3
	2012	10	66,7	66,7	100,0
	Total	15	100,0	100,0	

Tabel 2. Profil Partisipan Berdasarkan Jurusan

Jurusan Berdasarkan Partisipan Dengan Pengetahuan dan Tanpa Pengetahuan

Pengetahuan		Frequency	Percent	Valid Percent	Cumulative Percent
Dengan Pengetahuan	Valid Akuntansi Bisnis	5	33,3	33,3	33,3
	Manajemen Bisnis	2	13,3	13,3	46,7
	Jurusan Lainnya	8	53,3	53,3	100,0
	Total	15	100,0	100,0	
Tanpa Pengetahuan	Valid Akuntansi Bisnis	2	13,3	13,3	13,3
	Manajemen Bisnis	13	86,7	86,7	100,0
	Total	15	100,0	100,0	

Tabel 3. Uji Normalitas Pengukuran Umum Partisipan Keseluruhan

One-Sample Kolmogorov-Smirnov Test

Division		Common	
RadWear	N	30	
	Normal Parameters ^{a,b}	Mean	28.7923
		Std. Deviation	2.52423
		Absolute	.159
	Most Extreme Differences	Positive	.085
		Negative	-.159
	Kolmogorov-Smirnov Z	.870	
Asymp. Sig. (2-tailed)	.436		
WorkWear	N	30	
	Normal Parameters ^{a,b}	Mean	21.8930
		Std. Deviation	5.66974
		Absolute	.264
	Most Extreme Differences	Positive	.264
		Negative	-.153
	Kolmogorov-Smirnov Z	1.443	
Asymp. Sig. (2-tailed)	.031		

a. Test distribution is Normal.
b. Calculated from data.

Tabel 4. Uji Normalitas Pengukuran Unik Partisipan Keseluruhan

One-Sample Kolmogorov-Smirnov Test

Division		Unique	
RadWear	N	30	
	Normal Parameters ^{a,b}	Mean	53.5510
		Std. Deviation	4.02346
		Absolute	.156
	Most Extreme Differences	Positive	.063
		Negative	-.156
	Kolmogorov-Smirnov Z	.855	
Asymp. Sig. (2-tailed)	.458		
WorkWear	N	30	
	Normal Parameters ^{a,b}	Mean	45.2057
		Std. Deviation	6.35441
		Absolute	.233
	Most Extreme Differences	Positive	.233
		Negative	-.146
	Kolmogorov-Smirnov Z	1.274	
Asymp. Sig. (2-tailed)	.078		

a. Test distribution is Normal.
b. Calculated from data.

Data pengukuran umum divisi WorkWear sebesar 0,031 menunjukkan bahwa tidak berdistribusi normal. Untuk mengatasi hal tersebut dilakukan transformasi data ke dalam bentuk logaritma. Berikut adalah hasil transformasi logaritma:

Tabel 5. Uji Normalitas Pengukuran Umum Partisipan Keseluruhan Transformasi Logaritma

One-Sample Kolmogorov-Smirnov Test

Division		LOG_Common	
RadWear	N	30	
	Normal Parameters ^{a,b}	Mean	1.4575
		Std. Deviation	.04103
		Absolute	.183
	Most Extreme Differences	Positive	.107
		Negative	-.183
	Kolmogorov-Smirnov Z	1.004	
Asymp. Sig. (2-tailed)	.265		
WorkWear	N	30	
	Normal Parameters ^{a,b}	Mean	1.3273
		Std. Deviation	.10592
		Absolute	.243
	Most Extreme Differences	Positive	.243
		Negative	-.147
	Kolmogorov-Smirnov Z	1.333	
Asymp. Sig. (2-tailed)	.057		

a. Test distribution is Normal.
b. Calculated from data.

Tabel 6. Uji Normalitas Pengukuran Unik Partisipan Keseluruhan Transformasi Logaritma

One-Sample Kolmogorov-Smirnov Test

Division		LOG_Unique	
RadWear	N	30	
	Normal Parameters ^{a,b}	Mean	1.7275
		Std. Deviation	.03412
		Absolute	.174
	Most Extreme Differences	Positive	.077
		Negative	-.174
	Kolmogorov-Smirnov Z	.952	
Asymp. Sig. (2-tailed)	.325		
WorkWear	N	30	
	Normal Parameters ^{a,b}	Mean	1.6512
		Std. Deviation	.05949
		Absolute	.209
	Most Extreme Differences	Positive	.209
		Negative	-.148
	Kolmogorov-Smirnov Z	1.142	
Asymp. Sig. (2-tailed)	.147		

a. Test distribution is Normal.
b. Calculated from data.

Berdasarkan hasil diatas, data yang digunakan telah berdistribusi normal.

Tabel 7. Uji Hipotesis Pertama Partisipan Keseluruhan

Tests of Between-Subjects Effects						
Measure: MEASURE_1						
Transformed Variable: Average						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	
Intercept	142.457	1	142.457	64734.873	.000	
Common	.160	1	.160	72.685	.000	
Unique	1.323	1	1.323	601.058	.000	
Common * Unique	.011	1	.011	4.947	.028	
Error	.255	116	.002			

Tests of Within-Subjects Contrasts						
Measure: MEASURE_1						
Source	division	Type III Sum of Squares	df	Mean Square	F	Sig.
division	Linear	1.323	1	1.323	601.058	.000
division * Common	Linear	.011	1	.011	4.947	.028
division * Unique	Linear	142.457	1	142.457	64734.873	.000
division * Common * Unique	Linear	.160	1	.160	72.685	.000
Error(division)	Linear	.255	116	.002		

Hasil ini menyimpulkan bahwa hasil pengukuran umum pada kedua divisi berbeda signifikan dan hasil pengukuran unik pada kedua divisi berbeda signifikan.

Tabel 8. Uji Normalitas Pengukuran Umum Partisipan Berpengetahuan

One-Sample Kolmogorov-Smirnov Test			
Division			Common Dengan Pengetahuan
RadWear	N		15
	Normal Parameters ^{a,b}	Mean	28.7173
		Std. Deviation	3.10682
		Absolute	.232
	Most Extreme Differences	Positive	.141
		Negative	-.232
	Kolmogorov-Smirnov Z		.900
Asymp. Sig. (2-tailed)		.393	
WorkWear	N		15
	Normal Parameters ^{a,b}	Mean	21.8347
		Std. Deviation	5.50102
		Absolute	.230
	Most Extreme Differences	Positive	.230
		Negative	-.164
	Kolmogorov-Smirnov Z		.891
Asymp. Sig. (2-tailed)		.405	

a. Test distribution is Normal.
b. Calculated from data.

Tabel 9. Uji Normalitas Pengukuran Unik Partisipan Berpengetahuan

One-Sample Kolmogorov-Smirnov Test			
Division			Unik Dengan Pengetahuan
RadWear	N		15
	Normal Parameters ^{a,b}	Mean	53.5407
		Std. Deviation	4.50374
		Absolute	.141
	Most Extreme Differences	Positive	.087
		Negative	-.141
	Kolmogorov-Smirnov Z		.548
Asymp. Sig. (2-tailed)		.925	
WorkWear	N		15
	Normal Parameters ^{a,b}	Mean	44.8887
		Std. Deviation	6.74827
		Absolute	.249
	Most Extreme Differences	Positive	.249
		Negative	-.157
	Kolmogorov-Smirnov Z		.966
Asymp. Sig. (2-tailed)		.308	

a. Test distribution is Normal.
b. Calculated from data.

Pengukuran umum dan unik menghasilkan yang signifikan, dapat dilihat dari nilai signifikansi > 0,05 yaitu 0,393 dan 0,405 pada pengukuran umum, serta nilai signifikansi pada pengukuran unik sebesar 0,925 dan 0,308. Maka dapat disimpulkan bahwa data telah berdistribusi normal.

Tabel 10. Uji Hipotesis Pertama Partisipan Berpengetahuan

Tests of Between-Subjects Effects						
Measure: MEASURE_1						
Transformed Variable: Average						
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	
Intercept	41616.446	1	41616.446	3148.695	.000	
Common	452.486	1	452.486	34.235	.000	
Unique	4297.948	1	4297.948	325.182	.000	
Common * Unique	5.870	1	5.870	.444	.508	
Error	740.155	56	13.217			

Tests of Within-Subjects Contrasts						
Measure: MEASURE_1						
Source	division	Type III Sum of Squares	df	Mean Square	F	Sig.
division	Linear	4297.948	1	4297.948	325.182	.000
division * Common	Linear	5.870	1	5.870	.444	.508
division * Unique	Linear	41616.446	1	41616.446	3148.695	.000
division * Common * Unique	Linear	452.486	1	452.486	34.235	.000
Error(division)	Linear	740.155	56	13.217		

Nilai signifikansi F sebesar 0,508 > 0,05, dapat disimpulkan bahwa hasil pengukuran umum kedua divisi tidak berbeda signifikan. Nilai signifikansi F sebesar 0,000 < 0,05, dapat disimpulkan pengukuran unik kedua divisi berbeda signifikan.

Tabel 11. Uji Normalitas Pengukuran Umum Partisipan Non-Pengetahuan

One-Sample Kolmogorov-Smirnov Test			
Division			Common Tanpa Pengetahuan
RadWear	N		15
	Normal Parameters ^{a,b}	Mean	28.8673
		Std. Deviation	1.87996
		Absolute	.125
	Most Extreme Differences	Positive	.125
		Negative	-.123
	Kolmogorov-Smirnov Z		.484
Asymp. Sig. (2-tailed)		.973	
WorkWear	N		15
	Normal Parameters ^{a,b}	Mean	21.9513
		Std. Deviation	6.02657
		Absolute	.325
	Most Extreme Differences	Positive	.325
		Negative	-.178
	Kolmogorov-Smirnov Z		1.257
Asymp. Sig. (2-tailed)		.085	

a. Test distribution is Normal.
b. Calculated from data.

Tabel 12. Uji Normalitas Pengukuran Unik Partisipan Non-Pengetahuan

One-Sample Kolmogorov-Smirnov Test			
Division			Unique Tanpa Pengetahuan
RadWear	N		15
	Normal Parameters ^{a,b}	Mean	53.5613
		Std. Deviation	3.63990
	Most Extreme Differences	Absolute	.181
		Positive	.101
		Negative	-.181
	Kolmogorov-Smirnov Z		.701
Asymp. Sig. (2-tailed)		.710	
WorkWear	N		15
	Normal Parameters ^{a,b}	Mean	45.5227
		Std. Deviation	6.15522
	Most Extreme Differences	Absolute	.266
		Positive	.266
		Negative	-.146
	Kolmogorov-Smirnov Z		1.029
Asymp. Sig. (2-tailed)		.240	

a. Test distribution is Normal.
b. Calculated from data.

Nilai signifikansi 0,973 dan 0,085 pada pengukuran umum, sedangkan pengukuran unik sebesar 0,710 dan 0,240. Berdasarkan hasil tersebut dapat disimpulkan bahwa data telah berdistribusi normal.

Tabel 13. Uji Hipotesis Pertama Partisipan 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	42132.768	1	42132.768	3704.410	.000
Common	419.329	1	419.329	36.868	.000
Unique	4367.892	1	4367.892	384.035	.000
Common * Unique	2.363	1	2.363	.208	.650
Error	636.926	56	11.374		

Tests of Within-Subjects Contrasts						
Measure: MEASURE_1						
Source	division	Type III Sum of Squares	df	Mean Square	F	Sig.
division	Linear	4367.892	1	4367.892	384.035	.000
division * Common	Linear	2.363	1	2.363	.208	.650
division * Unique	Linear	42132.768	1	42132.768	3704.410	.000
division * Common * Unique	Linear	419.329	1	419.329	36.868	.000
Error (division)	Linear	636.926	56	11.374		

Nilai signifikansi F sebesar 0,650 menunjukkan bahwa pengukuran umum kedua divisi tidak berbeda signifikan. Sedangkan nilai signifikansi F sebesar 0,000, menunjukkan bahwa pengukuran unik kedua divisi berbeda signifikan. Hal tersebut menunjukkan bahwa kedua pengukuran belum dilakukan secara proporsional.

Tabel 14. Uji Normalitas Pengukuran Umum dan Unik Partisipan Berpengetahuan dan Non-Pengetahuan

One-Sample Kolmogorov-Smirnov Test					
Pengetahuan		Common	Unique		
Dengan Pengetahuan	N	15	15		
	Normal Parameters ^{a,b}	Mean	50.5520	98.4293	
		Std. Deviation	6.89179	8.33060	
	Most Extreme Differences	Absolute	.230	.087	
		Positive	.230	.080	
		Negative	-.157	-.087	
	Kolmogorov-Smirnov Z		.892	.336	
Asymp. Sig. (2-tailed)		.405	0.999		
Tanpa Pengetahuan	N	15	15		
	Normal Parameters ^{a,b}	Mean	50.8187	99.0840	
		Std. Deviation	6.88397	5.88783	
	Most Extreme Differences	Absolute	.256	.227	
		Positive	.256	.227	
		Negative	-.166	-.168	
	Kolmogorov-Smirnov Z		.992	.880	
Asymp. Sig. (2-tailed)		.279	.421		

a. Test distribution is Normal.
b. Calculated from data.

Partisipan yang diberi pengetahuan dan yang tidak diberi pengetahuan, menghasilkan nilai signifikansi pengukuran umum sebesar 0,405 dan 0,279, sedangkan pengukuran unik sebesar 0,999 dan 0,421. Dapat disimpulkan bahwa data telah berdistribusi normal.

Tabel 15. Uji Hipotesis Kedua Pengukuran Umum dan Unik Partisipan Berpengetahuan dan Non-Pengetahuan

Group Statistics					
CODE	N	Mean	Std. Deviation	Std. Error Mean	
C Pengetahuan	15	50.552	6.892	1.779	
	15	50.819	6.884	1.777	
U Non Pengetahuan	15	98.429	8.331	2.151	
	15	99.084	5.888	1.520	

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
C	Equal variances assumed	.002	.965	-1.06	28	.916	-.267	2.515	-5.419	4.885
	Equal variances not assumed			-1.06	28.000	.916	-.267	2.515	-5.419	4.885
U	Equal variances assumed	1.232	.276	-.249	28	.806	-.655	2.634	-6.050	4.741
	Equal variances not assumed			-.249	25.194	.806	-.655	2.634	-6.050	4.768

Nilai signifikansi sebesar 0,965 menggunakan metode *independent sample t test with equal variances assumed* sebesar 0,916. Maka dapat disimpulkan bahwa tidak terdapat perbedaan signifikan pada pengukuran umum pada partisipan yang diberi pengetahuan dan yang tidak. Sedangkan pada pengukuran unik nilai signifikansi sebesar 0,276 menggunakan metode *independent sample t test with equal variances assumed* sebesar 0,806. Maka dapat disimpulkan tidak terdapat perbedaan

signifikan antara partisipan yang diberi pengetahuan dan yang tidak.

Tabel 16. Uji Normalitas Hipotesis 3
One-Sample Kolmogorov-Smirnov Test

		Unstandardized Residual
N		30
Normal Parameters(a,b)	Mean	,0000000
	Std. Deviation	1,23435818
Most Extreme Differences	Absolute	,089
	Positive	,080
	Negative	-,089
Kolmogorov-Smirnov Z		,488
Asymp. Sig. (2-tailed)		,971

a. Test distribution is Normal.
b. Calculated from data.

Tabel ini menunjukkan kedua pengukuran berdistribusi normal dengan nilai signifikan sebesar 0,971.

Tabel 17. Uji Heteroskedastisitas Hipotesis 3
Coefficients(a)

Model		t	Sig.
1	(Constant)	-,966	,343
	RC	,418	,679
	RU	1,510	,143
	WC	,627	,536
	WU	-1,258	,220

a. Dependent Variable: YRatio

Dapat diketahui bahwa nilai signifikansi variabel bebas sebesar 0,679; 0,143; 0,536 dan 0,220. Maka disimpulkan bahwa tidak terjadi heteroskedastisitas pada kedua divisi.

Tabel 18. Uji Multikolinieritas Hipotesis 3
Coefficients(a)

Model		Collinearity Statistics	
		Tolerance	VIF
1	(Constant)		
	RC		
	RU	,251	3,986
	WC	,266	3,752
	WU	,160	6,257
		,147	6,786

a. Dependent Variable: YRatio

Variabel kedua divisi bebas dari multikolinieritas sebab memiliki nilai *tolerance* > 0,10 dan nilai VIF < 10.

Tabel 19. Uji F Hipotesis 3
ANOVA

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	30,827	4	7,707	4,360	,008 ^a
	Residual	44,186	25	1,767		
	Total	75,012	25			

a. Predictors: (Constant), WU, RU, RC, WC
b. Dependent Variable: Y Ratio

Nilai signifikansi sebesar 0,008 menyimpulkan bahwa kedua pengukuran

secara simultan berpengaruh signifikan terhadap alokasi kompensasi pada kedua divisi.

Tabel 20. Uji R Square Hipotesis 3
Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	,641 ^a	,411	,317	1,32944

a. Predictors: (Constant), WU, RU, RC, WC
b. Dependent Variable: Y Ratio

Nilai R Square sebesar 0,411 menunjukkan bahwa alokasi kompensasi pada kedua divisi dipengaruhi oleh kedua pengukuran sebesar 41,1%, sedangkan sisanya sebesar 58,9% dipengaruhi faktor lain.

Tabel 21. Uji t Hipotesis 3
Coefficients(a)

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-4,288	4,436		-,966	,343
	RC	,082	,195	,128	,418	,679
	RU	,180	,119	,449	1,510	,143
	WC	,068	,109	,241	,627	,536
	WU	-,127	,101	-,503	-1,258	,220

a. Dependent Variable: YRatio

Dari Tabel 4.30 diketahui bahwa kedua pengukuran divisi RadWear dan WorkWear, tidak mempengaruhi alokasi kompensasi secara signifikan.

Nilai signifikansi perbedaan antara divisi dengan pengukuran umum dan unik sebesar 0,000 dan 0,028. Dapat disimpulkan bahwa pendekatan *dissaggregated/ mechanically aggregated* dapat mengurangi bias pengukuran umum pada BSC, diterima. Nilai signifikansi pengukuran umum sebesar 0,806 dan unik sebesar 0,916. Dapat disimpulkan pendekatan pengetahuan akan mengurangi bias pengukuran umum dalam evaluasi kinerja menggunakan BSC, ditolak. Karena tidak ada perbedaan signifikan antara partisipan yang diberikan pengetahuan dengan yang tidak. Partisipan tidak terlalu memperdulikan informasi yang diberikan peneliti, sebab partisipan merasa bahwa informasi tersebut sudah cukup untuk melakukan evaluasi kinerja (Debusk et al, 2005). Jika dilakukan pengujian secara simultan, evaluasi kinerja menggunakan pengukuran di BSC mempengaruhi alokasi kompensasi tiap divisi karena hasilnya kurang dari 0,05. Hasil ini konsisten dengan penelitian Roberts, Albright, & Hibbets (2004), Banker et al (2004), serta Dilla dan Steinbart (2005). Dapat disimpulkan evaluasi kinerja

menggunakan BSC mempengaruhi alokasi kompensasi, diterima.

KESIMPULAN

Hipotesis 1 penelitian ini diterima, karena pengukuran unik BSC memiliki nilai yang signifikan dimana menunjukkan bahwa bias pengukuran umum berkurang. Hipotesis 2 penelitian ini ditolak, karena tidak ada perbedaan yang signifikan antara kelompok yang diberi pengetahuan dengan yang tidak, walaupun bias pengukuran umum berkurang. Hipotesis 3 penelitian ini diterima, karena secara simultan kedua pengukuran berpengaruh signifikan terhadap alokasi kompensasi kedua divisi.

Keterbatasan dan Saran untuk Penelitian Selanjutnya

Beberapa keterbatasan yang sekaligus merupakan saran untuk penelitian selanjutnya:

1. Menambah dan memilih kriteria partisipan yang memiliki pengalaman cukup menggunakan BSC.
2. Menentukan waktu yang disesuaikan dengan kondisi eksperimen.
3. Mengukur kembali pengetahuan partisipan sebelum dan sesudah pendekatan diberikan dengan memberikan beberapa pertanyaan terkait BSC.

DAFTAR PUSTAKA

- Azwar, S. (2005). *Metode penelitian* (cetakan VI). Yogyakarta: Pustaka Pelajar.
- Banker, R. D., Chang, H., & Pizzini, M. J. (2004). The balanced scorecard: judgemental effects of performance measures linked to strategy. *The Accounting Review*, 79 (1), 1-23.
- Bawono, I. R., Halim, A., & Lord, B. (2012, November). *Public sector performance measurement and budget allocation: an Indonesian experiment*.
- Bazerman, M. H., & Moore, D. A. (2009). *Judgement in Managerial Decision Making* (7th. ed.). New Jersey: John Wiley & Sons.
- Bone, H. & Sholihin, M. (2012). Pengaruh perspektif dan jenis ukuran dalam balanced scorecard terhadap evaluasi kinerja. *Thesis, Universitas Gadjah Mada*.
- Bowman, E. H. (1963). Consistency and optimality in managerial decision making. *Management Science*, 9 (1), 31-321.
- Dilla, W. N., Steinbart, P. J., (2005). Relative weighting of common and unique balanced scorecard measures by knowledgeable decision makers. *Behavioral Research in Accounting*, 17, 43-53.
- Fauzi, M. (2002). *Metode Penelitian Kuantitatif*. Semarang: Walisongo Press.
- Frijo, M. L., & Krumwiede, K. R. (2000, January). *The balanced scorecard: a winning performance measurement system*. *strategic finance*, 81(7), 50-54.
- Grevinga, K. H. M. (2013). Common measure bias in the balanced scorecard: an experiment with undergraduate students. Unpublished undergraduate thesis, University of Twente, Enschede.
- Holmstrom, B., Milgrom, P. (1991). Multitask principal-agent analyses: incentives contracts, asset ownership, and job design. *Journal of Law, Economics, and Organization*, 7, 24-52.
- Humphreys, K. A., & Trotman, K. T. (2011). The balanced scorecard: the effect of strategy information on performance evaluation judgments. *Journal of Management Accounting Research*, 23, 81-98.
- Indriantoro, N., & Supomo, B. (1999, Oktober), *Metodologi penelitian bisnis untuk akuntansi dan manajemen*. BPFY-Yogyakarta, Edisi Pertama.
- Islam, M. (2010). The link between perception of BSC implementation and corporate strategy and its impact on performance. *Management Accounting Section (MAS) Meeting Paper*. Retrieved from <http://ssrn.com/abstract=1659845//>
- Isnawijayani. (2011, Juni). Metode eksperimen dalam penelitian ilmu komunikasi., 4 (7), 75-88. Retrieved from <http://jodfisipunbara.files.wordpress.com/2012/07/5-Isnawijayani-oke-hal-1-8.pdf>
- Kahneman, D., Tversky, A. (1973). Availability: a heuristic for judging frequency and probability. *Cognitive Psychology*, 5, 207-232
- Kahneman, D., Tversky, A. (1974). Judgement under uncertainty: heuristic and biases. *Science*, 185, 1124-1131.
- Kaplan, R. S., & Norton, D. P. (1992). *The balanced scorecard*. Retrieved from [Harvard Business Review website: http://www.hbr.org//](http://www.hbr.org//)

- Kaplan, R. S., Norton, D. P. (1996, Fall). Linking the balanced scorecard to strategy. *California Management Review*, 39 (1), 53-79.
- 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. *PhD Dissertation. School of Business and Technology, Capella University, USA.*
- Libby, T., Salterio, S., & Webb, A. (2002). The balanced scorecard: the effects of assurance and process accountability on managerial judgement. Retrieved from SSRN Database.
- Lindberg, E., Schonfeldt, S. N. (2008). The balanced scorecard at skelleftea municipality.
- Lipe, M. G., Salterio S. E. (1998). The balanced scorecard: judgemental effects of information organization and diversity. Retrieved from SSRN Database.
- Lipe, M. G., Salterio S. E. (2000, July). The balancedscorecard: judgemental effects off common and unique performance measure. *The Accounting Review*, 75 (3), 283- 298. Retrieved from ProQuest Database.
- Malina, M. A., Selto, F. H. (2001). Communicating and controlling strategy: an empirical study of the effectiveness of the balanced scorecard. *Journal of Management Accounting Research*, 13, 47. Retrieved from SSRN Database..
- Roberts, M. L., Albright, T. L., & Hibbets, A. R. (2004). Debiasing balanced scorecard evaluation. *Behavioral Research in Accounting*, 16, 75-88.
- Sawalqa, F. A., Holloway, D., & Alam, M. (2011). Balanced scorecard implementation in jordan: an initial analysis. *International Journal of Electronic Business Management*, 9 (3), 196- 210.