

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

Unani, Yesyurun Benya dan Dra. Juniarti

Akuntansi Bisnis Universitas Kristen Petra

Email: yunie@peter.petra.ac.id

ABSTRAK

Lipe dan Salterio (2000) pertama kali menemukan adanya bias pengukuran umum ada Balanced Scorecard. Bias pengukuran umum ini berdampak pada penerapan BSC serta berpengaruh pada pengambilan keputusan manajer terutama dalam keputusan kompensasi. Penelitian terdahulu (Robert, Albright, Hibbets, 2004; Dilla dan Steinbart, 2005) berusaha untuk menanggulangi adanya bias pengukuran umum pada Balanced Scorecard. Namun, hasil dari kedua penelitian tersebut belum dapat mengurangi bias pengukuran umum BSC. Oleh karena itu, penelitian ini dengan menggabungkan pendekatan *disaggregated/ mechanically aggregated* dan pendekatan pengetahuan 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.

Dari hasil penelitian menunjukkan bahwa pendekatan *disaggregated/ mechanically aggregated* memiliki pengaruh signifikan mengurangi bias pengukuran umum dalam mengevaluasi kinerja menggunakan BSC, tetapi pendekatan pengetahuan tidak memiliki pengaruh signifikan dalam mengurangi bias pengukuran umum. Dari hasil penelitian ini juga terbukti bahwa evaluasi kinerja menggunakan BSC memiliki pengaruh signifikan terhadap alokasi kompensasi.

Kata kunci:

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

ABSTRACT

Lipe dan Salterio (2000) first discovered the existence of common measure bias Balanced Scorecard. Common measure bias has an impact on the implementation of the BSC and the effect on managers decision making, particularly in compensation decision. The previous research (Robert, Albright, Hibbets, 2004; Dilla dan Steinbart, 2005) attempted to overcome the common measurement bias of BSC. However, the resultsof both these studies have not been able to reduce the common measure bias. Therefore, this research approach by combining *disaggregated/ mechanically aggregated* and knowledge approaches and seen if the performance evaluation by using the BSC affected the allocation compensation, The methodology of research was a laboratory experiment. The hypothesis test of this study using repeated measure ANOVA, independent sample-t test, and regression analysis.

The reesults of the study showed that *disaggregated/ mechanically aggregated* approach have a significant affect in reducing common measurement bias in evaluating the performance measurement by using BSC, but knowledge approach did not have a significant affect in reducing common measurement bias. The other results of this study, proved that the evaluation performance by using BSC had a significant influence on the allocation of compensation.

Keywords:

Balanced Scorecard, common measure, unique measure, *debiasing*, common measure bias, *disaggregated/ mechanically aggregated*, knowledge, allocation of compensation.

PENDAHULUAN

Menurut Kaplan & Norton (1992, 1993), konsep dari Balance Scorecard telah dikembangkan sebagai sebuah alat ukur yang dapat digunakan oleh manajemen untuk menilai kinerja dari sebuah organisasi. Banyak perusahaan hanya mengutamakan kinerja finansial dalam mengevaluasi kinerja perusahaan (Newman, 1991). Balance Scorecard dibuat supaya perusahaan tidak hanya bergantung pada ukuran financial tetapi juga ukuran non-financial (Kaplan & Norton, 1996). Balance scorecard membawa perusahaan untuk fokus terhadap hal-hal yang penting bagi perusahaan (Kaplan & Norton, 1992, 1993). Berdasarkan strategi perusahaan, pengukuran Balance scorecard dapat dikategorikan dalam pengukuran umum dan pengukuran unik. Balanced Scorecard membantu perusahaan dalam menilai kinerja menggunakan pengukuran umum dan pengukuran unik yang terdapat dalam unit bisnis. Oleh karena itu, tiap unit bisnis harus mempunyai scorecard sendiri (Kaplan & Norton, 1996).

Didalam sebuah perusahaan terdapat berbagai unit bisnis yang memiliki strategi masing-masing. Pengukuran unik penting karena menunjukkan keunikan dari tiap unit bisnis. Pengukuran unik membantu mengukur keberhasilan strategi unit bisnis serta menunjukkan kinerja unit bisnis secara spesifik (Kaplan&Norton, 1996; Lipe&Salterio, 2000). Karena pengukuran unik ini penting maka perusahaan harus memperhatikan pengukuran unik dalam penilaian kinerja. Hal tersebut disebabkan yang menjadi pendorong keberhasilan strategi perusahaan adalah kinerja dari unit bisnis. Dengan begitu unit bisnis dapat membantu mencapai tujuan perusahaan.

Menurut Lipe dan Salterio (2000), evaluasi kinerja menggunakan balance scorecard dapat menimbulkan bias pengukuran umum. Lipe dan Salterio(2000) menunjukkan bahwa ukuran umum dan ukuran unik tidak digunakan secara seimbang. *Decision-maker* hanya menilai kinerja unit bisnis dari ukuran umum dan tidak mengindahkan ukuran unik. Setiap unit

bisnis memiliki ukuran yang unik, karena tiap unit bisnis memiliki fungsi yang berbeda. Dengan adanya bias pengukuran umum manfaat yang dapat diberikan dalam penerapan balance scorecard menjadi tidak optimal.

Lipe&Salterio (2000) menjelaskan bahwa penerapan balance scorecard mahal karena adanya teknologi dan pelatihan bagi karyawan. Apabila penerapan balance scorecard tidak optimal, maka biaya yang dikeluarkan sia-sia. Selain itu, bias pengukuran umum dapat menyebabkan hasil penilaian kinerja yang tidak tepat. Dengan hasil penilaian kinerja yang tidak tepat maka perusahaan tidak dapat menangkap strategi unit bisnis. Perusahaan yang tidak dapat menangkap strategi yang dijalankan unit bisnisnya, cenderung mengabaikan pengukuran unik dalam penilaian kinerja. Upaya unit bisnis untuk mendorong kinerja perusahaan menjadi sia-sia. Hal tersebut dapat mempengaruhi motivasi karyawan dalam bekerja.

Roberts et all (2004), penelitian ini menindaklanjuti penelitian yang dilakukan Lipe dan Salterio (2000) dengan tujuan untuk mengurangi bias yang muncul dalam penggunaan balance scorecard. Robert et all (2004) menggunakan pendekatan *Disaggregated/Mechanically Aggregated Balance Scorecard*.

Dilla and Steinbart (2005) juga menindaklanjuti penelitian Lipe dan Salterio (2000). Penelitian Dilla and Steinbart (2005) bertujuan mengurangi bias dengan melihat pengetahuan *decision-maker* tentang balance scorecard. Menurut Dilla and Steinbart (2005), pengetahuan tentang balance scorecard dapat membantu *decision-maker* untuk menggunakan ukuran umum dan ukuran unik dalam mengevaluasi kinerja unit bisnis.

Penelitian sebelumnya untuk mengurangi bias digunakan secara terpisah dan hasilnya belum optimal dalam mengurangi bias. Penelitian ini menggabungkan pendekatan *Disaggregated/Mechanically Aggregated* dan Pengetahuan tentang balanced scorecard

untuk mengurangi bias pengukuran umum dalam balanced scorecard.

Berdasarkan latar belakang di atas, maka rumusan masalah dalam penelitian ini adalah:

1. Apakah penggunaan Disagregated/Mechanically Aggregated Balance Scorecard dan Pengetahuan tentang Balance Scorecard dari partisipan dapat mengurangi bias pengukuran umum dalam Balance Scorecard?
2. Apakah evaluasi menggunakan BSC berpengaruh terhadap alokasi kompensasi?

Balance Scorecard

Menurut Kaplan dan Norton (1992), Balance Scorecard adalah sekumpulan ukuran yang dapat memberikan pandangan yang komprehensif bagi para top manager tentang area bisnisnya. Balance scorecard terdiri dari ukuran finansial dan non finansial. Ukuran non finansial dalam balance scorecard melengkapi ukuran finansial yang sebelumnya menjadi ukuran utama dalam evaluasi kinerja. Ukuran non finansial dapat dikatakan sebagai faktor pendorong dalam meningkatkan kinerja finansial dimasa mendatang. Balance scorecard membawa manager untuk fokus terhadap hal-hal yang penting dalam bisnisnya dan yang menjadi strategi perusahaan untuk dapat mencapai visi dan misi perusahaan. Balance Scorecard membagi ukuran kinerja menjadi empat perspektif yaitu perspektif finansial, perspektif customer, perspektif proses bisnis internal dan perspektif pengetahuan dan perkembangan.

Pengukuran Umum dan Pengukuran Unik

Pengukuran umum adalah pengukuran yang mengukur strategi organisasi secara keseluruhan baik pengukuran secara finansial maupun non finansial dan berlaku untuk seluruh unit bisnis dalam perusahaan (Kaplan&Norton, 1996; Lipe&Salterio, 2000; Grevinga, 2013). Pengukuran umum juga dibagi menjadi

empat perspektif balanced scorecard. Oleh karena itu, setiap perspektif mengandung ukuran umum.

Pengukuran unik adalah pengukuran yang mengukur strategi dari unit bisnis. Tiap unit bisnis mempunyai strategi dalam meningkatkan kinerja dan dengan pengukuran unik dapat diukur keberhasilan strategi unit bisnis tersebut (Kaplan&Norton, 1996; Lipe&Salterio, 2000; Grevinga, 2013). Oleh karena itu, tiap unit bisnis harus memiliki scorecard sendiri untuk dapat menilai kinerja unit tersebut.

Bias Pengukuran Umum

Bias pengukuran umum dalam Balance Scorecard dikemukakan oleh Lipe dan Salterio (2000). Bias pengukuran umum adalah keputusan dalam menilai kinerja untuk menggunakan salah satu ukuran lebih dominan dibandingkan dengan ukuran yang lain. Bias dalam balance scorecard terkait dengan pengukuran umum yang lebih dominan daripada pengukuran unik (Lipe dan Salterio, 2000). Lipe dan Salterio (2000) menemukan bahwa dalam menilai kinerja unit bisnis *decision maker* lebih fokus pada pengukuran umum dan mengabaikan ukuran unik. Dalam penelitian sebelumnya, Slovic dan McPhillamy (1994) membandingkan dua mahasiswa dalam mengelola informasi, hasilnya pengukuran umum memiliki bobot yang lebih tinggi dari pada pengukuran unik.

Penggunaan ukuran umum secara kognitif lebih mudah diproses daripada ukuran unik. Secara kognitif, *decision-maker* membutuhkan pemahaman lebih mendalam saat memproses pengukuran unik (Slovic&MacPhillamy, 1994; Lipe&Salterio, 1998, 2000). Selain itu, dalam membandingkan kinerja antar unit bisnis dalam suatu perusahaan *decision-maker* lebih sederhana dengan menilai pengukuran umum dari seluruh divisi. Dominasi ukuran umum tersebut terkait dengan strategi penyerdehanaan dalam penilaian kinerja yang dilakukan *decision-maker* (Humphreys&Trotman, 2011). Adanya bias pengukuran umum juga muncul karena keterbatasan tingkat rasionalitas dari

decision-maker untuk memahami pengukuran unik. Oleh karena itu, pengukuran umum menjadi indikator utama dalam penilaian kinerja.

Lipe dan Salterio (2000), mengatakan konsekuensi dari bias pengukuran dapat menghilangkan manfaat *balance scorecard*. Selain itu, strategi unit bisnis terabaikan dan menimbulkan ketidakadilan. Konsekuensi bias pengukuran umum juga berdampak pada pengambilan keputusan penting dalam perusahaan. Bias muncul pada penilaian kinerja, setelah itu penilaian kinerja dipakai sebagai alat ukur dalam mengambil keputusan. Dengan penilaian kinerja yang menyesatkan, maka keputusan yang diambil tidak tepat. Hal tersebut berpengaruh terhadap kinerja unit bisnis dan kinerja perusahaan di masa depan. Oleh karena itu, bias pengukuran umum harus dikurangi supaya penilaian kinerja optimal dan dapat meningkatkan kinerja perusahaan.

Pendekatan *Disaggregated* or *Mechanically Aggregated Balance Scorecard*

Einhorn (1972) menjelaskan bahwa untuk meningkatkan akurasi dari keputusan yang diambil, maka informasi yang diperoleh dibentuk secara kuantitatif dan output yang dihasilkan menggunakan kombinasi aturan mekanis. Kombinasi antar manusia dan sebuah model dengan menggunakan perpaduan klinis yaitu saat output dari model yang dibentuk menjadi input dalam pengambilan keputusan oleh *decision-maker* (Bowman, 1963).

Pendekatan *disaggregated/mechanically aggregated* membantu dalam menilai kinerja dengan cara memisahkan pengukuran *balanced scorecard* menjadi beberapa bagian. Pemisahan ini agar orang lebih fokus dalam penilaian kinerja. Selain itu, orang terbatas dalam mengelola informasi yang kompleks, oleh karena itu memisahkan pengukuran *balanced scorecard* akan menghindari penggunaan ukuran yang kompleks sehingga pengukuran yang digunakan tidak terabaikan. Hal tersebut agar penggunaan pengukuran umum dan

pengukuran unik dapat dipertimbangkan dalam penilaian.

Pendekatan yang disarankan untuk *balance scorecard* oleh Einhorn (1972) dan Bowman (1963) dilakukan dengan dua langkah. Langkah pertama yaitu dengan memisahkan keputusan evaluasi menjadi beberapa keputusan yang lebih kecil. *Decision-maker* dapat memahami pengukuran kinerja yang kompleks karena tiap pengukuran telah dibagi menjadi beberapa bagian. Dengan bentuk seperti ini, perhatian dari *decision-maker* dan kapasitas untuk memproses pengukuran kinerja mempermudah upaya yang dilakukan agar *decision-maker* menggunakan seluruh ukuran dalam menilai kinerja unit bisnis.

Langkah berikutnya adalah menentukan bobot untuk menjumlahkan hasil evaluasi untuk mendapatkan angka total dari hasil evaluasi. *Decision-maker* juga dapat membuat analisa secara menyeluruh untuk ukuran umum dan ukuran unik dengan lebih mudah karena pengukuran sudah dipisah menjadi beberapa bagian. Hal tersebut dapat memperkuat pentingnya ukuran unik dan ukuran umum dalam pengukuran kinerja.

Berdasarkan hal tersebut dapat kita buat hipotesis pertama sebagai berikut,

H1: Pendekatan *Disaggregated* or *Mechanically Aggregated Balance Scorecard* mengurangi bias pengukuran umum.

Pendekatan Pengetahuan tentang Balance Scorecard

Kesuksesan dalam menerapkan *Balance Scorecard* dibutuhkan pengetahuan dan pemahaman tentang *Balance Scorecard* (Niven, 2002). Bonner (1990) menunjukkan bahwa orang yang memiliki tingkat pengetahuan yang lebih akan mengukur secara berbeda dibandingkan orang dengan tingkat pengetahuan yang lebih rendah. Dalam penelitian tentang keputusan audit (Bonner, 1990; Bonner dan Lewis, 1990, Libby dan Tan 1994), *decision-maker* yang memiliki pengetahuan tentang suatu tugas yang spesifik akan mempengaruhi pengambilan

keputusan dalam menilai hasil dari tugas tersebut.

Penelitian tentang audit mengemukakan bahwa pengalaman kerja dari seorang auditor berkaitan dengan pelatihan yang didapat oleh auditor tersebut (Bonner, 1990; Bonner dan Pennington, 1991).

Hasil dari penelitian Dilla&Steinbart (2005), menunjukkan *decision-maker* menggunakan ukuran umum dan ukuran unik dalam penilaian kinerja unit bisnis, tetapi *decision-maker* memberi bobot yang lebih tinggi pada pengukuran umum. Hal tersebut sejalan dengan penelitian Lipe&Salterio (2000) dimana ukuran umum menjadi fokus dari *decision-maker*. Selain itu, hasil yang didapat oleh Dilla&Steinbart (2005) mendukung penelitian tentang psikologi dan perilaku konsumen (Kivets dan Simonson, 2000; Markman dan Medin, 1995; Slovic dan McPhilanny, 1974; Zhang dan Markman, 1998, 2001) menunjukkan bahwa *decision-maker* tetap menitikberatkan pada ukuran yang umum dari pada ukuran unik walaupun *decision-maker* telah menggunakan kedua ukuran saat menilai kinerja. Penelitian yang dilakukan Dilla dan Steinbart (2005), memiliki hasil yang sama dengan dengan hasil penelitian Psikologi dan perilaku konsumen tersebut. Berdasarkan hal tersebut dapat kita buat hipotesis kedua sebagai berikut,

H2: Pendekatan Pengetahuan tentang Balance Scorecard mengurangi bias pengukuran umum.

Hubungan Balance Scorecard dengan Alokasi Kompensasi

Balance scorecard membantu dalam mengevaluasi kinerja unit bisnis. *Decision-maker* kemudian menggunakan hasil evaluasi kinerja sebagai dasar untuk pengambilan keputusan, keputusan yang diambil berkaitan dengan kompensasi yang akan diterima karyawan. Milkovich et all (2011), pengertian kompensasi adalah segala bentuk manfaat baik itu uang, barang, maupun jasa yang diterima karyawan sebagai bagian dari hubungan kerja antara perusahaan dan karyawan. Pemberian kompensasi akan

berpengaruh terhadap sikap dan perilaku karyawan (Milkovich&Newman, 2008).

Hasil dari evaluasi kinerja menggunakan balance scorecard juga berpengaruh terhadap pengambilan keputusan kompensasi (Kaplan&Norton, 1996). Namun, Kaplan&Norton (1996) tidak dapat menghubungkan antara balance scorecard dengan alokasi kompensasi. Robert et all (2004), memberikan terobosan dalam menentukan alokasi kompensasi menggunakan pendekatan disaggregated/mechanically aggregated dalam balance scorecard. Penggunaan disaggregated/mechanically aggregated balance scorecard dapat menghasilkan analisa keseluruhan evaluasi kinerja. Berdasarkan analisa keseluruhan evaluasi kinerja, keputusan kompensasi dapat diambil. Hasil penelitian Robert et all (2004), menunjukkan bahwa keputusan kompensasi sangat didukung oleh pendekatan disaggregated/mechanically aggregated dalam balance scorecard. Berdasarkan hal tersebut maka dapat dibuat hipotesis ketiga sebagai berikut,

H3: Evaluasi kinerja menggunakan balance scorecard berpengaruh terhadap alokasi kompensasi

METODE PENELITIAN

Dalam penelitian eksperimen ini, kami menggunakan partisipan yang mau berpartisipasi secara sukarela. Terdapat 40 orang partisipan yang bersedia dan semua partisipan adalah mahasiswa/i S1 Fakultas Ekonomi Universitas Kristen Petra di Surabaya. Partisipan yang ikut berpartisipasi harus memenuhi syarat penelitian yaitu telah mengikuti dan menyelesaikan mata kuliah akuntansi manajemen. Dalam mata kuliah Akuntansi Manajemen telah diajarkan tentang konsep Balance Scorecard, sehingga partisipan yang telah mengikuti dan menyelesaikan sudah memahami konsep balance scorecard.

Desain Eksperimen

Rancangan penelitian eksperimen ini mengadaptasi dari Lipe & Salterio (2000), Robert et al. (2004) dan Dilla & Steinbart (2005). Desain dalam penelitian ini adalah $2 \times 2 \times 2$, dengan dua *between-subject factors* (pengukuran umum dan unik) dan satu *within-subject factor* (divisi). *Between-subject factors* yang pertama adalah pengukuran umum, yaitu perbandingan kinerja antara divisi RadWear dan WorkWear pada pengukuran umum. Dan *between-subject factors* yang kedua adalah pengukuran unik, merupakan perbandingan kinerja antara divisi RadWear dan WorWear pada pengukuran unik. Kemudian, untuk mengetahui apakah bias pengukuran umum dapat berkurang yaitu, dengan melihat interaksi antara *between subjeet factor* (pengukuran umum dan unik) dan *within subject factor* (divisi). Jika pengukuran umum dan unik menunjukkan hasil yang signifikan maka dapat dikatakan bahwa pengukuran tersebut telah digunakan dalam penilaian kinerja divisi.

Setiap partisipan akan mengevaluasi kinerja dari kedua manajer divisi, dan divisi merupakan *within-subject factor* sehingga komparabilitas penelitian akan maksimal karena penelitian menggunakan faktor yang sama. Dalam melakukan penelitian eksperimen ini, pengetahuan tentang Balance Scorecard tidak dikontrol secara langsung. Karena peneliti telah melakukan pemilihan terhadap partisipan eksperimen yaitu partisipan yang telah mengikuti dan menyelesaikan mata kuliah akuntansi manajemen.

Partisipan kemudian akan dikumpulkan dalam suatu ruangan. Sebelum partisipan mengerjakan kasus, 50% partisipan akan diberi review tentang balance scorecard dan 50% partisipan tidak diberi review tentang balanced scorecard. Partisipan yang akan diberi review akan masuk ke dalam ruangan dan diberi review, sedangkan partisipan yang tidak diberi review akan menunggu diluar ruangan hingga review yang dilakukan selesai. Setelah itu, 50% sisa partisipan dapat bergabung dengan partisipan yang sudah berada di dalam ruangan. Partisipan kemudian diminta

berperan sebagai eksekutif senior dari WCS *Incorporated*, dan kemudian menilai kinerja manager masing-masing divisi perusahaan secara bergantian.

Seluruh partisipan akan diberikan dan diminta untuk membaca kasus mengenai WCS *Incorporated*, dimana kasus ini menjelaskan sebuah perusahaan ritel dengan fokus utama mereka untuk memproduksi pakaian perempuan. Selama waktu membaca partisipan akan diberikan sejarah tentang WCS *Incorporated*. Dalam kasus tersebut partisipan juga diberikan informasi mengenai misi WCS,

"Kami akan menjadi penyalur pakaian yang luar biasa di setiap tingkat khusus yang dilayani oleh WCS."

serta informasi mengenai dua divisi WCS, yaitu RadWear (RAD) yang khusus untuk pakaian remaja, dan WorkWear (WORK) yang khusus untuk seragam kerja wanita. Dua divisi WCS memiliki strategi di masing-masing divisi.

Dalam kasus tersebut, kedua divisi memiliki scorecard masing-masing. Setiap *scorecard* dari masing-masing divisi terdiri dari enam belas pengukuran yang terpisah, masing-masing empat pengukuran dalam setiap empat kategori (*learning & growth, internal business process, customer, financial*). Dalam setiap kategori terdapat dua pengukuran yang umum di seluruh divisi dan dua pengukuran yang unik untuk masing-masing divisi.

Kemudian partisipan akan diberi instruksi untuk mengerjakan scorecard pada divisi RadWear (RAD) terlebih dahulu. Partisipan diminta untuk menyelesaikan dua langkah dari *Disaggregated/Mechanically Agregated* BSC yaitu; Langkah 1, partisipan menilai kinerja manajer RadWear (RAD), menggunakan skala dari 0 (*unacceptable*) sampai 100 (*excellent*),

Unacceptable : Tidak cukup memadai

Very poor : Jauh di bawah ekspektasi

- Poor* : Sedikit di bawah ekspektasi, memerlukan beberapa perbaikan
- Average* : Sesuai dengan ekspektasi
- Good* : Sedikit di atas ekspektasi
- Very Good* : Jauh di atas ekspektasi
- Excellent* : Jauh melampaui ekspektasi, manajer unggul.

Langkah 2, partisipan menjumlahkan bobot nilai untuk membuat total, mengagregatkan nilai pada setiap divisi. Kemudian setelah partisipan melakukan penilaian menyeluruh untuk masing-masing dari kinerja manajer, partisipan diminta untuk mengalokasikan total bonus akhir tahun sebesar Rp 100.000.000,- diantara dua manajer divisi. Alokasi tersebut digunakan untuk menguji H3.

HASIL PENELITIAN DAN PEMBAHASAN

Partisipan pada penelitian eksperimen ini adalah Mahasiswa/i Fakultas Ekonomi Universitas Kristen Petra kelompok Partisipan 1. Kelompok Partisipan 1 merupakan mahasiswa/i dari Program Manajemen Keuangan sebanyak 35 orang. Dalam penelitian ini partisipan dibagi kedalam dua kelompok yaitu Kelompok yang diberi Pengetahuan dan Kelompok yang tidak diberi Pengetahuan. Kemudian partisipan juga dibagi kedalam tiga kriteria yaitu sedang mengambil mata kuliah Akuntansi Manajemen (AM), sedang mengulang mata kuliah AM, dan sudah lulus mata kuliah AM.

Dalam pelaksanaannya, eksperimen dilakukan dua kali. Eksperimen pertama (09/06'2014) dilakukan dengan jumlah partisipan yang datang tidak mencapai target sehingga data yang diperoleh juga tidak banyak. Oleh karena itu, Ekperimen kedua dilakukan pada tanggal (17/06'2014), dalam eksperimen yang kedua jumlah partisipan lebih banyak dari ekperimen pertama sehingga data yang terkumpul dapat dipakai untuk melanjutkan penelitian ini.

Data Responden dapat dilihat, dari 35 orang responden terdapat 18 orang responden Perempuan dan 17 orang responden Laki-laki. Dari 18 orang responden perempuan, 10 orang responden termasuk dalam kelompok yang diberi pengetahuan dengan kriteria 40% sedang mengambil mata kuliah AM, 60% sudah lulus mata kuliah AM. Sedangkan 8 orang lain merupakan kelompok yan tidak diberi pengetahuan dengan kriteria sedang mengambil dan sedang mengulang AM sebesar 12,5% dan yang sudah lulus 75%.

Pada 17 responden Laki-laki, 10 orang termasuk dalam kelompok yang diberi pengetahuan dengan kriteria sedang mengambil mata kuliah AM 20% dan yang sudah lulus 80%. Sedangkan, 7 orang termasuk ke dalam kelompok yang tidak diberi pengetahuan dengan kriteria sedang mengambil mata kuliah AM 14,3%, sedang mengulang mata kuliah AM 28,6%, dan yang sudah lulus 57,1 %.

Tests of Between-Subjects Effects^a

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	33892.996	1	33892.996	2450.561	.000
Umum	.705	1	.705	.051	.822
Unik	37.969	1	37.969	2.745	.103
Umum * Unik	2708.070	1	2708.070	195.801	.000
Error	774.520	56	13.831		

a. Kode = Non_Pengetahuan

Tests of Within-Subjects Contrasts^a

Source	Division	Type III Sum of Squares	df	Mean Square	F	Sig.
Division	Linear	37.969	1	37.969	2.745	.103
Division *	Linear	2708.070	1	2708.070	195.801	.000
Division *	Linear	33892.996	1	33892.996	2450.561	.000
Division *	Linear	.705	1	.705	.051	.822
Error(Division)	Linear	774.520	56	13.831		

a. Kode = Non_Pengetahuan

Tests of Between-Subjects Effects^a

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.
Intercept	37650.803	1	37650.803	1440.480	.000
Umum	1.791	1	1.791	.069	.794
Unik	30.862	1	30.862	1.181	.281
Umum *	2900.124	1	2900.124	110.956	.000
Error	1986.464	76	26.138		

a. Kode = Pengetahuan

Tests of Within-Subjects Contrasts^a

Measure: MEASURE_1

Source	Division	Type III Sum of Squares	Df	Mean Square	F	Sig.
Division	Linear	30.862	1	30.862	1.181	.281
Division * Umum	Linear	2900.124	1	2900.124	110.956	.000
Division * Unik	Linear	37650.803	1	37650.803	1440.480	.000
Division * Umum * Unik	Linear	1.791	1	1.791	.069	.794
Error(Division)	Linear	1986.464	76	26.138		

a. Kode = Pengetahuan

Independent Samples Test

	Levene's Test for Equality of Variances	t-test for Equality of Means								
								95% Confidence Interval of the Difference		
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Umum	Equal variances assumed	.148	.703	-1.588	33	.122	-3.89150	2.45074	-8.87756	1.08456
	Equal variances not assumed			-1.575	29.358	.126	-3.89150	2.47079	-8.94216	1.15816
Unik	Equal variances assumed	2.814	.103	-1.585	33	.127	-7.83000	5.00842	-18.02572	2.36372
	Equal variances not assumed			-1.653	32.720	.108	-7.83000	4.74114	-17.48505	1.81305

Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.788 ^a	.620	.570	.36823

a. Predictors: (Constant), Work_unik, Work_umum, Rad_unik, Rad_umum

b. Dependent Variable: Rasio_Alokasi

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	.850	.426		1.995	.055		
	Rad_umum	.072	.028	.560	2.551	.016	.263	3.807
	Rad_unik	.040	.015	.531	2.609	.014	.305	3.278
	Work_umum	-.068	.021	-.477	-3.271	.003	.595	1.680
	Work_unik	-.035	.011	-.560	-3.320	.002	.445	2.248

a. Dependent Variable: Rasio_Alokasi

Pengujian hipotesis 1 dilakukan berdasarkan kelompok yang diberi

pengetahuan tentang BSC dan kelompok yang tidak diberi pengetahuan tentang BSC. Tabel 4.5 menunjukkan hasil pengujian hipotesis 1 pada kelompok yang diberi pengetahuan tentang BSC. Hasil pengujian hipotesis 1 berdasarkan kelompok yang diberi pengetahuan, pada tabel *between subject*, interaksi antara pengukuran umum dan pengukuran unik menunjukkan hasil yang signifikan (umum*unik, sig<0,05). Pada tabel *within subject*, penilaian kinerja divisi jika dikaitkan dengan pengukuran umum menunjukkan hasil yang signifikan (divisi*common, sig<0,05). Hasil yang sama juga terjadi ketika pengukuran unik dikaitkan dengan penilaian kinerja divisi (divisi*unik, sig<0,05). Hasil yang serupa juga ditemukan pada kelompok yang tidak diberi pengetahuan (tabel 4.6). Dapat disimpulkan bahwa pada kelompok yang diberi pengetahuan dan yang tidak diberi pengetahuan, responden menggunakan pengukuran umum dan pengukuran unik dalam menilai kinerja divisi. Berdasarkan hasil pengujian tersebut, hipotesis pertama dalam penelitian ini dapat diterima. Dengan kata lain, pendekatan *Dissagregated/Mechanically Aggregated* dapat mengurangi bias pengukuran umum dalam Balanced Scorecard.

Uji *Independent Sample T-test* yang digunakan untuk menguji hipotesis 2. Dari hasil pengujian, evaluasi kinerja dengan menggunakan pengukuran umum dan unik pada kelompok yang diberi pengetahuan dan kelompok yang tidak diberi pengetahuan menunjukkan nilai yang tidak signifikan (sig>0,05). Dari hasil ini maka dapat disimpulkan bahwa adanya pengetahuan tentang BSC tidak memberikan perbedaan yang signifikan pada evaluasi kinerja manajer baik dengan menggunakan pengukuran umum maupun pengukuran unik oleh partisipan penelitian, karena partisipan tetap memperhatikan kedua pengukuran baik yang diberi pengetahuan dan yang tidak diberi pengetahuan. Berdasarkan hal tersebut hipotesis 2 tidak dapat diterima.

Hipotesis 3 diuji dengan menggunakan analisis regresi dengan variabel dependen berupa Y rasio, dimana Y

rasio merupakan rasio pembagian antara alokasi bonus divisi RadWear dan alokasi bonus divisi WorkWear. Variabel independen yang digunakan adalah Rad Umum, Rad Unik, Work Umum, dan Work Unik. Rad Umum dan Rad Unik adalah bagian dari skor agregat dari divisi RadWear. Sedangkan Work Umum dan Work Unik adalah bagian dari skor agregat divisi WorkWear. Dari hasil uji regresi linear berganda, menunjukkan pengaruh yang signifikan pada seluruh variabel independen yang digunakan. Selain itu kita dapat melihat pada R Square sebesar 62%. Dapat disimpulkan bahwa skor agregat pada masing-masing divisi memiliki pengaruh yang besar untuk mendasari keputusan alokasi kompensasi di kedua divisi tersebut. Oleh karena itu, hipotesis ketiga dapat diterima, yaitu evaluasi kinerja dengan BSC terbukti mempengaruhi keputusan alokasi kompensasi.

KESIMPULAN, KETERBATASAN, DAN SARAN

Berdasarkan hasil pengujian yang telah dilakukan, maka dapat diperoleh beberapa kesimpulan. Hipotesis 1 pada penelitian ini diterima, hal ini berarti pendekatan *disaggregated/mechanically aggregated* dapat mengurangi bias pengukuran umum. Hipotesis 2 pada penelitian ini tidak diterima, hal ini berarti pendekatan pengetahuan tidak dapat mengurangi bias pengukuran umum. Hipotesis 3 pada penelitian ini diterima, hal ini berarti penilaian kinerja menggunakan BSC mempengaruhi alokasi kompensasi.

Saran yang dapat diberikan terkait penelitian ini adalah dalam penelitian selanjutnya, diharapkan peneliti dapat menggunakan partisipan yang mempunyai pengalaman menggunakan BSC dalam penilaian kinerja perusahaan, sehingga pengetahuan tentang BSC lebih baik. Peneliti selanjutnya diharapkan dapat meninjau ulang tingkat pengetahuan tentang Balanced Scorecard melalui pelatihan-pelatihan mengenai penerapan Balanced Scorecard.

Keterbatasan dalam penelitian ini adalah sebagai berikut: Partisipan

eksperimen merupakan mahasiswa Jurusan Manajemen Keuangan. Dalam eksperimen ini partisipan dibagi ke dalam tiga kriteria yaitu partisipan yang minimal sedang mengambil mata kuliah akuntansi manajemen, mengulang mata kuliah Akuntansi Manajemen, atau sudah lulus mata kuliah Akuntansi Manajemen. Pengetahuan tentang Balanced scorecard yang diperoleh partisipan hanya berupa teori dan belum ada pelatihan tentang cara menggunakan Balanced Scorecard dalam penilaian kinerja

DAFTAR REFERENSI

- 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. and B. L. Lewis. 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., Kilough, L. N., & Brown, R. N. (2005). Intolerance of Ambiguity Effect of Evaluation with The Balances Scorecard. *The Innovation Congress: Issues in Innovation*, vol. 3, issue 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.
- Dumond, E. J. (1994). Making Best Use of Performance-Measures and Information. *International Journal of Operation & Production Management* 14(9), pp. 16-31
- Edwards, W. . and J. R. Newman. 1982. *Multiaattribute Evaluation*. Beverly Hills, CA: Sage Publications.
- Einhorn, H. J. 1972. Expert Measurement and Mechanical Combination. *Organizational Behavior and Human Decision Process* 19 (Feb): 86-106.
- Frijo, M. L. and Krumwiede, K. R. (1999). Balanced Scorecards: A Rising Trend in Strategic Performance Measurement. *Journal of Strategic Performance Measurement* 3(1), pp. 42-48
- 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.
- 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. (1993). Putting the Balanced Scorecard to Work. *Harvard Business Review* (September-October): 134-147.

- 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.
- Kennedy, J. (1995). Debiasing the Curse of Knowledge in Audit Judgment. *The Accounting Review*, 70(2), 249-273.
- Lawrie, G. and Cobbold, I. (2004). Evolution of the Balanced Scorecard into An Effective Strategic Performance Management Tool. *Development of the 3rd Generation Balanced Scorecard*.
- Lawson, R., Stratton, W., and Hatch, T. (2003). The Benefits of a Scorecard System. *CMA Management June/July*, pp. 24-26
- 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.
- Lingle, J. H. And Schiemann, W. A. (1996). From The Balanced Scorecard to Strategic Gauges: Is Measurement worth it?. *Management Review*, 85 (3), pp. 56-61
- 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., and 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. and 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. The Role of Knowledge in R&D Efficiency. *Oxford University Press, 97(3)*, 453-470.
- Niven, P. R. 2002. *Balanced Scorecard Step-by-Step Maximizing Performance and Maintaining Results*. New York. NY: John Wiley & Sons.
- Roberts, M. L., Albright, T. L., & Hibbets, A. R. (2004). Debiasing Balanced Scorecard Evaluations. *Behavioral Research in Accounting, 16(1)*, 75-88.
- Sandt, J., Schaeffer, U., and Weber, J. (2001). Balanced Performance Measurement System and Manager Satisfaction – Empirical Evidence from a German Study. *WHU-Otto Beisheim Graduated School of Management*.
- Tversky, A. and Kahneman D. (1973). Availability: A Heuristic for Judging Frequency and Probability. *Cognitive Psychology, 5*, 207-232.
- Tversky, A. and 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*.