

Studi Eksperimental: Mengurangi Bias Pengukuran Umum Balanced Scorecard Dalam Penilaian Kinerja Pada Mahasiswa S1 Program Manajemen Pemasaran

Lana Aristya Prabowo dan Juniarti
Akuntansi Bisnis Universitas Kristen Petra
Email: yunie@peter.petra.ac.id

ABSTRAK

Pada penelitian terdahulu ditemukan bahwa terdapat bias pengukuran umum pada *Balanced Scorecard*. Bias pengukuran umum menyebabkan ketidakakuratan penilaian kinerja, berkurangnya manfaat BSC, serta berpengaruh pada keputusan alokasi kompensasi kepada karyawan. Beberapa penelitian terdahulu telah berhasil mengurangi bias pengukuran umum menggunakan pendekatan *disaggregated/ mechanically aggregated* dan pendekatan pengetahuan. Penelitian-penelitian tersebut belum dapat mengurangi bias pengukuran umum BSC secara optimal. Untuk itu, penelitian dilakukan untuk melakukan penelitian dengan menggabungkan kedua pendekatan tersebut.

Penelitian ini dilakukan dengan laboratorium eksperimen. Penelitian menggunakan uji hipotesis *repeated measures ANOVA*, *independent sample t test*, dan analisis regresi linear berganda. Hasil penelitian ini membuktikan 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. 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

In the previous study, it was found that there was a common measure bias on the Balanced Scorecard. Common measure bias caused inaccuracy in the performance evaluation, reduced the benefits of BSC, also affected on the allocation of compensation to the employee. Several previous studies succeeded in reducing the common measure bias by using disaggregated / mechanically aggregated and knowledge approaches. These studies were not able to reduce common measure bias of BSC optimally. For that reason, a research conducted by combining both approaches.

This study was conducted with laboratory experiment. This research using hypothesis test with repeated measures ANOVA, independent sample t test, and multiple linear regression analysis. The results showed that the disaggregated/ mechanically aggregated could reduce the common measure bias in the performance evaluation by using the BSC. It was also found that the knowledge approach proven to reduce the common measure bias. Performance evaluation by using BSC proved to affect the allocation of compensation.

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

PENDAHULUAN

Balanced Scorecard (BSC) merupakan suatu alat pengukuran yang digunakan perusahaan untuk melakukan penilaian kinerja. Para pengambil keputusan dapat menggunakan dua pengukuran BSC, yaitu pengukuran umum yang digunakan dalam mengukur performa perusahaan secara umum dan pengukuran unik yang digunakan dalam penilaian kinerja per divisi perusahaan (Kaplan & Norton, 1996b; Dilla & Steinbart, 2005; Grevinga, 2013). Dominasi penggunaan ukuran umum dalam penilaian kinerja akan menimbulkan bias pada pengukuran umum BSC (Lipe & Salterio, 2000), sehingga ukuran umum dan unik haruslah digunakan keduanya secara proporsional.

Dalam penerapannya *Balanced Scorecard* (BSC) membutuhkan biaya yang relatif mahal dan juga rumit untuk dikembangkan (Lipe & Salterio, 2000; Lindberg & Schonfelt, 2008; Lawrie & Cobbold, 2004). Mahalnya BSC disebabkan karena mahalnya instalasi software dan banyaknya biaya yang dikeluarkan untuk investasi pada karyawan dengan pengadaan training, seminar, workshop untuk membangun fondasi yang kuat dengan pemahaman konsep akan BSC (Marr & Neely, 2003; Lawrie & Cobbold, 2004). Bias dalam pengukuran umum BSC menyebabkan berkurangnya manfaat dan kegunaan BSC (Lipe & Salterio, 2000; Roberts, Albright, & Hibbets, 2004).

Dominasi ukuran umum yang menimbulkan bias pengukuran umum pada BSC ditemukan Lipe & Salterio (2000). Penelitian Lipe & Salterio menunjukkan bahwa pengambil keputusan yang belum memiliki pengetahuan yang cukup akan BSC cenderung lebih bergantung pada komponen umum dan mengabaikan komponen unik yang terdapat dalam divisi tersebut.

Penelitian-penelitian setelah Lipe & Salterio sudah berusaha mengurangi bias dalam pengukuran umum pada BSC, tetapi belum berhasil mengurangi bias secara optimal disebabkan pendekatan dilakukan secara parsial. Roberts, Albright, & Hibbets (2004) melakukan penelitian dengan mencari pendekatan untuk mencoba mengurangi bias pengukuran umum pada BSC menggunakan *disaggregated/mechanically aggregated*. *Disaggregated/mechanically aggregated* ini dilakukan dengan penilaian kinerja perusahaan menggunakan penilaian umum dan unik pada tiap divisi secara terpisah. Roberts, Albright, dan Hibbets juga menguji

keterkaitan penilaian kinerja menggunakan BSC dan alokasi kompensasi. Hasil penelitian ini menyimpulkan bahwa *disaggregated/mechanically aggregated* mempengaruhi alokasi kompensasi dan bias dalam pengukuran umum BSC masih ditemukan dalam penelitian ini.

Kemudian, Dilla & Steinbart (2005) melakukan penelitian dengan mengadopsi penelitian Lipe dan Salterio, tetapi dengan perbedaan pada pengetahuan partisipan akan BSC. Hasil dari penelitian ini menyimpulkan bahwa dengan pengetahuan, bias pengukuran umum dapat dieliminasi meskipun belum maksimal karena partisipan masih menitik beratkan pada pengukuran umum daripada pengukuran unik.

Berdasarkan penelitian-penelitian sebelumnya yang belum optimal dalam mengurangi bias pengukuran umum, maka peneliti tertarik melakukan penelitian dengan mencoba menggabungkan pendekatan yang sebelumnya dilakukan secara parsial menggunakan pendekatan *disaggregated/mechanically aggregated* dan pendekatan pengetahuan. Peneliti berharap dapat mengurangi bias secara lebih optimal sehingga BSC dapat bermanfaat bagi perusahaan.

Pengukuran Umum dan Pengukuran Unik *Balanced Scorecard*

Disamping menggunakan pengukuran keuangan dan non-keuangan, terdapat pengukuran umum dan unik dalam *Balanced Scorecard* (BSC). Pengukuran umum merupakan pengukuran keuangan dan non-keuangan yang berlaku dalam seluruh divisi perusahaan. Ukuran umum ini disesuaikan dengan strategi perusahaan (Grevinga, 2013). Pada umumnya ukuran umum diterapkan di seluruh unit dalam perusahaan (Kaskey, 2008).

Bias Pengukuran Umum

Bias pengukuran umum diartikan sebagai ketidakmampuan pengambil keputusan untuk menyertakan informasi unik dalam penilaian kinerja sehingga ukuran unik diabaikan, yang disebabkan karena untuk memproses informasi ini dibutuhkan usaha kognitif yang lebih besar (Slovic & MacPhillamy 1974; Lipe & Salterio, 2005).

Bias pengukuran umum timbul karena pengambil keputusan cenderung menggunakan ukuran umum dan mengabaikan ukuran unik untuk menilai kinerja perusahaan (Kaplan & Norton, 1996;

Lipe & Salterio, 2000). Bias ukuran umum ini juga disebabkan karena keterbatasan seseorang dalam mengolah informasi (Dilla & Steinbart, 2005).

Terjadinya bias pengukuran umum ini berdampak pada penilaian kinerja dan pengambilan keputusan kompensasi oleh para pengambil keputusan untuk para manajer divisi atau karyawan yang tidak adil (Gibbs, Merchant, Wim A Van der Stede, & Vargus, 2005).

Mengurangi bias pengukuran umum bermanfaat pada penilaian kinerja dan pengambilan keputusan kompensasi yang lebih adil.

Mengurangi Bias Pengukuran Umum (Debiasing)

Debiasing merupakan sebuah prosedur untuk mengurangi atau mengeliminasi bias dari strategi kognitif para pengambil keputusan (Bazerman, 2006). Beberapa pendekatan untuk mengurangi bias yaitu dengan pengetahuan, pengalaman, dan keahlian, insentif dan akuntabilitas, dan pelatihan dan alat (Serfas, 2011). Terdapat pendekatan lain yang dapat digunakan untuk mengurangi bias yaitu pendekatan *disagregat/mechanically aggregated* (Bowman, 1963; Einhorn, 1972 (dalam Roberts, Albright, & Hibbets, 2004); Roberts, Albright, & Hibbets, 2004).

Disagregat / mechanically aggregated merupakan alat untuk mengurangi bias dengan menyederhanakan informasi yang perlu diproses (Roberts, et al., 2004). Einhorn (dalam Roberts, Albright, & Hibbets, 2004) dan Bowman (1963) mengemukakan pendekatan *disagregat/mechanically aggregated* yang dapat diaplikasikan dalam BSC melibatkan dua langkah proses, yaitu: (1) *disaggregating* atau memisahkan keputusan evaluasi menjadi beberapa keputusan yang lebih kecil dan (2) agregat keputusan kecil menjadi skor keseluruhan berdasarkan bobot yang telah ditentukan (Einhorn, 1972 ; Lyness & Cornelius, 1982 ; Edwards & Newman, 1982 (dalam Roberts, Albright, & Hibbets, 2004); Roberts, Albright, & Hibbets, 2004).

Penilaian *disagregated / mechanically aggregated*, akan menurunkan tuntutan kognitif karena jumlah informasi yang harus dipertimbangkan untuk mengevaluasi setiap dimensi individu berkurang.

Pendekatan lain yang dapat digunakan untuk mengurangi bias yaitu

pendekatan pengetahuan, pendekatan ini sudah dilakukan oleh Dilla & Steinbart (2005). Semakin baik pembuat keputusan memahami teori dan struktur BSC, maka akan semakin mudah untuk menggabungkan ukuran-ukuran baik umum dan unik ketika membandingkan dan mengevaluasi kinerja masing-masing divisi dan perusahaan (Dilla & Steinbart, 2005).

Hubungan *Balanced Scorecard* dan Alokasi Kompensasi

Kaplan dan Norton (1996b, 2001) menyarankan perusahaan untuk menghubungkan keputusan kompensasi dengan BSC untuk meningkatkan manfaat BSC tersebut. BSC perlu dikaitkan dengan alokasi kompensasi agar pengukuran kinerja dapat berubah menjadi adil, dan diharapkan dapat merubah perilaku karyawan serta meningkatkan motivasi karyawan atas pemberian *reward* yang sudah adil tersebut.

Selain itu, Roberts, Albright, & Hibbets (2004) juga mengemukakan bahwa keputusan kompensasi dari para pengambil keputusan sangat didukung oleh evaluasi kinerja menggunakan pendekatan *disagregated/mechanically aggregated* BSC.

METODE PENELITIAN

Penelitian menggunakan pendekatan laboratorium eksperimen. Penelitian laboratorium eksperimen adalah penelitian yang menguji hubungan sebab akibat pada lingkungan yang artifisial atau buatan, dalam penelitian campur tangan peneliti sangat tinggi (Indriyantoro & Supomo, 1999). Peneliti menggunakan model eksperimen ini karena peneliti dapat memiliki kontrol terhadap eksperimen dan dapat memanipulasi situasi dalam eksperimen. Dalam eksperimen ini, partisipan akan ditugaskan menjadi seorang manajer senior yang mengevaluasi kinerja manajer dua divisinya, yaitu divisi RadWear dan WorkWear pada perusahaan WCS.

Partisipan dalam penelitian ini adalah 32 mahasiswa S1 Jurusan Manajemen Pemasaran 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 dan Salterio (2000), Roberts, Albright, & Hibbets (2004), dan Dilla dan Steinbart (2005). Peneliti menggunakan desain penelitian 2 x 2 x 2, meliputi penilaian antar subjek (umum dan unik antar divisi), dan penilaian dalam subjek (divisi). Faktor pertama antar subjek, umum, mengindikasikan apakah RadWear atau WorkWear tampil lebih baik pada ukuran umum. Faktor antar subjek kedua, unik, mengindikasikan apakah RadWear atau WorkWear tampil lebih baik pada ukuran unik. Faktor dalam subjek mengindikasikan apakah ukuran umum dan unik RadWear lebih baik dari pada ukuran umum dan unik WorkWear.

Partisipan diminta berperan sebagai seorang eksekutif senior dari perusahaan WCS untuk menilai kinerja kedua manager divisi. Setelah itu, partisipan akan diberi informasi tentang latar belakang perusahaan WCS serta visi misi dan strategi masing-masing divisi. Selanjutnya partisipan akan diberikan kesempatan untuk membaca dan memahami kasus selama 15-20 menit. Lalu partisipan akan diberikan 16 ukuran BSC 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 manager, dan partisipan diminta untuk mengalikan penilaian individu dengan bobot yang telah ditentukan (*pre-determined weight*), menjumlahkan bobot nilai sebagai nilai 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 sebagai 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 manager divisi, partisipan mengalokasikan

dana pada akhir tahun sebesar Rp 100.000.000,00 di antara dua manager divisi. Sebagai langkah terakhir, partisipan akan menyelesaikan pertanyaan lanjutan terkait kasus, memberikan informasi demografis.

Variabel dependen penelitian ini adalah perbedaan penilaian kinerja dari kedua manager divisi dan alokasi kompensasi.

Teknik analisis yang digunakan untuk menguji hipotesis yaitu *repeated measures ANOVA*, *independent sample t test*, dan analisis regresi linier berganda.

HASIL PENELITIAN DAN PEMBAHASAN

Profil Responden

Tabel 1

Jenis Kelamin Berdasarkan Responden yang Diberi Pengetahuan dan Tidak Diberi Pengetahuan

	Frequency	Percent	Valid Percent	Cumulative Percent
Pengetahuan Valid	Laki-laki	7	43,8	43,8
	Perempuan	9	56,2	100,0
	Total	16	100,0	100,0
Tanpa Pengetahuan Valid	Laki-laki	6	37,5	37,5
	Perempuan	10	62,5	100,0
	Total	16	100,0	100,0

Tabel 2

Responden yang Sudah Mengambil AM, Mengulang AM, dan Sedang Mengambil AM

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Sudah lulus AM	16	50,0	50,0
	Mengulang AM	3	9,4	59,4
	Sedang mengambil AM	13	40,6	100,0
Total	32	100,0	100,0	

Repeated Measures ANOVA Pada Partisipan Dengan Pengetahuan

a. Uji Normalitas

Tabel 3

Uji Normalitas Data Pengukuran Umum Pada Partisipan Dengan Pengetahuan

Division		Common Dengan Pengetahuan	
RadWear	N	16	
	Normal Parameters ^{a,b}	Mean	26.6231
		Std. Deviation	1.75778
		Absolute	.138
	Most Extreme Differences	Positive	.089
		Negative	-.138
Kolmogorov-Smirnov Z		.551	
Asymp. Sig. (2-tailed)		.922	
WorkWear	N	16	
	Normal Parameters ^{a,b}	Mean	24.4950
		Std. Deviation	3.33917
		Absolute	.375
	Most Extreme Differences	Positive	.375
		Negative	-.200
Kolmogorov-Smirnov Z		1.500	
Asymp. Sig. (2-tailed)		.022	

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

Division		LOG2 Common Dengan Pengetahuan	
RadWear	N	16	
	Normal Parameters ^{a,b}	Mean	.1535
		Std. Deviation	.00900
		Absolute	.137
	Most Extreme Differences	Positive	.092
		Negative	-.137
Kolmogorov-Smirnov Z		.550	
Asymp. Sig. (2-tailed)		.923	
WorkWear	N	16	
	Normal Parameters ^{a,b}	Mean	.1415
		Std. Deviation	.01526
		Absolute	.332
	Most Extreme Differences	Positive	.332
		Negative	-.153
Kolmogorov-Smirnov Z		1.327	
Asymp. Sig. (2-tailed)		.059	

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

Tabel 4
Uji Normalitas Data Pengukuran Unik Pada Partisipan Dengan Pengetahuan

Division		Unique Dengan Pengetahuan	
RadWear	N	16	
	Normal Parameters ^{a,b}	Mean	47.3219
		Std. Deviation	3.47199
		Absolute	.169
	Most Extreme Differences	Positive	.169
		Negative	-.145
Kolmogorov-Smirnov Z		.678	
Asymp. Sig. (2-tailed)		.748	
WorkWear	N	16	
	Normal Parameters ^{a,b}	Mean	46.3406
		Std. Deviation	2.14641
		Absolute	.115
	Most Extreme Differences	Positive	.115
		Negative	-.111
Kolmogorov-Smirnov Z		.461	
Asymp. Sig. (2-tailed)		.984	

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

Tabel 4 menunjukkan pengukuran umum pada divisi WorkWear tidak berdistribusi normal karena bernilai < 0,05 yaitu 0,022. Untuk mengatasi ketidaknormalan data dilakukan transformasi logaritma. Berikut merupakan hasil uji normalitas setelah dilakukan transformasi logaritma:

Tabel 5
Uji Normalitas Data Pengukuran Umum Pada Partisipan Dengan Pengetahuan Setelah Transformasi Logaritma

Tabel 6
Uji Normalitas Data Pengukuran Unik Pada Partisipan Dengan Pengetahuan Setelah Transformasi Logaritma

Division		LOG2 Unique Dengan Pengetahuan	
RadWear	N	16	
	Normal Parameters ^{a,b}	Mean	.2237
		Std. Deviation	.00857
		Absolute	.159
	Most Extreme Differences	Positive	.150
		Negative	-.159
Kolmogorov-Smirnov Z		.636	
Asymp. Sig. (2-tailed)		.814	
WorkWear	N	16	
	Normal Parameters ^{a,b}	Mean	.2215
		Std. Deviation	.00519
		Absolute	.122
	Most Extreme Differences	Positive	.114
		Negative	-.122
Kolmogorov-Smirnov Z		.487	
Asymp. Sig. (2-tailed)		.972	

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

Tabel 5 dan 6 menunjukkan bahwa setelah dilakukan transformasi logaritma, data pengukuran umum dan unik pada masing-masing divisi telah berdistribusi normal dengan nilai signifikansi *kolmogorov smirnov* > 0,05.

b. Uji Hipotesis

Tabel 7
Analisis *Repeated Measures* ANOVA Pada Partisipan Dengan Pengetahuan

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	1.096	1	1.096	21166.288	.000
Common	.000	1	.000	7.743	.007
Unique	.045	1	.045	870.902	.000
Common * Unique	.000	1	.000	3.779	.057
Error	.003	60	5.177E-005		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	Division	Type III Sum of Squares	df	Mean Square	F	Sig.
Division	Linear	.045	1	.045	870.902	.000
Division * Common	Linear	.000	1	.000	3.779	.057
Division * Unique	Linear	1.096	1	1.096	21166.288	.000
Division * Common * Unique	Linear	.000	1	.000	7.743	.007
Error(Division)	Linear	.003	60	5.177E-005		

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

Analisis Repeated Measures ANOVA Pada Partisipan Tanpa Pengetahuan

a. Uji Normalitas

Tabel 8

Uji Normalitas Data Pengukuran Umum Pada Partisipan Tanpa Pengetahuan

One-Sample Kolmogorov-Smirnov Test

Division	Common Tanpa Pengetahuan
N	16
Normal Parameters ^{a,b}	Mean: 24.9931 Std. Deviation: 2.37857
RadWear	Absolute: .134
Most Extreme Differences	Positive: .122 Negative: -.134
Kolmogorov-Smirnov Z	.535
Asymp. Sig. (2-tailed)	.937
N	16
Normal Parameters ^{a,b}	Mean: 22.9419 Std. Deviation: 2.97122
WorkWear	Absolute: .110
Most Extreme Differences	Positive: .110 Negative: -.084
Kolmogorov-Smirnov Z	.439
Asymp. Sig. (2-tailed)	.990

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

Tabel 9

Uji Normalitas Data Pengukuran Unik Pada Partisipan Tanpa Pengetahuan

One-Sample Kolmogorov-Smirnov Test

Division	Unique Tanpa Pengetahuan
N	16
Normal Parameters ^{a,b}	Mean: 44.6231 Std. Deviation: 4.26837
RadWear	Absolute: .118
Most Extreme Differences	Positive: .103 Negative: -.118
Kolmogorov-Smirnov Z	.472
Asymp. Sig. (2-tailed)	.979
N	16
Normal Parameters ^{a,b}	Mean: 44.0888 Std. Deviation: 2.98329
WorkWear	Absolute: .121
Most Extreme Differences	Positive: .121 Negative: -.085
Kolmogorov-Smirnov Z	.484
Asymp. Sig. (2-tailed)	.973

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

Data diatas menunjukkan bahwa data pengukuran umum dan unik pada masing-masing divisi telah berdistribusi normal.

b. Uji Hipotesis

Tabel 10

Analisis Repeated Measures ANOVA Pada Partisipan Tanpa Pengetahuan

Tests of Between-Subjects Effects

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	37344.737	1	37344.737	7180.866	.000
Common	13.371	1	13.371	2.571	.114
Unique	3325.507	1	3325.507	639.448	.000
Common * Unique	4.602	1	4.602	.885	.351
Error	312.035	60	5.201		

Tests of Within-Subjects Contrasts

Measure: MEASURE_1

Source	Division	Type III Sum of Squares	df	Mean Square	F	Sig.
Division	Linear	3325.507	1	3325.507	639.448	.000
Division * Common	Linear	4.602	1	4.602	.885	.351
Division * Unique	Linear	37344.737	1	37344.737	7180.866	.000
Division * Common * Unique	Linear	13.371	1	13.371	2.571	.114
Error(Division)	Linear	312.035	60	5.201		

Dari hasil diatas menyimpulkan bahwa hasil pengukuran umum pada divisi RadWear dan divisi WorkWear tidak berbeda signifikan. Sedangkan hasil pengukuran unik pada kedua divisi berbeda signifikan.

Analisis Repeated Measures ANOVA pada Partisipan Keseluruhan

a. Uji Normalitas

Tabel 10

Uji Normalitas Data Pengukuran Umum Pada Partisipan Keseluruhan

One-Sample Kolmogorov-Smirnov Test

Division	Common
N	32
Normal Parameters ^{a,b}	Mean: 25.8081 Std. Deviation: 2.21771
RadWear	Absolute: .112
Most Extreme Differences	Positive: .092 Negative: -.112
Kolmogorov-Smirnov Z	.634
Asymp. Sig. (2-tailed)	.816
N	32
Normal Parameters ^{a,b}	Mean: 23.7184 Std. Deviation: 3.20770
WorkWear	Absolute: .181
Most Extreme Differences	Positive: .181 Negative: -.076
Kolmogorov-Smirnov Z	1.023
Asymp. Sig. (2-tailed)	.246

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

Tabel 11

Uji Normalitas Data Pengukuran Unik Pada Partisipan Keseluruhan

One-Sample Kolmogorov-Smirnov Test

Division	Unique
N	32
Normal Parameters ^{a,b}	Mean: 45.9725 Std. Deviation: 4.06548
RadWear	Absolute: .116
Most Extreme Differences	Positive: .116 Negative: -.108
Kolmogorov-Smirnov Z	.658
Asymp. Sig. (2-tailed)	.779
N	32
Normal Parameters ^{a,b}	Mean: 45.2147 Std. Deviation: 2.80077
WorkWear	Absolute: .071
Most Extreme Differences	Positive: .066 Negative: -.071
Kolmogorov-Smirnov Z	.403
Asymp. Sig. (2-tailed)	.997

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

Dari data diatas menunjukkan bahwa data pengukuran umum dan unik pada masing-masing divisi telah berdistribusi normal.

b. Uji Hipotesis

Tabel 12
Analisis *Repeated Measures* ANOVA Pada Partisipan Keseluruhan

Measure: MEASURE_1
Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	79201.438	1	79201.438	16008.353	.000
Common	32.433	1	32.433	6.555	.012
Unique	6942.431	1	6942.431	1403.218	.000
Common * Unique	7.096	1	7.096	1.434	.233
Error	613.491	124	4.948		

Measure: MEASURE_1

Source	Division	Type III Sum of Squares	df	Mean Square	F	Sig.
Division	Linear	6942.431	1	6942.431	1403.218	.000
Division * Common	Linear	7.096	1	7.096	1.434	.233
Division * Unique	Linear	79201.438	1	79201.438	16008.353	.000
Division * Common * Unique	Linear	32.433	1	32.433	6.555	.012
Error(Division)	Linear	613.491	124	4.948		

Dari hasil diatas menyimpulkan bahwa hasil pengukuran umum pada divisi RadWear dan divisi WorkWear tidak berbeda signifikan. Sedangkan hasil pengukuran unik pada kedua divisi berbeda signifikan.

Analisis Independent Sample T Test

a. Uji Normalitas:

Tabel 13
Uji Normalitas Data Pengukuran Umum dan Unik Berdasarkan Pemberian Pengetahuan Tentang BSC

One-Sample Kolmogorov-Smirnov Test

Pengetahuan		Common	Unique	
Dengan Pengetahuan	N	16	16	
	Normal Parameters ^{a,b}	Mean	51.1181	93.6625
		Std. Deviation	4.29648	4.58526
	Most Extreme Differences	Absolute	.209	.138
		Positive	.209	.103
	Negative	-.135	-.138	
Kolmogorov-Smirnov Z		.836	.554	
Asymp. Sig. (2-tailed)		.486	.919	
Tanpa Pengetahuan	N	16	16	
	Normal Parameters ^{a,b}	Mean	47.9350	88.7119
		Std. Deviation	5.01052	6.89651
	Most Extreme Differences	Absolute	.150	.126
		Positive	.150	.126
	Negative	-.093	-.105	
Kolmogorov-Smirnov Z		.601	.502	
Asymp. Sig. (2-tailed)		.863	.963	

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

Berdasarkan hasil diatas dapat disimpulkan data yang digunakan untuk *independent sample t test* telah berdistribusi normal.

b. Uji Hipotesis

Hasil *independent sample t test* pengukuran umum pada kelompok partisipan yang diberi pengetahuan tentang BSC dan partisipan yang diberi pengetahuan tentang BSC adalah sebagai berikut:

Tabel 14
Independent Sample T Test Pengukuran Umum

Group Statistics

Pengetahuan	N	Mean	Std. Deviation	Std. Error Mean
Common Dengan Pengetahuan	16	51.118	4.296	1.074
Common Tanpa Pengetahuan	16	47.935	5.011	1.253

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	1.610	.214	1.929	30	.063	3.183	1.650	-.187	6.553
		Equal variances not assumed			1.929	28.318	.063	3.183	1.650	-.190

Hasil diatas menyimpulkan bahwa terdapat perbedaan signifikan hasil pengukuran umum pada partisipan yang diberi pengetahuan tentang BSC dan partisipan yang tidak diberi pengetahuan tentang BSC.

Hasil *independent sample t test* pengukuran unik pada kelompok partisipan yang diberi pengetahuan tentang BSC dan partisipan yang tidak diberi pengetahuan tentang BSC adalah sebagai berikut:

Tabel 15
Independent Sample T Test Pengukuran Unik

Group Statistics

Pengetahuan	N	Mean	Std. Deviation	Std. Error Mean
Unique Dengan Pengetahuan	16	93.663	4.585	1.146
Unique Tanpa Pengetahuan	16	88.712	6.897	1.724

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	
Unique	Equal variances assumed	1.488	.232	2.391	30	.023	4.951	2.070	.722	9.179
		Equal variances not assumed			2.391	26.094	.024	4.951	2.070	.696

Hasil diatas menunjukkan bahwa terdapat perbedaan signifikan hasil pengukuran unik pada partisipan yang diberi pengetahuan

tentang BSC dan partisipan yang tidak diberi pengetahuan tentang BSC.

Analisis Regresi Linier Berganda

a. Uji Asumsi Klasik

1. Uji Normalitas

Tabel 16
Uji Normalitas Regresi RadWear dan WorkWear

One-Sample Kolmogorov-Smirnov Test

		Unstandardize d Residual
N		32
Normal Parameters(a,b)	Mean	.0000000
	Std. Deviation	.30738104
Most Extreme Differences	Absolute	.118
	Positive	.118
	Negative	-.104
Kolmogorov-Smirnov Z		.667
Asymp. Sig. (2-tailed)		.764

a. Test distribution is Normal.

b. Calculated from data.

Hasil diatas menyimpulkan bahwa residual regresi pada RadWear dan WorkWear telah berdistribusi normal.

2. Uji Heteroskedastisitas

Tabel 17
Uji Heteroskedastisitas Regresi RadWear dan WorkWear

Correlations						
		RadWear Common_ Log	RadWear Unique_ Log	WorkWear Common_ Log	WorkWear Unique_ Log	Abnormal Residual
Spearman's rho	RadWear Common_Log	Correlation Coefficient	1.000	.821**	.719**	.582**
		Sig. (2-tailed)	.	.000	.000	.000
		N	32	32	32	32
RadWear Unique_Log		Correlation Coefficient	.821**	1.000	.744**	.680**
		Sig. (2-tailed)	.000	.	.000	.000
		N	32	32	32	32
WorkWear Common_Log		Correlation Coefficient	.719**	.744**	1.000	.627**
		Sig. (2-tailed)	.000	.000	.	.000
		N	32	32	32	32
WorkWear Unique_Log		Correlation Coefficient	.582**	.680**	.627**	1.000
		Sig. (2-tailed)	.000	.000	.000	.
		N	32	32	32	32
Abnormal Residual		Correlation Coefficient	.237	.104	-.014	.044
		Sig. (2-tailed)	.192	.572	.939	.809
		N	32	32	32	32

** . Correlation is significant at the 0.01 level (2-tailed).

Tabel diatas menunjukkan bahwa korelasi rank spearman menghasilkan nilai yang signifikan pada abnormal residual RadWear Common, RadWear Unique, WorkWear Common, dan WorkWear Unique. Dari hasil pengujian diatas, maka disimpulkan tidak terjadi heteroskedastisitas dalam regresi kedua divisi.

3. Uji Multikolinieritas

Tabel 18
Uji Multikolinieritas Regresi RadWear dan WorkWear

Coefficients ^a								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-1.211	1.195		-1.013	.320		
	RadWear Common_Log	1.501	.858	.368	1.749	.092	.341	2.934
	RadWear Unique_Log	4.344	1.122	1.078	3.872	.001	.194	5.148
	WorkWear Common_Log	-1.809	.573	-.617	-3.158	.004	.394	2.538
	WorkWear Unique_Log	-3.361	.970	-.576	-3.463	.002	.543	1.842

a. Dependent Variable: Y ratio_Log

Berdasarkan data diatas diketahui bahwa keempat variabel bebas memiliki nilai tolerance > 0,10 dan nilai VIF < 10, sehingga disimpulkan regresi RadWear dan WorkWear bebas dari multikolinieritas.

b. Uji Hipotesis

1. Uji Simultan (Uji F)

Tabel 19
Uji F Regresi RadWear dan WorkWear

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3.793	4	.948	8.740	.000 ^b
	Residual	2.929	27	.108		
	Total	6.722	31			

a. Predictors: (Constant), WorkWear Unique, WorkWear Common, RadWear Common, RadWear Unique

b. Dependent Variable: Y ratio

Dari hasil uji F diatas, maka disimpulkan pengukuran umum dan unik secara simultan berpengaruh signifikan terhadap alokasi kompensasi pada divisi RadWear dan WorkWear.

Tabel 20
Nilai R Square Regresi RadWear dan WorkWear

Model Summary ^a				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.751 ^a	.564	.500	.32936

a. Predictors: (Constant), WorkWear Unique, WorkWear Common, RadWear Common, RadWear Unique

b. Dependent Variable: Y ratio

Nilai R Square sebesar 0,564 menunjukkan bahwa alokasi kompensasi pada kedua divisi dipengaruhi oleh pengukuran umum dan unik sebesar 56,4%, sedangkan sisanya sebesar 43,6% dipengaruhi faktor lain.

2. Uji Parsial (Uji t)

Tabel 21
Uji t Regresi RadWear dan WorkWear

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	
	B	Std. Error	Beta			
1	(Constant)	.263	.983		.267	.791
	RadWear Common	.083	.047	.394	1.769	.088
	RadWear Unique	.114	.034	.993	3.325	.003
	WorkWear Common	-.092	.028	-.631	-3.274	.003
	WorkWear Unique	-.093	.029	-.557	-3.211	.003

a. Dependent Variable: Y ratio

Dari tabel diatas diketahui pengukuran secara parsial berpengaruh signifikan terhadap alokasi kompensasi pada kedua divisi.

Berdasarkan hasil penilaian seluruh responden, maka dapat dilihat bahwa interaksi antara divisi dan pengukuran unik bernilai signifikan sebesar $0,000 < 0,05$. Dengan demikian, dapat disimpulkan bahwa *disaggregated/mechanically aggregated* dapat mengurangi bias pengukuran umum dalam penilaian kinerja menggunakan BSC.

Pengujian hipotesis kedua menghasilkan nilai signifikansi sebesar $0,063 < 0,1$ pada ukuran umum dan $0,023 < 0,05$ pada ukuran unik. Hasil ini menyimpulkan pengetahuan akan mengurangi bias pengukuran umum dalam penilaian kinerja menggunakan BSC diterima.

Dalam hasil pengujian penelitian ini, diketahui uji simultan (uji f) menghasilkan nilai signifikansi sebesar $0,000 < 0,05$. Dari hasil ini, maka dapat disimpulkan bahwa pengukuran umum dan unik secara simultan berpengaruh signifikan terhadap alokasi kompensasi pada divisi RadWear dan WorkWear. Berdasarkan uji regresi, maka dapat ditunjukkan bahwa alokasi kompensasi pada kedua divisi dipengaruhi oleh pengukuran umum dan unik sebesar 56,4%, sedangkan sisanya sebesar 43,6% dipengaruhi faktor lain. Berdasarkan uji parsial, maka alokasi kompensasi dipengaruhi oleh penilaian kinerja menggunakan BSC, ditunjukkan dengan nilai ukuran umum dan ukuran unik yang signifikan pada kedua divisi. Dengan demikian, dapat disimpulkan bahwa hipotesis ketiga yang menyatakan evaluasi kinerja menggunakan BSC mempengaruhi alokasi kompensasi diterima.

KESIMPULAN

1. Pendekatan *disaggregated/mechanically aggregated* dapat terbukti mengurangi bias pengukuran umum dalam penilaian kinerja menggunakan BSC, hipotesis 1 diterima.

2. Pendekatan pengetahuan terbukti mengurangi bias pengukuran umum dalam penilaian kinerja menggunakan BSC, hipotesis 2 diterima.
3. Penilaian kinerja menggunakan BSC terbukti mempengaruhi alokasi kompensasi, hipotesis 3 diterima

Keterbatasan dan Saran untuk Penelitian Selanjutnya

Beberapa keterbatasan yang sekaligus merupakan saran untuk penelitian selanjutnya:

1. Penelitian mendatang disarankan dapat menggunakan partisipan yang sudah berpengalaman menggunakan BSC sehingga partisipan dapat melakukan penilaian kinerja menggunakan BSC dengan baik dan menghasilkan penilaian kinerja yang lebih signifikan.
2. Peneliti selanjutnya diharapkan dapat mengukur pengetahuan partisipan dengan memberikan pertanyaan-pertanyaan terkait teori dan penerapan BSC.
3. Untuk penelitian mendatang, sebaiknya menggunakan pendekatan-pendekatan lain yang diasumsikan dapat mengurangi bias pengukuran umum dalam penilaian kinerja menggunakan BSC.

DAFTAR REFERENSI

- Bazerman, M., & Moore, D. A. (2009). *Judgment in Managerial Decision Making* (7th.ed.). New Jersey: John Wiley & Sons.
- Bonner, S. E. (1990). Experience Effects in Auditing: The Role of Task-Specific Knowledge. *The Accounting Review*, 65 (1), 72-92
- Bonner, S. E., & Lewis, B. L. (1990). Determinants of Auditor Expertise. *Journal of Accounting and Auditing*, 28, 1-20
- Bowman, E. H. (1963). Consistency and Optimality in Managerial Decision Making. *Management Science*, 9 (1), 31-321.
- 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.
- 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, Netherlands.*
- Indriantoro, N., & Supomo, B. (1999, Oktober). Metode 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. (2001). 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. (2001b). 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. *PhD Dissertation. School of Business and Technology, Capella University, USA.*
- 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. (2004). The Balanced Scorecard: The Effect 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.
- Lindberg, E., & Schonfeldt, S. N. (2008). The Balanced Scorecard at Skelleftea Municipality. *Bachelor Thesis, UMEA School of Business, Sweden.*
- 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.
- 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.
- Nelson, R. R. (1987). The Role of Knowledge in R & D Efficiency. *Oxford University Press*, 97 (3), 453-470.
- Roberts, M. L., Albright, T. L., & Hibbets, A. R. (2004). Debiasing 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. *Dissertation, Chemnitz University of Technology, Germany.* Retrieved from <http://books.google.co.id/books?id=i7OJWje1JgQC&printsec=frontcover#v=onepage&q&f=false>.
- 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 atau Pengungkapan Informasi Management Guidance.