

Het meten van gebruik van de balanced scorecard: wegen naar de evaluatie van een management accounting innovatie

M. Gelderman¹

1 Introductie

In 1987 werd de discipline management accounting door Johnson en Kaplan nog zo goed als dood verklaard. De jaren daarna toonden dat dit vakgebied meer levenskracht bezat dan menigeen had vermoed: de nieuwe management accounting innovaties volgden elkaar in hoog tempo op. Alhoewel de diffusie van sommige innovaties een punt van aandacht blijft (Schoute, 1999), kunnen we stellen dat technieken als activity-based costing, economic value addedTM en de balanced scorecard (BSC) wel degelijk hun impact hebben op de praktijk. Kijken we naar het aandachtspunt van dit artikel, de BSC, dan wordt deze claim bevestigd door meerdere empirische onderzoeken (Frigo en Krumwiede, 2000; Ittner en Larcker, 1998).

Met het aanvaarden van deze management accounting innovaties in de praktijk, wordt de beantwoording van de vraag of zij daadwerkelijk bijdragen aan het succes van de adopterende organisatie belangrijk. In de wetenschappelijke literatuur is dan ook bij herhaling opgeroepen tot het uitvoeren van dit type onderzoek voor management accounting innovaties in het algemeen (Atkinson et al., 1997) en voor de BSC in het bijzonder (Ittner en Larcker, 1998). Ook Kaplan (1998, p. 95) geeft aan dat het de taak van management accounting wetenschappers is 'to refute the null hypothesis that changes in existing management accounting systems cannot produce improved performance.'

Dit artikel gaat in op het evalueren van het succes van de BSC. Aan de hand van de system-to-value chain van Doll en Torkzadeh (1991)

betoog ik dat het gebruik van de BSC niet alleen een doel op zich is, maar dat wanneer gebruik op adequate wijze wordt gemeten, de mate van gebruik tevens bruikbaar is als benadering voor de bijdrage van de BSC aan de prestaties van de organisatie. Voor het langs meerdere dimensies meten van het gebruik van management informatie systemen is door Doll en Torkzadeh (1998) een vragenlijst (meetinstrument) ontwikkeld. Dit instrument dient om vast te stellen in welke mate een systeem wordt gebruikt voor decision making, work integration en customer service. Met dit instrument is in een negental organisaties het gebruik van de BSC door managers gemeten. In het artikel wordt een analyse van de antwoorden op dit meetinstrument gepresenteerd. Deze analyse richt zich op het vaststellen van de onderliggende dimensies van BSC-gebruik en het beoordelen van de betrouwbaarheid van het meetinstrument. Ik besluit het artikel met suggesties voor verder onderzoek.

2 Evaluatie van de balanced scorecard

Uit de wens het succes van management accounting innovaties te beoordelen, volgt automatisch de volgende vraag: hoe kunnen we een relatie laten zien tussen veranderingen in een management accounting systeem - hier de introductie van de BSC - en performance. Als wetenschappers zouden we natuurlijk het liefst een veldexperiment uitvoeren (Brownell, 1995). We maken twee willekeurige groepen van organisa-

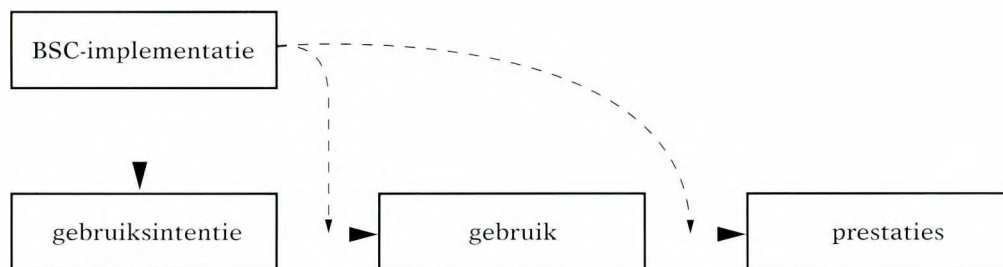
¹ M. Gelderman is als universitair hoofddocent management accounting werkzaam bij de Vrije Universiteit Amsterdam. Tevens is hij verbonden aan het Limperginstituut

ties of organisatieonderdelen en in de ene groep voeren we een BSC in, terwijl we dat in de andere achterwege laten. Afgezien van potentiële problemen met de interne validiteit, zal het iedereen in eerste oogopslag duidelijk zijn dat dit voor de wetenschappers weliswaar een potentieel Walhalla oplevert, maar dat de bereidheid uit de praktijk om aan dit type onderzoek mee te werken erg gering zal zijn.

Een van de oplossingen die is aangedragen voor dit probleem is het uitvoeren van action research. De onderzoeker werkt zelf mee aan een implementatie en door zijn nauwe betrokkenheid bij die implementatie is de onderzoeker niet alleen in staat de merites van het nieuwe systeem te beoordelen, maar ook te werken aan de verdere ontwikkeling van de innovatie (Kaplan, 1998). De vraag die zich bij deze onderzoeksstrategie opdringt is of de onderzoeker in dit geval nog wel als een onpartijdige beoordelaar van het succes van de innovatie kan worden beschouwd. Zoals Hopwood (1974, p. 78) vijftientig jaar geleden al wist op te merken: 'What a number of studies gain in persuasiveness, they loose in objectivity.' Daarbij komt dat, zelfs als de verbetering in performance objectief zou kunnen worden vastgesteld, het niet duidelijk is of deze is veroorzaakt door de management accounting innovatie, of door de aanwezigheid van de consultant/wetenschapper/action researcher in de organisatie. Ik kan mij vrij goed voorstellen dat als de heren Kaplan en Norton een organisatie binnenkomen, het niet eens zo veel meer uit maakt wat voor innovatie zij invoeren. Hun naamsbekendheid en hun kwaliteiten als consultant zullen er vermoedelijk voor zorgen dat de implementatie van vrijwel iedere willekeurige vernieuwing tot verbetering van de prestaties leidt. De kritieke onderzoeksvraag is echter of, door de bank genomen, invoering van de BSC leidt tot verbetering en die vraag wordt op deze wijze niet beantwoord. Ten opzichte van action research vormen 'vanaf enige afstand' uitgevoerde casestudies een duidelijke verbetering, het veelal kwalitatieve karakter van dergelijke studies maakt naar mijn mening het trekken van harde conclusies echter problematisch.

Kiezen we voor een kwantitatieve benadering, dan zijn twee manieren van onderzoek naar de bijdrage van bijvoorbeeld de BSC aan het succes van een organisatie mogelijk. Allereerst kan tijdserie-analyse gebruikt worden om vast te stellen of invoering van een BSC leidt tot een trendbreuk in een prestatie maatstaf. Een punt van aandacht is hierbij overigens de keuze van die prestatie maatstaf: deze dient niet zo gekozen te worden dat de enige conclusie van het onderzoek kan zijn dat prestaties op die maatstaven die in de BSC zijn opgenomen - ten gevolge van manipulatie? - verbeteren². Gebruik van een niet-manipuleerbare maatstaf als aandeelhoudersrendement lijkt echter redelijk valide. Een tweede mogelijkheid is het uitvoeren van cross-sectioneel onderzoek waarbij we de prestaties van ondernemingen met en zonder BSC vergelijken. Voor de gekozen prestatie maatstaf geldt hetzelfde probleem als hiervoor genoemd. Verder zijn in verband met survival- en selectie-effecten vergelijkbare correcties in de analyse nodig als Shaver (1998) voorstelt voor het evalueren van strategie-onderzoek.

Kaplan (1998) is van beide hiervoor genoemde werkwijzen niet gecharmeerd en stelt zelfs dat zij ongewenst zijn. Zij houden geen rekening met de vraag of de implementatie valide is en gaan ook niet in op de vraag hoe goed de BSC door de adopterende organisatie wordt gebruikt. Helaas voorziet de door Kaplan gekozen formulering niet direct in de door hemzelf eerder in het aangehaalde artikel genoemde behoefte aan een verwerpbare nul-hypothese. De begrippen 'valide' en 'goed' blijven ongedefinieerd. Zo lang niet duidelijk is wat een 'valide' implementatie en 'goed' gebruik van de BSC is, valt niet te testen of de BSC bijdraagt aan prestaties van de organisatie. Immers de meest voor de hand liggende definitie van goed gebruik van een valide implementatie is dat de implementatie bijdraagt aan de prestaties en dat leidt tot cirkelredeneringen waar we niet verder mee komen. Om in dit proces vooruitgang te boeken, zijn een uitwerking van het begrip valide implementatie (bijvoorbeeld in de vorm van een checklist, waarmee het mogelijk is een validiteitscore vast te stellen) en het begrip goed gebruik noodzakelijk. In dit artikel concentreer ik me op het laatstgenoemde aspect.



Figuur 1: De relatie tussen de kenmerken van de balanced scorecard (BSC) implementatie, gebruiksintentie, gebruik van de balanced en prestaties van de organisatie

3 The system-to-value chain: de relatie tussen gebruik en prestaties

Een belangrijke reden om de aandacht in eerste instantie op dit tweede en niet op het eerste aspect te concentreren, is dat we in redelijkheid mogen aannemen dat, als de BSC bijdraagt aan betere prestaties van de organisatie, deze, zoals is weergegeven in figuur 1, het meest direct gerelateerd zullen zijn aan gebruik. De ontwerpkenmerken beïnvloeden weliswaar de invloed die het systeem op de prestaties van de gebruikers heeft, maar doen dit slechts indirect. Om invloed te kunnen hebben, dient de BSC daadwerkelijk te worden gebruikt. Van gebruikers (en zeker wanneer dit managers zijn) mogen we verwachten dat zij in staat zijn in te schatten of gebruik van de BSC een positieve invloed op hun prestaties zal hebben³. Indien de manager van mening is dat gebruik van het systeem voor hem/haar positieve gevolgen zal hebben, zal dit leiden tot een gebruiksintentie, die sterker zal zijn naarmate de verwachte baten van gebruik groter zijn. De gebruiksintentie zal in het algemeen weer leiden tot gebruik van het systeem zelf. Het gebruik van het systeem is de factor die uiteindelijk tot betere prestaties van de organisatie moet leiden. De keten van causale factoren (in dit geval kenmerken van de BSC-implementatie), naar beliefs en attitude (in dit geval samengevat in gebruiksintentie), naar gedrag (gebruik), naar impact (prestaties) is systematisch weergegeven in figuur 1. Door Doll en Torkzadeh (1991, 1998) wordt deze keten aangeduid met de term system-to-value chain.

Overigens dient hier wel te worden opgemerkt dat de relatie tussen gebruiksintentie en gebruik en de relatie tussen gebruik en performance kan

worden beïnvloed (gemodereerd) door implementatiekenmerken. Zo kan gebruik van het systeem de facto verplicht zijn of juist vanwege implementatieproblemen onmogelijk blijken. In dit soort situaties kan het gebeuren dat de relatie tussen gebruiksintentie en gebruik niet bestaat. Eveneens zijn situaties denkbaar waarin gebruik niet gerelateerd is aan prestaties. Traditioneel wordt bij informatiesysteemonderzoek bijvoorbeeld vaak gewerkt met een gebruiksmaatstaf die gebaseerd is op duur of frequentie van gebruik. Situaties waarin beide variabelen stijgen zonder dat sprake is van een prestatieverbetering zijn zeer wel denkbaar. Dit probleem kunnen we echter, evenals het vorige, grotendeels ondervangen door keuze van een 'slimmere' gebruiksmaatstaf.

Als terzijde zij hier opgemerkt dat de meting van gebruik niet alleen gehanteerd kan worden voor het leggen van een relatie met performance. Gegeven bovengenoemde overwegingen kan gebruiksintentie ook worden beschouwd als een proxy voor de kwaliteit van de BSC. De eerder genoemde 'checklist voor kwaliteit van de implementatie' zou vastgesteld kunnen worden door kenmerken van de implementatie te relateren aan scores op dit gebruiksinstrument. De score op het gebruiksinstrument zal immers in geringere mate dan de ondernemingsprestaties beïnvloed worden door in dit kader niet relevante factoren. De geringere omvang van de storings-term maakt het mogelijk een zuiverder inschatting van het belang van verschillende implementatiekenmerken te maken.

4 Keuze van een gebruiksmaatstaf

Hierboven werd al een eerste keer verwezen naar de informatiesysteem-literatuur. In deze hoek is veel onderzoek verricht naar het succes van informatiesystemen en de keuze van geschikte variabelen in dergelijk onderzoek (DeLone en

McLean, 1992). Gezien de grote overeenkomsten tussen de BSC en management informatie systemen, lijkt het onverstandig om als management accounting onderzoekers zelf het wiel opnieuw uit te vinden. Het is wel zo handig om leentjebuur te spelen bij dit vakgebied en te leren van de daar gemaakte fouten en opgedane ervaring. Als startpunt voor het vinden van een slimme gebruiksmaatstaf is dan ook gebruikgemaakt van de literatuur uit dit vakgebied.

Een eerste conclusie uit deze literatuur is dat het meten van frequentie en duur van gebruik - bij geautomatiseerde systemen overigens wellicht een stuk gemakkelijker te realiseren dan bij een BSC - geen goed idee is. Gemeten in deze termen leidt meer gebruik niet noodzakelijkerwijze tot een verbetering van de prestaties (Gelderman, 1998). Een alternatieve benadering is dan ook recentelijk voorgesteld door Doll en Torkzadeh (1998). Deze auteurs ontwikkelden een meetinstrument dat 'nuttig' gebruik meet langs meerdere dimensies. Er wordt in deze vragenlijst niet gevraagd hoe lang of hoe vaak een informatiesysteem wordt gebruikt, maar in hoeverre de respondent het eens is met de stelling dat hij het systeem gebruikt voor een bepaald doel. Een voorbeeld van een dergelijke stelling is: 'Ik gebruik [het systeem] om het besluitvormingsproces rationeler te maken'. In totaal bestaat het volledige Doll en Torkzadeh-instrument uit 30 van dit soort vragen. Samen dekken deze 30 vragen een drietal dimensies af.

- 1 *Decision making*: deze dimensie bestaat in totaal uit dertien vragen. Van deze vragen hebben er zes betrekking op gebruik ten behoeve van het oplossen van problemen ('use to analyze cause and effect relationships') en zeven op 'decision rationalization': het uitleggen, verdedigen en meer rationeel maken van beslissingen ('to improve decision making processes or explain/justify the reasons for decisions').
- 2 *Work integration*: deze dimensie bestaat in totaal uit twaalf vragen. Vier vragen hebben betrekking op horizontale integratie en acht vragen hebben betrekking op verticale integratie. De vragen richten zich op zaken als communicatie, coördinatie, plannen, monitoren, etc.
- 3 *Customer service*: deze dimensie bestaat uit vijf vragen. Zij hebben allemaal betrekking op

het gebruik van het systeem voor het verbeteren van de dienstverlening aan klanten.

De aan dit meetmodel ten grondslag liggende claim, is dat deze drie (of vijf als we beide subdimensies meetellen) dimensies volstaan voor het beschrijven van gebruik. Ook voor een ten bate van management control gebruikt systeem als de BSC zullen deze drie dimensies moeten volstaan. Later in dit artikel zal deze claim, voorzover mogelijk, met behulp van een factoranalyse op het instrument, worden geverifieerd.

5 Gegevensverzameling

In overleg met praktijkdeskundigen werd van alle dertig vragen vastgesteld of zij van toepassingen waren op de balanced scorecard, of zij begrijpelijk waren voor de doelgroep en hoe sterk de overlap met de andere vragen uit het meetinstrument was. Van het originele instrument werden acht items geëlimineerd. De belangrijkste reden voor eliminatie was gelegen in het feit dat de vragen niet van toepassing waren op de BSC, te veel overlap met andere vragen vertoonden, of niet automatisch voor alle respondenten van toepassingen waren. Bij het verwijderen van vragen werd er zorg voor gedragen dat per dimensie minstens drie items over bleven⁴. De vragen werden door de auteur dezes uit het Engels vertaald naar het Nederlands en door een native speaker - die dankzij enige jaren consulting-ervaring in Nederland bekend was met de gebruikte terminologie - terugvertaald naar het Engels. Waar nodig werden correcties in de vertaling aangebracht. Vervolgens werd het instrument gepilot-test onder vijf managers, die tevens commentaar op de vragen gaven. Aanpassingen bleken niet noodzakelijk, alhoewel er wel commentaar was op de overlap tussen verschillende vragen. De uiteindelijk gebruikte vragen zijn terug te vinden in tabel 1. De vragen over decision rationalization ten bate van het oplossen van problemen zijn gemerkt met een P (problem solving), die over decision rationalization met een R. De vragen over horizontale en verticale integratie enerzijds met respectievelijk een H en een V en de vragen over customer service anderzijds met een C.

Vervolgens is de vragenlijst uitgezet onder BSC-gebruikende managers in een negental organisaties⁵. Waar mogelijk werd de vragenlijst

aangepast aan de organisatie door niet de term 'balanced scorecard' te gebruiken, maar direct te verwijzen naar de naam van het in die organisatie geïmplementeerde systeem. De vragenlijsten werden uitgezet met medewerking van, en soms

door, de betrokken organisaties. Er is niet van alle organisaties bekend hoeveel lijsten zij hebben uitgezet, maar voor het merendeel van de organisaties (meer dan 95% van de terugontvangen enquêtes) zijn deze gegevens wel bekend. Hier

Tabel 1: Pattern matrix na oblimin rotatie van de extractie van een vijftal factoren met behulp van principale componenten-analyse (n = 79). Ladingen boven de 0.3 zijn cursief weergegeven, ladingen boven de 0.4 vet. De nummering komt overeen met de nummering van de originele vragenlijst. Omdat de originele vragenlijst langer is dan het hier gebruikte instrument ontbreken er nummers.

		<i>1 (PR)</i>	<i>2 (HC)</i>	<i>3 (VR)</i>
P1	Ik gebruik de BSC om te beslissen hoe ik een probleem het best kan benaderen	.478	-.113	.091
P4	Ik gebruik de BSC om mijn gedachtegang te verifiëren aan de hand van de gegevens	.858	.133	.020
P5	Ik gebruik de BSC om wijs te worden uit de gegevens	.617	.171	.223
P6	Ik gebruik de BSC om te analyseren waarom problemen optreden	.625	-.028	-.113
R6	Ik gebruik de BSC om de effectiviteit en de efficiëntie van het besluitvormingsproces te verbeteren	.667	-.225	-.056
R7	Ik gebruik de BSC om het besluitvormingsproces rationeler te maken	.612	-.164	.112
V3	Ik gebruik de BSC om mijn werk te plannen	.412	-.175	.182
C4	Ik gebruik de BSC om klanten op creatievere wijze van dienst te zijn	.552	-.362	-.031
C1	Ik gebruik de BSC om strategischer om te gaan met interne en/of externe klanten	.230	-.736	-.173
C2	Ik gebruik de BSC om interne en/of externe klanten van dienst te zijn	.350	-.560	.048
C3	Ik gebruik de BSC om de kwaliteit van de dienstverlening aan klanten te verbeteren	.345	-.457	.189
C5	Ik gebruik de BSC om informatie uit te wisselen met interne en/of externe klanten	-.053	-.747	.013
H1	Ik gebruik de BSC om te communiceren met andere mensen binnen mijn organisatie-eenheid	.000	-.539	.298
H2	Mijn organisatie-eenheid en ik gebruiken de BSC om onze activiteiten op elkaar af te stemmen	-.071	-.775	.144
H3	Ik gebruik de BSC om activiteiten met anderen in mijn organisatie-eenheid te coördineren	-.031	-.721	.101
V2	Ik gebruik de BSC om mijn eigen prestaties in de gaten te houden	.075	.098	.746
V4	Ik gebruik de BSC om te communiceren met de mensen die aan mij rapporteren	.109	-.278	.618
V5	Ik gebruik de BSC om te communiceren met de mensen aan wie ik rapporteer	-.211	-.154	.805
V8	Ik gebruik de BSC om feed back te krijgen over mijn eigen prestaties	.106	-.067	.660
R1	Ik gebruik de BSC om me te helpen mijn beslissingen uit te leggen	.000	-.091	.719
R2	Ik gebruik de BSC om me te helpen mijn beslissingen te verdedigen	.359	.040	.434
R3	Ik gebruik de BSC om me te helpen de redenen voor mijn beslissingen expliciet te maken	.393	.045	.473

varieert de response van 40% tot 72% (gewogen naar het aantal uitgezette enquêtes bedraagt de gemiddelde response 51%). In totaal werden 107 vragenlijsten geretourneerd, van deze vragenlijsten waren er 91 adequaat ingevuld⁶. Van deze 91 respondenten maakten er acht geen gebruik van de in hun organisatie geïmplementeerde BSC⁷, terwijl vier respondenten niet alle vragen van het gebruiksinstrument invulden of op een vraag meer dan één antwoord gaven. De resterende 79 vragenlijsten zijn gebruikt in de hierna gepresenteerde analyse.

6 Analyse

Middels factoranalyse kan vastgesteld worden welke onderliggende dimensies (ook wel aangeduid als latente variabelen of factoren) in de set van vragen (ook wel aangeduid als items of manifeste variabelen) gevonden kunnen worden. Voor het analyseren van dit meetinstrument is common factor analyse de aangewezen techniek (Hair, Jr. et al., 1998). Doll en Torkzadeh (1998) gebruiken echter principale componenten analyse. Alhoewel toepassing van deze techniek in principe onjuist is (Johnson, 1998), zullen de uitkomsten over het algemeen sterke overeenkomst vertonen met die van common factoranalyse (Hair, Jr. et al., 1998). Teneinde mijn analyse en die van Doll en Torkzadeh (1998) vergelijkbaar te houden, is een principale componenten analyse uitgevoerd. Doll en Torkzadeh (1998) voeren na de factorextractie een oblimin-rotatie uit omdat zij verwachten dat de verschillende gebruiksdimensies samenhang zullen vertonen. Ook dit voorbeeld is gevolgd⁸.

Het is a priori niet geheel duidelijk uit hoeveel factoren⁹ het gebruikte meetinstrument zal bestaan. Doll en Torkzadeh (1998) vonden drie factoren die overeenkomen met de drie eerder besproken dimensies. Zoals bij de bespreking van die dimensies al werd geconstateerd bestonden de eerste twee dimensies ieder uit twee subdimensies. A priori lijkt het dan ook niet onmogelijk dat voor het beschrijven van de gegevens vijf dimensies noodzakelijk zijn. Op deze grond werd besloten de kwaliteit van oplossingen met twee tot vijf factoren te onderzoeken.

De eerste uitkomsten van de principale componentenanalyse suggereren dat drie factoren niet volstaan voor het beschrijven van de gege-

vens. Vijf eigenwaarden liggen boven de één en een inspectie van de screenplot suggereert dat de meest aangewezen oplossing bestaat uit vier (volgens het smallest-large-eigenvalue criterium zoals weergegeven door Johnson, 1998) of vijf factoren (volgens de interpretatie van Hair, Jr. et al., 1998). Het criterium waarop de uitkomst van een factoranalyse uiteindelijk echter beoordeeld dient te worden is de mate waarin de oplossing leidt tot een duidelijke en interpreteerbare factorstructuur. De drie-factoroplossing blijkt hier het best te voldoen. In zijn totaliteit verklaart deze oplossing iets meer dan 54% van de variantie. De factorladingen zijn weergegeven in tabel 1. Er zijn geen items die met meer dan 0.40 op meer dan één factor laden. Vijf items laden met lading van boven de 0.30 (maar onder de 0.40) op een tweede factor. Deze verkeerd ladende items worden hieronder bij de betrokken factoren besproken. Alle primaire factorladingen liggen boven de 0.40.

Alhoewel de factorstructuur helder is, is de interpretatie niet volledig in overeenstemming met de uitkomsten van Doll en Torkzadeh. De eerste factor komt in hoofdlijnen overeen met de decision making factor uit het originele instrument. De problem solving component domineert deze factor echter. De rationalization component heeft twee items die primair laden. De overige drie items vormen samen met een viertal verticale communicatie-items een factor. Twee van deze drie items vertonen echter een kruislading van meer dan 0.3 op de eerste factor, hetgeen aangeeft dat er reden is om deze items bij beide factoren te laten horen¹⁰. Naast de genoemde items, belanden twee 'vreemde' items in deze factor. Bij het eerste item ('Ik gebruik de BSC om mijn werk te plannen') wekt dit geen verbazing. Classificatie binnen de dimensie decision making ligt gegeven de verwoording meer voor de hand dan classificatie binnen de dimensie verticale communicatie, zoals voorgesteld door Doll en Torkzadeh. Ten slotte is er de vraag uit de klantdimensie. Deze laadt onverwacht, en onverklaarbaar, op de eerste factor.

De tweede factor combineert de horizontale communicatiedimensie met de klantdimensie. Opvallend is dat drie van de vragen uit de klantdimensie betrekking hebben op interne en/of externe klanten. Allicht vormt het noemen van de

interne klanten een verklaring voor het onverwachte samengaan van beide dimensies. Het verschil tussen informatiewisseling met interne klanten en horizontale communicatie lijkt klein. Bij de andere items is het verschil groter, maar het ontstaan van verbanden zeker niet ondenkbaar. Twee vragen uit de klantdimensie laden tevens op de eerste factor, terwijl de klantdimensievraag uit de eerste factor tevens op de tweede factor laadt. Dit suggereert dat een vierde factor noodzakelijk is voor het adequaat beschrijven van de gegevens. Het uitvoeren van de analyse met vier factoren maakt, zoals aangegeven, de structuur echter niet helderder, de rotatie kampt zelfs met convergentieproblemen hetgeen duidt op een 'slechte' oplossing. Op het probleem van het samengaan van de klant- en horizontale communicatiedimensie kom ik hieronder terug.

De derde factor uit tabel 1 combineert items die betrekking hebben op rationalization van beslissingen en verticale communicatie¹¹. Bij Doll en Torkzadeh horen de genoemde items bij twee aparte dimensies. Bij gebruik van de BSC zijn verticale communicatie (onder andere het afleggen van verantwoording) en dat deel van rationalization dat te maken heeft met het verdedigen en

Samenvattend, is het opvallend dat het beschreven meetinstrument het wel anders, maar niet slechter doet dan in de originele toepassing. In de toepassing van Doll en Torkzadeh vormen horizontale en verticale communicatie enerzijds en decision making en problem solving anderzijds beide samen een enkele factor, terwijl te verwachten valt dat dit aparte dimensies zijn. In het hier beschreven onderzoek zijn verticale en horizontale communicatie wel als aparte dimensies te onderkennen. Problem solving en rationalization vormen weliswaar samen één factor, maar alleen de rationalization items zijn geassocieerd met de verticale communicatiefactor, hetgeen eveneens duidt op het bestaan van een aparte dimensie. De klantendimensie en de horizontale communicatiedimensie zijn daarentegen in het hier beschreven onderzoek weer niet van elkaar te onderscheiden. De vraagstelling waarbij naast externe ook interne klanten worden genoemd, vormt hiervoor een mogelijke verklaring.

In tabel 2 zijn nog enige beschrijvende statistieken van de hierboven gevonden schalen weergegeven. Zoals gegeven de factorladingen te verwachten viel, is Cronbach's alpha ruimschoots acceptabel (Nunnally, 1967). Ook bereik, gemid-

Tabel 2: Beschrijvende gegevens voor de drie gevonden schalen. Op de diagonaal van de correlatiematrix staat de correlatie tussen de volgens de maximum likelihood methode berekende factorscore en de schaalscore (die berekend is als het ongewogen gemiddelde van de items). Onder de diagonaal staat de correlatie tussen de schaalscores vermeld.

<i>Schaal</i>	<i>Bereik</i>	<i>Gemidd.</i>	<i>sd</i>	<i>alpha</i>
Decision Making (PR)	1.75-4.50	2.72	0.62	0.83
Horizontal work integration en klanten (HC)	1.29-4.86	2.59	0.72	0.87
Verticale communicatie en beslissingsrationalisatie (VR)	1.14-4.86	2.43	0.68	0.85

	<i>PR</i>	<i>HC</i>	<i>VR</i>
Decision Making (PR)	.97		
Horizontal work integration en klanten (HC)	.64	.97	
Verticale communicatie en beslissingsrationalisatie (VR)	.60	.59	.97

uitleggen van beslissingen niet meer van elkaar te onderscheiden. Dit is een opmerkelijk, en eigenlijk niet eens heel erg onverwacht, patroon. Het zou interessant zijn vast te stellen of deze overlapping van twee dimensies optreedt voor alle management accounting informatiesystemen.

delde en standaarddeviatie van de schalen voldoen. Het minimumniveau van de decision making schaal is enigszins aan de hoge kant¹². De correlatie tussen de schaalscores en de factorscores benadert de 1. De correlatie tussen de drie schalen is vrij hoog, hetgeen gegeven het feit dat zij allemaal gebruik meten, verwacht mag wor-

den. Onderlinge samenhang tussen de factoren is sterk, maar niet perfect. Dit is, evenals overigens het geringe aantal kruisladingen en de eigenvalues in de factoranalyse, een aanwijzing dat de discriminantvaliditeit voldoende is.

7 Verder onderzoek

Het hier gepresenteerde meetinstrument biedt langs een drietal lijnen mogelijkheden voor verder onderzoek. Allereerst is het wenselijk de hier gepresenteerde analyses te herhalen op een grotere gegevensset en daarbij geen exploratieve, maar confirmatieve factoranalyse te gebruiken. Kruisladingen kunnen dan expliciet gemodelleerd worden en bovendien is het mogelijk om de vijf oorspronkelijke dimensies te onderscheiden, terwijl ook restricties aan de covariantie tussen de factoren kunnen worden opgelegd¹³. In plaats van een tweetal dimensies - zoals de klantendimensie en horizontale communicatie - samen te voegen in één factor is het dan mogelijk te kijken naar de invloed die het veronderstellen van een perfecte correlatie tussen beide factoren heeft. Ook het vaststellen van de significantie van de discriminantvaliditeit is op deze manier mogelijk.

Een tweede lijn is het relateren van gebruik van de BSC aan organisationele performance. Dit is immers het doel waarvoor de meetinstrumentontwikkeling is opgezet en op deze manier wordt ons de mogelijkheid geboden om de bruikbaarheid van de BSC te evalueren.

In een geheel andere richting, ligt de eerder in dit artikel opgeroepen vraag of het samengaan van de verticale communicatie en beslissingsrationalisatie kenmerkend is voor systemen die ten bate van management accounting worden gebruikt. Hierbij dringt zich tevens de vraag op of bij gebruik voor verticale communicatie nog wel sprake is van vrijwillig gebruik van de BSC. Aanvullend onderzoek zou zich in het algemeen kunnen richten op de vraag wat management accounting en control systemen/prestatie metingsystemen (eventueel) uniek maakt in vergelijking met managementondersteunende systemen.

Tot slot is, zoals hierboven al is aangegeven, onderzoek denkbaar (en zeker indien de veronderstelde relatie tussen gebruik van de BSC en organisationele performance empirisch is bevestigd zelfs wenselijk) waarin het gebruiksinstru-

ment als afhankelijke variabele wordt gebruikt. Op deze manier kunnen ontwerpkenmerken van de balanced scorecard worden gerelateerd aan gebruik en, indirect, aan succes van de organisatie.

LITERATUUR

- Atkinson, Anthony A., Ramji Balakrishnan, Peter Booth, Jane M. Cote, Tom Groot, Teemu Malmi, Hanno Roberts, Enrico Uliana en Anne Wu, (1997), New directions in management accounting research, *Journal of Management Accounting Research*, Vol. 9, pag. 79-108.
- Brownell, P., (1995), Research methods in management accounting, Vol. 2 van *Accounting Research Methodology*. Coopers & Lybrand.
- DeLone, William H. en Ephraim R. McLean, (1992), Information systems success: the quest for the dependent variable, *Information Systems Research*, Vol. 3, nr. 1, pag. 60-95.
- Doll, William J. en Gholamreza Torkzadeh, (1991), On end-user computing satisfaction, *MIS Quarterly*, pag. 5-10, Mrt.
- Doll, William J. en Gholamreza Torkzadeh, (1998), Developing a multidimensional measure of system-use in an organizational context, *Information & Management*, Vol. 33, pag. 171-185.
- Frigo, Mark L. en Kip R. Krumwiede, (2000), The balanced scorecard: a winning performance measurement system, *Strategic Finance*, pag. 50-54, Jan.
- Gelderman, M., (1998), Lisrel: een waardevol hulpmiddel voor accounting onderzoek, *FMA-kroniek*, Rotterdam.
- Gelderman, M., (1995), The relation between user satisfaction, usage of information systems and performance, *Information & Management*, Vol. 34, nr. 1, pag. 11-18.
- Hair, Jr., J.F., R.E. Anderson, R.L. Tatham en W.C. Black, (1998), *Multivariate data analysis*, Prentice Hall, Upper Saddle River, New Jersey, vijfde druk.
- Helden, G. J. van, en C. P. Lewy, Klaar voor nieuwe prestatie metingssystemen? Barrières voor de balanced scorecard, *Management Control & Accounting*, Vol. 2, nr. 3, pag. 42-46, 1998.
- Hopwood, A., (1974), *Accounting and Human Behaviour*, Modern Finance Series, Haymarket Publishing.
- Ittner, C.D. en D.F. Larcker, (1998), Innovations in performance measurement: trends and research

- implications, *Journal of Management Accounting Research*, Vol. 10, pag. 205-238.
- Johnson, D.E., (1998), *Applied multivariate methods for data analysts*. Duxbury.
- Johnson, H.T. en R.S. Kaplan, (1987), *Relevance Lost: the rise and fall of management accounting*, HBS Press, Boston, Mass.
- Jöreskog, K.G. en D. Sörbom, (1989), *Lisrel7: a guide to the program and applications*. SPSS Inc., tweede druk.
- Kaplan, R.S., (1998), *Innovation action research: creating new management theory and practice*, *Journal of Management Accounting Research*, Vol. 10, pag. 90-118.
- Nunnally, J.C., (1967), *Psychometric theory*, McGraw-Hill.
- Schmitt, N. en R.J. Klimoski, (1991), *Research methods in human resources management*, *Human Resources Management*, South Western, Cincinnati, Ohio.
- Schoute, M., (1999), *Het waarom van de abc-paradox*, *Maandblad voor Accountancy en Bedrijfseconomie*, pag. 590-599, nov.
- Shaver, J.M., (1998), *Accounting for endogeneity when assessing strategy performance: does entry mode choice affect fdi survival*, *Management Science*, Vol. 44, nr. 4, pag. 571-585, Apr.

NOTEN

1 Het in dit artikel beschreven onderzoek werd financieel mogelijk gemaakt door het Limperg Instituut. Ik wil Tom Groot, Paul Jansen, Eelke Wiersma en de deelnemers aan het Vrije Universiteit Accounting Seminar bedanken voor in eerdere fasen van dit onderzoek geleverd commentaar en voor opmerkingen over dit artikel. De anonieme referent van dit tijdschrift ben ik dankbaar voor zijn uitgebreide beoordeling van mijn artikel. Ik ben André de Waal en Maarten Heere dankbaar voor samenwerking in het kader van het onderzoeksproject prestatie-meting. De in dit artikel geanalyseerde gegevens vormen een onderdeel van de in het kader van dit project aangelegde database.

2 Dat dit een punt van aandacht is mag wel blijken uit Schmitt en Klimoski (1991) waar een voorbeeld van een meta-analyse wordt besproken waarvan de conclusie is dat de prestatiebeloning de prestaties verhoogt indien als succescriterium dezelfde maatstaf wordt gekozen die voor het bepalen van de beloning wordt gebruikt.

3 Overigens zou een directe relatie tussen ontwerpkenmerken van de BSC slechts te leggen zijn indien alle gebruikers ongeacht de eigenschappen van de implementatie in gelijke mate gebruik zouden maken van het systeem. Dit lijkt geen erg reële veronderstelling.

4 Factoroplossingen met drie indicatoren per factor zijn geïdentificeerd (Jöreskog en Sörbom, 1989). Grondige validatie van meetinstrumenten met minder dan drie items per dimensie is derhalve niet mogelijk.

5 Bij het selecteren van deze organisaties is ervoor gezorgd dat de implementaties minstens twee jaar geleden waren begonnen en dat deze door verschillende organisatie-adviesbureaus waren begeleid. Een van de organisaties heeft de implementatie geheel intern (zonder hulp van consultants) uitgevoerd. Onder genoemde organisaties bevinden zich profit- en non-profitorganisatie en niet alleen producenten, maar ook dienstverleners.

6 Van de 16 overige vragenlijsten was een deel niet (volledig) ingevuld, en een ander deel evident niet serieus beantwoord.

7 Het lijkt niet onaannemelijk dat onder de non-respondenten ook een groep non-adopters zit.

8 Alle analyses die hieronder gepresenteerd worden zijn herhaald met maximum likelihood factoranalyse. Qua interpretatie zijn de resultaten niet verschillend. Daar sommige auteurs (Johnson, 1998) van mening zijn dat oblique rotatie ongewenst is - factorextractie is gebaseerd op orthogonaliteit, het is dan enigszins vreemd om deze zelfde orthogonaliteit vervolgens prijs te geven ten behoeve van interpretatie van de uitkomsten - is ook varimax-rotatie toegepast. Het aantal kruisladingen neemt toe, maar verder is ook hier het resultaat vergelijkbaar.

9 Strict genomen componenten.

10 De aard van deze items is vergelijkbaar met die van het antwoord op de vraag 'Speelt u basket-ball' in een meetinstrument dat zowel sportiviteit als lengte meet.

11 Het samengaan van deze dimensies is niet uniek voor de driefactoroplossing. De rationalization en verticale communicatie items vormen ook bij andere aantallen factoren samen een dimensie.

12 Alle items zijn negatief gecodeerd. Gebruik ten behoeve van beslissingsondersteuning is dus relatief gering.

13 Een uitgebreidere discussie van de mogelijkheden van confirmatieve factoranalyse is te vinden in Gelderman (1995).