

'Good to Great' – of alleen maar 'Good'?

Jim Collins schreef geschiedenis met zijn bestseller *Good to Great*, waarin hij aan de hand van de historische prestaties van ondernemingen vijf succesprincipes formuleerde. Zijn bevindingen lijken de wetenschappelijke toets der kritiek echter niet te kunnen doorstaan.

De auteurs zijn beiden
hoofddocent financiën
aan de universiteit van
Wisconsin Oshkosh.¹

Het boek *Good to Great* (hierna: GTG) heeft langdurig op de bestsellerlijsten gestaan van de *New York Times*, de *Wall Street Journal* en *Business Week*. Sinds de verschijning van boek zijn er drie miljoen gebonden exemplaren van verkocht. Bovendien is het boek in 35 talen vertaald.² Sinds de publicatie ervan heeft GTG wereldwijde populariteit verworven, niet alleen in bestuurskamers – tot en met die van het ziekenhuis Barnes-Jewish St. Peters in St. Louis – maar ook in het managementonderwijs in het hele land, onder meer aan de Wharton School van de Universiteit van Pennsylvania.³ Ook heeft GTG grote indruk gemaakt. De *Financial Times* vroeg lezers ter gelegenheid van de uitverkiezing van het managementboek van 2007 om uit een lijstje van vijf titels het 'beste boek aller tijden' te kiezen. Die lijst was samengesteld aan de hand van suggesties van topondernemers, onder wie Jeff Immelt van GE en Meg Whitman van eBay. GTG stond op het lijstje, samen met *Wealth of Nations* uit 1776 van Adam Smith en *The Effective Executive* uit 1966 van Peter Drucker, een duidelijk teken dat het één van de invloedrijkste managementboeken uit de geschiedenis is.

De onderzoeksaanpak van Collins

Jim Collins en zijn 21-koppige onderzoeksteam hebben in *Good to Great*⁴ aan de hand van een aantal criteria uit in totaal 1.435 ondernemingen een elftal ondernemingen geselecteerd die in hun ogen 'de sprong' hadden gemaakt naar een 'geweldig' ('great') prestatieniveau. Deze ondernemingen werden onder meer gekozen op grond van hun rendement voor de aandeelhouder, hun koppositie in de eigen bedrijfstak, hun stevige positie en het feit dat zij zelfs vijftien jaar na het omslagjaar

naar een geweldig prestatieniveau nog steeds een stijgende lijn vertoonden. De volgende ondernemingen belandden in het lijstje van Collins: Abbott Laboratories, Circuit City, Fannie Mae, Gillette, Kimberly-Clark, Kroger, Nucor, Philip Morris (nu Altria), Pitney Bowes, Walgreens en Wells Fargo. Vervolgens vergeleek Collins deze elf ondernemingen met een selectie van ondernemingen die de sprong van 'goed' naar 'geweldig' niet hadden weten te maken. Collins inventariseerde de eigenschappen die de geselecteerde ondernemingen met elkaar gemeen hadden maar die juist ontbraken bij de controlegroep. Zo kwam hij uit op vijf GTG-principes. In de tabel in figuur 1 benoemen we deze GTG-principes, waarmee een onderneming volgens Collins 'structureel geweldige resultaten' zou kunnen bereiken.⁵

De fouten in de onderzoeksaanpak van Collins

Ons inziens zijn er destijds twee fundamentele fouten geslopen in het onderzoek van Collins. Als gevolg daarvan wordt de pretentie, als zouden de vijf GTG-principes tot 'structureel geweldige resultaten' leiden, niet waargemaakt. Uit het onderzoek van Collins blijkt enkel dat de elf GTG-ondernemingen de genoemde principes met elkaar gemeen hadden gedurende de specifieke periode waar Collins naar heeft gekeken. Dat is een volstrekt andere conclusie, die vooral betekent dat GTG geen enkel bewijs biedt dat andere ondernemingen in andere tijden aan de hand van

**Onder topmanagers werd
Good to Great in één adem
genoemd met onder
meer *Wealth of Nations* van
Adam Smith**

1. De auteurs (te bereiken via niendorf@uwosh.edu respectievelijk beck@uwosh.edu) zijn Peter Cappelli, een anonieme reviewer en Barbara Rau, erkentelijk voor hun nuttige commentaar.

2. <http://www.jimcollins.com/biol/index.html>.

3. Zie 'Good to Great, a B-School Staple' (*Good to Great: verplichte kost op de business schools*), *Businessweek Online*, 19 januari 2006.

4. J. Collins, *Good to Great*, HarperCollins, New York, 2001. In het Nederlands verschenen onder dezelfde titel (*Good to Great – de Nederlandstalige editie*) bij Business Contact.

Figuur 1. Gemeenschappelijke kenmerken van de elf *Good to Great*-ondernemingen van Collins

– Leiderschap op 'niveau 5'
• Leiders die vooral kijken naar het belang van de onderneming in plaats van hun eigen positie/belang
– Eerst wie, dan Wat
• 'Zorg eerst voor de juiste mensen in de bus alvorens je bestemming te kiezen'
– Zie de naakte feiten onder ogen
• Je kunt geen goede beslissingen nemen als je de feiten niet kent
– Egelconcept
• Het is beter om één ding te weten dan om te proberen veel dingen te weten te komen
– Ontwikkel de visie voor je onderneming
• Houd vast aan de fundamentele waarden maar wees bereid om aanpak en bedrijfsstrategieën aan te passen

de benoemde principes bovengemiddeld kunnen presteren. Met andere woorden, Collins heeft geen 'tijdloze, universele antwoorden gevonden die door iedere onderneming kunnen worden toegepast'.⁶ De twee fatale fouten in GTG betreffen de 'datamining' en de toekenning van oorzakelijk verband aan correlaties. We gaan in dit artikel in op deze twee fouten en bekijken hoe de elf GTG-ondernemingen sinds het onderzoek van Collins hebben gepresteerd. Op grond daarvan vinden wij bewijs met betrekking tot de algemene toepasbaarheid van de GTG-principes. De uitkomsten van onze toetsingen bieden empirisch bewijs ten aanzien van de stelling van Collins dat toepassing van de GTG-concepten leidt tot 'structureel geweldige resultaten'. In het navolgende bespreken we eerst de datamining door Collins, en vervolgens het oorzakelijk verband dat in zijn onderzoek is toegekend aan de geconstateerde correlaties. Daarna beschrijven wij de opzet van onze empirische toets en de uitkomsten ervan.

Datamining

De datamining ('gegevensdelving') is een fundamentele tekortkoming van het GTG-onderzoek. Datamining houdt in dat er gegevens worden verzameld; vervolgens worden die gegevens onderzocht op patronen; zijn die patronen eenmaal in kaart, dan worden verklaringen geformuleerd die dienen als achterliggende oorzaken of principes. De gevonden patronen kunnen echter volledig worden bepaald door de onderzochte periode en

de onderzochte gegevens. Daarom biedt datamining geen legitiem bewijs voor een toepasbaarheid buiten de steekproef of periode. Dit probleem met datamining (ook wel: data 'dredging', 'grubbing' of 'fishing' genoemd – ofwel: 'baggeren', 'wroeten' of 'vissen' naar gegevens) wordt goed omschreven door Hand:⁷

Niet-uniforme gegevens verschillen per definitie van elkaar, en deze verschillen kunnen worden geïnterpreteerd als patronen. Het probleem is dat deze 'patronen' vaak gewoon het gevolg zijn van willekeurige fluctuaties en geen afspiegeling van een onderliggende structuur. Gegevensanalyse is niet bedoeld om het tijdelijke willekeurige patroon van het moment in een model onder te brengen, maar juist de onderliggende structuur die tot consistente en herhaalbare patronen leidt. Voor statistici spreekt uit het begrip datamining dan ook een naïeve hoop, een vergeefse strijd tegen de kille realiteit van het toeval.

Onderzoekers die met datamining werken, zoals Hand, kunnen dus conclusies trekken die ofwel volledig gebaseerd zijn op louter willekeurige patronen dan wel op patronen die zich enkel voordoen bij de ondernemingen in de steekproef of gedurende de gekozen onderzoeksperiode. Collins bedient zich van datamining om aan zijn vijf GTG-principes te komen. Volgens zijn eigen beschrijving heeft hij de ondernemingen in de *Fortune 500*-ranglijst op grond van vier selectiecriteria teruggebracht tot de elf GTG-ondernemingen. Vervolgens vergeleek hij die selectie met elf andere ondernemingen om 'in kaart te brengen welke essentiële en onderscheidende factoren hier in het spel waren'.⁸ Dit is datamining. In de woorden van Collins zelf: 'Het is belangrijk om voor ogen te houden dat we alle concepten, die we in dit boek presenteren, hebben ontwikkeld door middel van empirische deducties, direct uit de gegevens. We zijn dit onderzoek niet begonnen met een bepaalde theorie die we wilden toetsen of bewijzen. Ons doel was om van de grond af aan een theorie te ontwikkelen, direct afgeleid van het feitenmateriaal'.⁹ Gelet op de voor GTG gehanteerde methode van datamining heeft Collins niet bewezen dat zijn vijf principes meer zijn dan vijf 'tijdelijke willekeurige patronen'. Met andere woorden, als Collins gelijk heeft, dan moet op grond van de door hem gehanteerde statistiek worden afgeleid dat zijn gelijk louter toeval is. Zoals Walker al opmerkte, krijg je met de door Collins toegepaste methode 'een cirkelredenering: dit is wat geweldige ondernemin-

5. Collins, *Good to Great*, p. 190.

6. *Idem*, p. 5.

7. D.J. Hand, 'Datamining: Statistics and more?', *The American Statistician*, 52, 1998, p. 112-119; citaat op p. 112.

8. Collins, *Good to Great*, p. 3.

9. *Idem*, p. 10.

gen doen, dus doe die dingen en dan wordt je onderneming geweldig'.¹⁰

Mits goed gebruikt kan datamining een buitengewoon effectief instrument zijn. Dat goede gebruik houdt in dat de onderzoeker a) gegevens onderzoekt op patronen, b) verklaringen formuleert voor die patronen, en c) die verklaringen vervolgens statistisch toetst om te onderzoeken in hoeverre ze ook gelden buiten de steekproef en tijdsperiode van het onderzoek. In medisch onderzoek wordt deze methode bijvoorbeeld veelvuldig toegepast. Het cruciale en noodzakelijke ingrediënt voor het succes van de methode is een uitgebreide en nauwgezette toetsing van patronen die op grond van de gegevens zijn geconstateerd. Collins heeft een dergelijke toetsing niet gegeven in GTG. In plaats daarvan heeft hij zijn (in potentie) 'tijdelijke willekeurige patronen' gepresenteerd als vijf principes waarmee een onderneming in het algemeen geweldig kan gaan presteren.

Correlatie versus oorzakelijk verband

Een tweede probleem met GTG is dat correlaties zijn aangezien voor oorzakelijke verbanden. Collins koos in zijn GTG-onderzoek ondernemingen die hij als geweldig presterend bestemde. Vervolgens onderzocht hij de kenmerken van die ondernemingen in vergelijking met een groep andere ondernemingen. Zo vond hij vijf kenmerken die de 'geweldige' ondernemingen met elkaar gemeen hadden. Met andere woorden, hij constateerde een *verband* tussen de geweldige ondernemingen en de vijf kenmerken. Hij constateerde dat de elf GTG-ondernemingen geweldig presteerden in een periode waarin zij ook de gevonden principes vertoonden. Hij leverde echter geen enkel bewijs voor het door hem opgevoerde *oorzakelijk* verband: gevestigde of startende onderneming + GTG-concepten = Structureel Geweldig Resultaat.¹¹

Op grond van de voor GTG gehanteerde methode is niet aangetoond dat de geselecteerde ondernemingen als gevolg van de benoemde kenmerken tot geweldige prestatieniveaus waren gekomen. Aangetoond is alleen dat de elf ondernemingen deze kenmerken met elkaar gemeen hadden. Deze fout kan in de praktijk veel schadelijker zijn dan datamining. Immers, managers zouden hieruit de conclusie kunnen trekken dat zij geweldige resultaten kunnen boeken als ze de vijf principes toepassen in hun bedrijfsvoering. In werkelijkheid zijn er minstens twee alternatieve verklaringen.

Een eerste alternatieve verklaring is deze: een on-

derneming kan zich aan de principes houden doordat ze geweldig presteert; bijvoorbeeld – een onderneming die winst maakt, kan meer investeren in werving & selectie om de juiste mensen aan te trekken. Een tweede alternatieve verklaring: er kan een externe factor in het spel zijn, bijvoorbeeld een onopgemerkte factor in het ondernemingsklimaat, die tegelijkertijd zowel structureel goede resultaten als de principes in de hand werkt. Opnieuw biedt GTG geen bewijs voor een oorzakelijk verband; er wordt alleen gesteld dat er sprake is van een oorzakelijk verband.

In de epiloog van GTG gaat Collins in op veel gestelde vragen. De derde vraag betrof de statistische significantie gelet op het beperkte aantal van 28 ondernemingen voor het hele onderzoek. Collins antwoordde hierop dat hij 'twee vooraanstaande hoogleraren had geraadpleegd ten aanzien van deze vraag' en dat zij hadden 'geconcludeerd dat we geen statistisch probleem hebben'.¹² In een aanvullende toelichting citeerde hij een van de hoogleraren, William P. Briggs van de Universiteit van Colorado:

*Hoe groot is de kans dat je toevallig stuit op een groep van elf ondernemingen die elk de door u geconstateerde primaire eigenschappen hebben, terwijl die ontbreken bij de ondernemingen waarmee ze zijn vergeleken?*¹³

Vervolgens schrijft Collins:

*Hij (Briggs) concludeerde dat die kans kleiner is dan 1 op 17 miljoen. Er is vrijwel uitgesloten dat we simpelweg zijn gestuit op elf willekeurige gebeurtenissen (de elf ondernemingen; vert.) die toevallig het 'good to great'-patroon vertonen waar we naar zochten. We kunnen met zekerheid concluderen dat de door ons gevonden eigenschappen nauw verband houden met de omslag van een goed naar een geweldig prestatievermogen.*¹⁴

Deze redenering heeft geen betrekking op de onderzoeksmethode die voor GTG is gehanteerd. Collins heeft niet eerst de 'primaire eigenschappen' benoemd om vervolgens de kenmerken te beoordelen van iedere onderneming waar hij die primaire eigenschappen aantrof, zoals zijn toelichting lijkt te suggereren. Hij begon met de ondernemingen en onderzocht ze vervolgens op gemeenschappelijke kenmerken. De redenering van de hoogleraren is dus niet van toepassing op de door Collins gehanteerde methode. Gelet op zijn onderzoeksmethode

10. G. Walker, 'Good to Great: Why some companies make the leap—and others don't', [recensie van *Good to Great* in] *Academy of Management Perspectives*, 20(1), 2006, p. 120-122; citaat op p. 120.

11. Collins, *Good to Great*, p. 190.

12. Idem, p. 211-212.

13. Idem, p. 212.

14. In deze formulering klinkt door dat de hoogleraren eigenlijk een voorwaardelijke kans wilden kwantificeren. De vraag had ook als volgt kunnen worden gesteld: 'Hoe groot is de kans dat elf geselecteerde ondernemingen vijf kenmerken met elkaar gemeen hebben, gelet op het feit dat ze allemaal een geweldig beleggingsrendement bieden?'

waren de 'elf willekeurige gebeurtenissen' geen willekeurige gebeurtenissen – maar gebeurtenissen/ondernemingen die eerst doelbewust en zorgvuldig zijn uitgekozen en vervolgens zijn doorwroet op wat ze met elkaar gemeen hadden. De professoren hadden de volgende vraag moeten stellen:

*Hoe groot is de kans dat je in een steekproef van ondernemingen, gebaseerd op overeenkomstige prestatiekenmerken, vijf gemeenschappelijke kenmerken vindt?*¹⁴

Collins heeft niet aangetoond dat de geselecteerde ondernemingen dankzij de benoemde kenmerken tot geweldige prestatieniveaus waren gekomen

Volgens ons ligt deze kans veel dichterbij honderd procent dan bij één op zeventien miljoen. Hoewel Collins elf ondernemingen heeft geselecteerd voor zijn vergelijking, maakt zijn methode niet duidelijk of de werkwijzen bij de elf GTG-ondernemingen verschillen van de werkwijzen bij andere ondernemingen in de *Fortune 500*. Misschien hebben sommige ondernemingen die slecht presteren ook wel een leiding op 'niveau 5', maar daar kom je met dit onderzoek niet achter.

De juiste aanpak

Het valt te waarderen in Collins dat hij ervoor koos om niet het geijkte pad te volgen en dus 'dit onderzoek niet (is) begonnen met een bepaalde theorie die we wilden toetsen of bewijzen' maar in plaats daarvan 'van de grond af aan een theorie (wilde) ... ontwikkelen'.¹⁵ Om de door Collins geponeerde stellingen te kunnen rechtvaardigen, is het echter nodig om een theorie te formuleren, hypothesen op te stellen, een steekproef te doen, toetsingen uit te voeren en conclusies te trekken over de validiteit van de theorie. Voor alle duidelijkheid: wij beweren hier niet dat de conclusies en stellingen van Collins onjuist zijn; wij beweren dat hij geen bewijs biedt dat zijn bevindingen meer zijn dan slechts willekeurige patronen.

Om zijn stellingen te onderbouwen, zal Collins (of wie dan ook) een vervolgonderzoek dienen te doen op GTG om de validiteit van wat hij poneert te toetsen met behulp van de klassieke statistische analyse. De eerste stap zou daarbij zijn om hypothesen op te stellen met betrekking tot de relatie tussen de vijf GTG-principes en het succes van een onderneming (afgemeten aan, bijvoorbeeld, het rendement voor de aandeelhouder). Die hypothe-

sen zouden niet alleen gebaseerd dienen te zijn op het inductieve werk van Collins maar ook op eerder managementonderzoek naar dergelijke relaties. Er is bijvoorbeeld veel literatuur over het belang van valide selectieprocedures om 'de juiste mensen in de bus te krijgen'. Nadat adequate hypothesen zijn opgesteld, zou vervolgens een selectie van beursgenoteerde ondernemingen moeten worden gemaakt die representatief is voor alle ondernemingen. Ten slotte zou met behulp van een van de statistische methoden, bijvoorbeeld de regressieanalyse, moeten worden onderzocht in hoeverre de relaties, die in de hypothesen worden verondersteld, zich voordoen onder de geselecteerde ondernemingen.

Onze aanpak

Methode

Behalve onze uiteenzetting over de statistische fouten in GTG testen we ook hoe de GTG-ondernemingen hebben gepresteerd. We doen dit om niet alleen op grond van statistische argumenten maar ook empirisch aan te tonen dat er geen basis is voor de beweringen van Collins. We hebben de veralgemeenbaarheid van de GTG-uitkomsten op twee manieren onderzocht. Eerst hebben we bekeken hoe de GTG-ondernemingen in de tien jaar na het GTG-onderzoek hebben gepresteerd in vergelijking met de ondernemingen in de *Fortune 500*. Als de elf GTG-ondernemingen werkelijk geweldig waren, dan zouden zij ook buiten de onderzoeksperiode aanmerkelijk beter hebben moeten presteren dan andere ondernemingen. We hebben de resultaten afgemeten aan het rendement voor de aandeelhouder over die periode van tien jaar.¹⁶ Om precies te zijn: we hebben de elf GTG-ondernemingen en de ondernemingen in de *Fortune 500*-lijst van 2006 gerangschikt op grond van het rendement voor de aandeelhouder om te zien hoe die elf ondernemingen het in de periode 29 december 1995 - 29 december 2005 (hierna: 1995 – 2005) hebben gedaan. Eén van de selectiecriteria voor de GTG-groep van ondernemingen was dat ze aan het eind van de onderzoeksperiode voor GTG nog steeds 'een stijgende lijn' moesten vertonen. Voor zover dit een meetbaar en effectief criterium is, zou dit tot een opwaartse tendens moeten leiden in de rangschikking van de elf GTG-ondernemingen, waardoor ze er in onze rangschikking nog beter uit zouden moeten springen.

In een tweede test hebben we het rendement op de elf GTG-aandelen vergeleken met het rendement op de S&P 500 voor de periode na het GTG-

15. Collins, *Good to Great*, p. 10.

16. Collins en zijn team hebben de elf ondernemingen destijds eveneens mede geselecteerd op grond van het rendement voor de aandeelhouder over een langere periode.

onderzoek.¹⁷ Indien de conclusies van Collins algemeen toepasbaar waren, dat wil zeggen, ook buiten de onderzoeksperiode, dan zouden de geselecteerde ondernemingen ook in andere perioden betere resultaten hebben geboekt dan de overige ondernemingen. Onze hypothese was dat er geen significant verschil zou worden gevonden tussen het rendement voor elk van de elf GTG-ondernemingen en het rendement voor de S&P 500 in de periode volgend op de selectieperiode van Collins. Deze hypothese kan als volgt worden geformuleerd:

H_0 : *Er is geen verschil tussen het gemiddelde rendement voor iedere GTG-onderneming en het gemiddelde rendement voor de S&P 500-index.*

H_1 : *Het gemiddelde rendement voor iedere GTG-onderneming is groter dan het gemiddelde rendement voor de S&P 500-index.*

Gegevens

We hebben deze hypothese getoetst voor twee perioden: de periode direct volgend op de selectieperiode voor GTG (1995-2005) en een zo lang mogelijke periode die niet overlapt met de selectieperiode voor GTG. Eén van de selectiecriteria voor GTG was dat een onderneming over een periode van vijftien jaar een geweldig ('great') cumulatief rendement voor de aandeelhouder moest hebben gegenereerd: meer dan drie keer het marktgemiddelde. Bovendien moest die periode zijn voorafgegaan door een periode van vijftien jaar met een goed ('good') cumulatief rendement (minder dan een derde van het 'geweldige' cumulatieve rendement). Het punt tussen deze twee perioden werd gezien als het omslagpunt ('transition point') – het punt waarop de onderneming de omslag maakte van een 'goed' naar een 'geweldig' ('good to great') prestatieniveau. De GTG-ondernemingen beleefden deze omslag op uiteenlopende momenten tussen 1964 en 1984. Om een zo lang mogelijke testperiode te kunnen krijgen, hebben wij het aandelenrendement voor iedere GTG-onderneming afzonderlijk vergeleken met het rendement op de S&P 500 voor een periode van vijftien jaar vanaf het omslagpunt voor de betreffende onderneming.

Uitkomsten

We hebben de elf GTG-ondernemingen en de ondernemingen in de *Fortune 500*-ranglijst van 2006 gerangschikt op grond van het rendement voor de aandeelhouder in de periode 1995-2005. Als de elf GTG-ondernemingen waren geleid door bij uitstek kundige managementteams, dan zouden ze,

op grond van de theorie van Collins, ook in perioden buiten de onderzoeksperiode voor GTG een 'geweldig' ('great') resultaat te zien moeten geven in vergelijking met de andere ondernemingen. De tabel in figuur 2 laat de tien best presterende ondernemingen uit de *Fortune 500* zien, gerangschikt naar het rendement voor de aandeelhouder. De ondernemingen in deze lijst, die aangevoerd wordt door NVR en Dell, behaalden in de onderzochte periode van tien jaar gemiddeld een rendement voor de aandeelhouder variërend van 53 procent (NVR) tot 31,6 procent (Kinder Morgan Energy). Het gemiddelde rendement voor deze top-10 bedroeg 37,8 procent. De elf GTG-ondernemingen, daarentegen, behaalden in dezelfde periode gemiddeld een rendement van 10,9 procent – nog geen derde van het gemiddelde voor de top-10. Hoewel een aantal GTG-ondernemingen een goed rendement behaalden (Walgreens bijvoorbeeld 20,2 procent) liep de bandbreed-

te in deze groep af tot een rendement van 6,2 procent voor Kimberly-Clark. Van de elf GTG-ondernemingen kwam Walgreens het hoogste uit op de ranglijst voor de *Fortune 500* (plaats 62); Kimberly-Clark eindigde met een 302^e plaats als laatste. Gemiddeld kwam de GTG-selectie als groep uit op de 202^e plaats van de ranglijst. Met andere woorden, noch in de plaatsing op de ranglijst, noch in het rendement vinden we enige aanwijzing voor een 'geweldige' prestatie in de onderzochte tien jaar volgend op de steekproefperiode voor GTG.

De uitkomsten in de tabel in figuur 2 zijn vooral veelzeggend als we rekening houden met de opwaartse tendens in de rangschikking van de elf GTG-ondernemingen die Collins mede als voorwaarde had gesteld voor opname in de groep van 'geweldig' presterende ondernemingen: ze moesten aan het eind van de meetperiode nog steeds 'in de lift' zitten. De GTG-ondernemingen zijn duidelijk sterke ondernemingen die geleid worden door een competent management. Een gemiddelde 202^e plaats op de ranglijst (mede dankzij de genoemde opwaartse tendens), in combinatie met een gemiddeld rendement dat slechts een derde bedraagt van dat van de top-10, rechtvaardigt echter niet het etiket 'geweldig' ('great') van GTG.

De plaatsing op de ranglijst noch het rendement bieden enige aanwijzing voor een 'geweldige' prestatie in de tien jaar volgend op de steekproefperiode voor Good to Great

17. We hebben de S&P 500-index gebruikt omdat er voor de Fortune 500 geen consequent geciteerde rendementsindex is, wat voor de hand ligt aangezien die lijst veelvuldig van samenstelling verandert. Veel ondernemingen die in enig jaar te vinden zijn in de Fortune 500 zitten ook in de S&P 500. Bovendien wees een anonieme reviewer erop dat Collins geprobeerd heeft de ondernemingen in zijn selectie te vergelijken met soortgelijke ondernemingen ('matched pair analysis'); helaas bleek bij nadere beschouwing dat de 'paren' zich niet leenden voor een goede vergelijking.

Figuur 2. De Fortune 500 van 2006, gerangschikt naar het rendement voor de aandeelhouder voor de periode ultimo 1995 - 2005¹

Rangschikking naar rendement	Plaats op de ranglijst van de Fortune 500 van 2006	De tien beste Fortune 500 ondernemingen	Rendement voor de aandeelhouder, ultimo 1995 - 2005 %
1	132	NVR	53,0
2	225	Dell	39,4
3	10	Jabil Circuit	38,7
4	448	Best Buy	37,7
5	335	Frontier Oil	37,6
6	131	Ryland Group	36,8
7	123	Whole Foods Market	36,6
8	96	MDC Holdings	33,8
9	181	Qualcomm	32,5
10	330	Kinder Morgan Energy	31,6
		Gemiddeld rendement over tien jaar	37,8%
		<i>Good to Great-onderneming</i>	
62	45	Walgreens	20,2
98	46	Wells Fargo	17,1
122	20	Philip Morris (nu Altria)	15,2
223	177	Nucor	10,3
235	226	Circuit City	9,8
240	93	Abbott Laboratories	9,4
247	394	Pitney Bowes	9,1
280 ²	Niet in de ranglijst	Fannie Mae	7,6
283 ³	Niet in de ranglijst	Gillette	7,5
289	21	Kroger	7,3
302	140	Kimberly-Clark	6,2
202		Gemiddelde	10,9%

¹. Meetperiode voor het rendement is van 2 januari 1996 tot en met 30 december 2005.

Dezelfde periode is gebruikt door Forbes voor de ondernemingen in de Fortune 500-ranglijst van 2006

². Fannie Mae stond niet in de Fortune 500 voor 2006, maar het rendement en de positie op de ranglijst zijn hier weergegeven alsof dat wel zo was.

³. Rendement voor Gillette alleen tot en met 2004, toen de onderneming werd overgenomen door Procter & Gamble.

18. We hebben ook de periode van twintig jaar van 1986 tot en met 2005 onderzocht. Ondanks een sterke tendens naar ontkrachting van de nul-hypothese als gevolg van de overlap van tien jaar met de onderzochte periode voor GTG konden we de nul-hypothese ook voor deze periode in zes van de elf gevallen niet ontkrachten.

De uitkomsten in de tabel in figuur 3 werpen ook meer licht op de vraag omtrent de algemene toepasbaarheid van de bevindingen van Collins voor andere perioden. Deel 1 van figuur 3 geeft de uitkomsten van gepaarde t-toetsen voor de nul-hypothese (geen verschil tussen het gemiddelde rendement van de GTG-ondernemingen en het gemiddelde rendement van de S&P 500-index voor de periode 1995 tot en met 2005). Zoals de uitkomsten laten zien, konden wij de nul-hypothese in tien van de elf gevallen niet ontkrachten. We moeten derhalve wel concluderen dat er in die tien gevallen geen significant verschil bestaat tussen het rendement van de GTG-ondernemingen en de gemiddelde onderneming in de S&P 500.

Slechts één GTG-onderneming, Walgreens, deed het in de onderzochte periode beter dan de S&P 500. Kortom, we hebben nauwelijks aanwijzingen gevonden voor het 'geweldige' prestatievermogen dat in GTG wordt beschreven.¹⁸

We hebben ook het rendement van elk van de elf ondernemingen afzonderlijk vergeleken met dat van de S&P 500 voor de periode vanaf vijftien jaar na het omslagpunt tot eind 2005. Dat levert een langere meetperiode op dan de tien jaar voor deel 1 van figuur 3; gemiddeld veertien jaar per onderneming. Ook de uitkomsten in deel 2 van tabel 3 bieden geen ondersteuning voor een algemene geldigheid van de bevindingen van Collins. In negen van de elf gevallen konden we de nul-hypothese

Figuur 3. Toetsing van de *Good to Great*-rendementen in vergelijking met de S&P 500

Deze tabel toont een vergelijking van het rendement voor de elf GTG-ondernemingen met dat van de S&P 500. De nul-hypothese van vergelijkbare rendementen wordt ontkracht indien het rendement van de afzonderlijke aandelen voor de onderzochte periode significant hoger uitvalt dan het rendement van de S&P 500. Als de nul-hypothese wordt ontkracht, mogen we concluderen dat het gemiddelde aandelenrendement hoger is dan het gemiddelde rendement van de S&P 500. Wordt de nul-hypothese niet ontkracht, dan is er geen significant verschil in rendement. Significante p-waarden zijn vetgedrukt.

Deel 1: Ultimo 1995 t/m 2005				
Onderneming	Uitkomst van gepaarde toets		p-waarde	
Abbott Laboratories	0,427		0,335	
Altria (voorheen Philip Morris)	1,261		0,105	
Circuit City Stores	1,098		0,137	
Federal National Mortgage Association	0,672		0,251	
Gillette*	0,436		0,332	
Kimberly-Clark	0,532		0,298	
Kroger	0,457		0,324	
Nucor	0,567		0,286	
Pitney Bowes	0,651		0,270	
Walgreens	1,703		0,045	
Wells Fargo & Co.	1,554		0,061	
Deel 2: Omslagjaar + 15 jaar - ultimo 2005				
Onderneming	Omslagjaar	Aantal getoetste jaren	Uitkomst van gepaarde toets	p-waarde
Abbott Laboratories	1974	16	1,055	0,146
Altria (voorheen Philip Morris)	1964	26	2,605	0,005
Circuit City Stores	1982	8	1,220	0,113
Federal National Mortgage Association	1984	6	0,273	0,393
Gillette*	1980	9	0,324	0,373
Kimberly-Clark	1972	18	1,269	0,103
Kroger	1973	17	1,066	0,144
Nucor	1975	15	1,449	0,074
Pitney Bowes	1973	17	0,729	0,233
Walgreens	1975	15	1,930	0,028
Wells Fargo & Co.	1983	7	1,109	0,135

* Rendement voor Gillette alleen tot en met 2004, toen de onderneming werd overgenomen door Procter & Gamble.

niet ontcrachten. Op grond daarvan concluderen we derhalve met nog meer zekerheid dat het prestatieniveau van de GTG-ondernemingen nauwelijks verschilde van dat van de gemiddelde onderneming in de S&P 500. Met andere woorden, we hebben geen bewijs gevonden dat de GTG-ondernemingen het 'geweldig' deden in de periode na de selectieperiode van Collins.

De zwakke punten van ons onderzoek

Zoals elk statistisch onderzoek zou ook ons onderzoek om een aantal redenen een onvoldoende kunnen krijgen. Het is bijvoorbeeld in theorie mogelijk dat de elf GTG-ondernemingen in de

door ons onderzochte periode de GTG-principes hebben losgelaten. Als dit op de een of andere manier waar was, dan zou het door ons gemeten gemiddelde prestatievermogen van deze ondernemingen in overeenstemming zijn met wat Collins beweert. Als we echter bedenken hoe sterk de krachten tegen verandering zijn in een onderneming, dan lijkt de kans zeer klein dat zoiets zou zijn gebeurd bij tien van de elf GTG-ondernemingen (met Walgreens als uitzondering). Echter, ook al zouden onze uitkomsten om welke reden dan ook onaanvaardbaar zijn, dan nog kan daaruit niet worden geconcludeerd dat de door Collins gehanteerde methode dat wel is. Wij stelden ons ten doel

om zijn bevindingen te toetsen. Ook al zouden wij dat niet naar behoren hebben gedaan, dan nog voldoet zijn methode niet.

Conclusie

We hebben zowel statistische argumenten als empirisch bewijs voorgelegd met betrekking tot de stelling van Collins dat de GTG-concepten tot 'structureel geweldige resultaten' leiden. Het empirische bewijs biedt geen ondersteuning voor zijn stellingname. Wanneer de elf GTG-ondernemingen samen met de *Fortune 500* van 2006 worden gerangschikt op grond van het rendement voor de aandeelhouder gemeten over de periode 1995 tot en met 2005, dan komen de elf GTG-ondernemingen uit op een gemiddelde 202^e plaats. Bovendien bedroeg hun rendement voor de aandeelhouder in die periode nog niet een derde van dat van de top-10. Ook uit gepaarde t-toetsen voor de rendementsprestatie van de elf GTG-ondernemingen ten opzichte van

de S&P 500-index blijkt geen noemenswaardig verschil tussen de door Collins geselecteerde ondernemingen en de gemiddelde onderneming in de S&P 500.

Onze verklaring hiervoor is dat de selectiemethode voor de 'geweldig' presterende ondernemingen in GTG een klassiek voorbeeld is van datamining. Deze fout is verergerd doordat er een oorzakelijk verband is toegekend aan gesignaleerde correlaties. Bij de in GTG 'ontdekte' managementprincipes lijkt het dus niet te gaan om 'tijdloze en algemeen geldende antwoorden die door iedere organisatie kunnen worden toegepast'.¹⁹ Er kon geen bewijs worden geleverd dat ze ook voor andere ondernemingen en andere tijden zouden gelden. Op grond van de onderzoeksmethode en het bewijs in GTG kan niet worden gesteld dat de vijf principes tot 'geweldige' prestaties zouden leiden. De elf GTG-ondernemingen zijn zeker goed, maar we hebben geen bewijs gevonden dat ze geweldig zijn.

Reprinted from Bruce Niendorf and Kristine Beck, 'Good to Great, or Just Good?', *Academy of Management Perspectives*, November 2008, pp. 13-20, with permission from the publisher. Copyright © 2008 by the Academy of Management. All rights reserved. Vertaling: Raymond Gijsen.

REACTIE

HIERONDER TWEE REACTIES OP HET ARTIKEL 'GOOD TO GREAT - OF ALLEEN MAAR 'GOOD'? VAN DR. ANDRÉ DE WAAL MBA EN PROF. DR. J. STRIKWERDA CMC. DE WAAL IS ASSOCIATE PROFESSOR AAN DE MAASTRICHT SCHOOL OF MANAGEMENT EN ACADEMIC DIRECTOR VAN HET CENTER FOR ORGANIZATIONAL PERFORMANCE. STRIKWERDA IS ALS MANAGEMENT CONSULTANT VERBONDEN AAN NOLAN, NORTON & CO EN HOOGLEERAAR ORGANISATIE EN VERANDERING AAN DE FACULTEIT VOOR ECONOMIE EN BEDRIJFSKUNDE VAN DE UNIVERSITEIT VAN AMSTERDAM.

Van 'Good to Great' door de gehaktmolen

André de Waal

Het heeft lang geduurd, maar nu is toch het onvermijdelijke gebeurd: Jim Collins is van zijn voetstuk gevallen. Of beter: gestoten. Niendorf/Beck maken, het kan niet anders gezegd worden, in hun artikel gehakt van Collins' onderzoeksmethode en daarmee van zijn onderzoeksresultaten. Ik zeg 'onvermijdelijk' omdat we sinds de publicatie in 1982 van *In Search of Excellence* van Peters en Waterman weten dat het niet verstandig is om bedrijven te noemen in onderzoek naar 'high performance organisaties' (HPO's). Het maakt onderzoeken kwetsbaar omdat de onderzoeker niet kan voorzien hoe het de beschreven organisaties zal vergaan.

In het geval van Peters en Waterman gingen bijna alle genoemde bedrijven in de jaren na de publicatie van het boek kopje onder of bijna failliet, met IBM als het bekendste voorbeeld. Sindsdien wordt door vele critici smalend opgemerkt dat het onderzoek van beide heren geen waarde kan hebben omdat zoveel van de beschreven bedrijven kopje onder

zijn gegaan, en dat daarom geen aandacht hoeft te worden gegeven aan de onderzoeksresultaten. Dat dit ook met de door Collins geselecteerde organisaties zou gaan gebeuren, was een kwestie van tijd.

Maar de kritiek op Collins' onderzoek, zoals geleverd in het artikel van Niendorf/Beck is fundamenteeler. Het grootste bezwaar tegen de onderzoeksmethode die Collins heeft gebruikt, is dat hij datamining heeft toegepast. Dit komt er op neer dat Collins die data heeft gebruikt die hem het beste uitkwamen. Uit het oogpunt van de onderzoeksbetrouwbaarheid is dit een doodzonde, het veroorzaakt namelijk dat het onderzoek niet replicerbaar is – immers, andere data geven andere resultaten – en dat de uitkomsten toevalligheden zijn. Daarbij komt nog dat selecteren aan het begin van het onderzoek veel risico's met zich meebrengt. Zo maakte Collins op basis van een financiële analyse een selectie tussen excellente en minder goed

19. Collins, *Good to Great*, p. 5.

presterende organisaties en gebruikte hij dit als uitgangspunt voor de vergelijking. Hieruit werden dan conclusies getrokken met betrekking tot de onderscheidende kenmerken. De vraag is echter of de conclusies valide zijn: zijn data verzameld op basis van de juiste criteria en zijn deze data geanalyseerd met behulp van een wetenschappelijk valide vergelijkingstechniek? En zijn de conclusies van toepassing op 'alle' organisaties, tijdloos en universeel?

Daarbij komt altijd een element van toeval kijken: was de juiste informatie überhaupt beschikbaar? Uit de kritiek in het artikel blijkt dat in dit opzicht op de aanpak van Collins veel aan te merken is. Zo kun je je afvragen of de door Collins gekozen indicator voor 'greatness', de beurskoers, wel de juiste is. De markt heeft immers niet altijd gelijk, kijk maar naar de kredietcrisis of de internetzeepbel. Betekent dit dan ook dat *Good to Great* niet langer als relevant beschouwd kan worden? Dat ligt eraan hoe de lezer het boek wil gebruiken. Het mag niet worden gezien als een universeel receptenboek waarmee de lezer, als hij de vijf stappen maar nauwgezet volgt, zijn organisatie per definitie 'great' zal maken. Wordt het boek daarentegen als inspiratiebron gebruikt, als een 'gedachtenboek'

waarvan de inhoud door de lezer zelf voor zijn bedrijf toepasbaar wordt gemaakt, dan zal het nog immer tot zijn recht kunnen komen en waardevol kunnen zijn.

De auteurs van het artikel houden het overigens gelukkig niet alleen bij kritiek, ze komen ook met advies over hoe toekomstig HPO-onderzoek robuuster kan worden uitgevoerd. De eerste stap in dat onderzoek is het opstellen van hypothesen over de mogelijke relatie tussen in de literatuur geïdentificeerde succesfactoren en organisatieresultaten. Daarna moeten deze hypothesen, met behulp van financiële en niet-financiële data afkomstig van willekeurig gekozen organisaties, worden geverifieerd met één van de vele beschikbare statistische analysetechnieken (zoals factoranalyse of regressieanalyse). Tot slot moeten de beperkingen van het onderzoek goed worden aangegeven: wat kan de lezer daar wel en niet mee doen. Door dit advies is nu een meetlat neergelegd waarmee toekomstige studies naar excellente organisaties de maat kan worden genomen. Daarbij sluit ik me aan bij de oproep van de auteurs aan Collins om deze methode alsnog los te laten op zijn 'Good to Great'-onderzoek.

Jede Konsequenz führt zum Teufel ...¹

Hans Strikwerda

En dus ook elk succesmodel voor managers. Het principiële verschil tussen de natuurwetenschappen en de gedragswetenschappen is dat de natuur zelf geen weet heeft van haar wetmatigheden. Als Newtons appel valt door de werking van de zwaartekracht, dan is dat voor de andere appels geen reden om samen te zweren tegen de zwaartekracht. In weerwil van de vele publicaties die de natuurwetenschappen als rolmodel voor de gedragswetenschappen relativeren c.q. verwerpen, is er een onophoudelijke stroom van publicaties op het gebied van management die dat rolmodel blijft hanteren. *Good to Great* van Collins is er daar één van. Het begon in 1982 met *In Search of Excellence* van Peters en Waterman. De tijd heeft geleerd dat hun 'wetmatigheden' niet klopten. Twee academische publicaties in het tijdschrift *Academy of Management Perspectives* relativeren nu de conclusies van Collins.² De in het onderzoek van Collins als *great* geïdentificeerde ondernemingen zouden dat toch niet zijn. Als al die natuurkundeboeken

over succesvol management zijn gebaseerd op uitgebreide, gedegen statistische analyses, waarom lukt het dan niet *sustainable* succesformules te formuleren?

Allereerst, wat beweert Collins in essentie in zijn boek *Good to Great*?

1. De *Good-to-Great-leider* onderscheidt zich door bescheidenheid, niet hoog van de toren blazen, wilskracht en doorzettingsvermogen. Dit type leider is niet bezig met zijn eigen ego, maar met het belang van de onderneming en zijn mensen, op langere termijn;
2. Het werven van goede mensen (met een mentaliteit van aanpakken en niet zeuren) en het wegsturen van mensen die niet goed zijn, is belangrijker dan een briljante visie en dito strategie en organisatie;
3. De harde feiten moeten onder ogen worden gezien, zonder daarbij het geloof in het uiteindelijke succes te verliezen;

1. Johan Wolfgang von Goethe (1749-1832).

2. Niendorf, B. & K. Beck (2008). *Good to Great, or Just Good?* *Academy of Management Perspectives*, 22(4), pp. 13-20.
Resnick, B. G., & T.L. Smunt, T. L. (2008). *From Good to Great to ...* *Academy of Management Perspectives*, 22(4), pp. 6-12.

4. Concentratie op één enkel principe (*hedgehog principle*) in een cultuur van discipline;
5. Technologie als versneller van het hedgehog principe;
6. Gestage volharding.

Op basis van deze principes kon ik zelf in 2005 stellen dat *Good to Great* een voortreffelijk boek is. Maar in de praktijk blijkt, zo is mijn ervaring, dat de meeste managers en bestuurders van het boek onthouden dat je niet moet zeuren over strategie en organisatie. Je moet eerst 'get the right people on the bus and the wrong people off the bus' en vervolgens bepalen wat je gaat doen. Het gevolg daarvan is dat te weinig of zelfs géén aandacht wordt besteed aan de systemische aspecten van de organisatie. Te stellen dat je mensen moet hebben die niet zeuren maar aanpakken, is niet voldoende. Zoals de Amerikaanse psycholoog McCall heeft aangetoond, blijken op basis van hun geleverde prestaties goede managers vaak de verkeerde personen te zijn om innovaties, nieuwe businessmodellen etc. door te voeren, zoals die doorgaans nodig zijn voor de continuïteit op de langere termijn. En dan nog, als de goede mensen gevonden zijn, dan is dat slechts een eerste stap naar succes: 'Because the capabilities of organizations are a function of resources other than people, and of elements beyond just resources, namely, processes and values.'³ Geen enkele organisatie kan zonder een aantal managers die voldoen aan de karakteris-

tiek van altruïstisch-politieke persoonlijkheid. Dit zijn mensen die niet vragen wat er moet gebeuren, maar zien wat er gedaan moet worden en dit aanvaarden, ongeacht hun bevoegdheden en middelen. Maar dit type persoon heeft een broertje, de op macht beluste politieke persoonlijkheid, die meestal wel zo slim is om zijn machtsstreven te verhullen met altruïstisch gedrag.⁴ Daaraan gaat menig onderneming ten onder.

Juist het vermogen om te experimenteren, open te staan voor serendipiteit en bewust *tight control* vermijden, zo laten Christensen & Raynor in hun *The Innovator's Solution* en Simons in zijn *Levers of Organization Design*⁵ zien, geeft een onderneming meer kans op overleven dan één principe in een cultuur van discipline (dat werkt alleen in een eerste groeifase).

In 2005, zo realiseer ik mij, heb ik mij vooral laten leiden door de principes 1, 3 en 6, een reflectie van mijn eigen waarden en van mijn observatie dat die werken. Ik had beter moeten opletten en moeten zien, dat een stroming snelle managers (dus niet echte leiders) aan de haal zou gaan met Collins' principes 2, 4 en 5, de 'short cuts'.

Wat leren we van het onderzoek van Collins en de twee academische papers? Wijsheid in management is een kwestie van principes, niet van vaste oorzaak-gevolgrelaties of correlaties. Niet wat werkt is een constante, de wil om te slagen is een constante. Hoe dat ingevuld moet worden hangt altijd weer af van concrete omstandigheden.

3. Christensen, C. M. & M.E. Raynor (2003). *The innovator's solution: creating and sustaining successful growth*. Boston, Mass.: Harvard Business School Press, p. 183.

4. Jensen, M. C., & Meckling, W. H. (1994). *The Nature of Man*: SSRN.

5. Simons, R. (2005). *Levers of Organization Design: How Managers Use Accountability Systems for Greater Performance And Commitment*. Boston, Mass.: Harvard Business School Press.