

***De invloed van universitaire aanwezigheid op de kwaliteit van lokaal
rekenkamer onderzoek***

Wybe D.J.H. Kloppenburg

Universiteit Twente

Begeleider: Dr. P-J. Klok

Meelezer: Dr. H. van der Kolk

27-aug-12

Voorwoord

Voor u ligt de scriptie welke tot stand is gekomen in het kader van de bacheloropdracht ter afsluiting van de bachelor bestuurskunde aan de Universiteit Twente.

Mijn dank gaat uit naar mijn begeleider Dr. P-J Klok die mij bijgestaan heeft tijdens het gehele traject van deze scriptie. Verder wil ik ook Dr. H. van der Kolk bedanken voor het begeleiden van het onderzoeksvoorstel in het kader van het inleidende vak als ook het meelezer zijn van de scriptie in de uitvoerende en afrondende fase. Zonder de begeleiding en (bij)sturing van beiden zou deze scriptie niet zo in zijn huidige vorm voor u liggen.

Ook mijn familie en vrienden wil ik bij deze bedanken voor het vertrouwen en de steun.

Wybe Kloppenburg

24 Augustus 2012

Samenvatting

Lokaal rekenkameronderzoek ondersteunt de raad met gevraagd en ongevraagd onderzoek. De lokale rekenkamer kan bestaan in meerdere vormen, daarbij is het mogelijk dat wetenschappers zitting hebben in de lokale rekenkamer. Het is denkbaar dat de aanwezigheid van wetenschappers invloed heeft op de kwaliteit van het geleverde onderzoek en concreet de kwaliteit van geproduceerde rapporten. In dit onderzoek wordt getracht vast te stellen of dit effect bestaat. Dit wordt gedaan door het eisenkader van Hoogerwerf en Zoutendijk (1990) toe te passen op een selectie rapporten van lokale rekenkamers. De conclusie is dat alle rapporten vrij hoog scoren en dat de spreiding vrij klein is. Na statistische toetsing blijkt dat het te onderzoeken effect niet aanwezig is. Wel blijkt de grootte van de gemeente een positief effect te hebben op de mate waarin aan de eisen uit de beleidspraktijk wordt voldaan. Verder lijkt een verklaring van de vrij hoge scores het vrij simpele type procesonderzoek te zijn dat voornamelijk wordt uitgevoerd.

Inhoudsopgave

1.	Inleiding en onderzoeksvraag	6
2.	Theorie	8
2.1.	Kwaliteit beleidsevaluatie	8
2.1.1.	Wetenschappelijke eisen	9
2.1.2.	Eisen uit de beleidspraktijk.....	9
2.1.3.	Andere theorieën	12
2.1.4.	Positieve invloed wetenschappers.....	13
2.2.	Interveniërende variabelen	13
2.2.1.	Type onderzoek.....	13
2.2.2.	Grootte van de gemeente	14
2.2.3.	Externe of Interne uitvoering rapport	15
3.	Methoden en Technieken	15
3.1.	Onderzoeksontwerp.....	15
3.1.1.	Doelgerichte Steekproef.....	16
3.1.2.	Bedreigingen	17
3.2.	Data verzameling.....	17
3.3.	Case selectie	19
3.4.	Data analyse	20
3.5.	Operationalisatie.....	21

4.	Resultaten.....	31
4.1.	Kwalitatieve resultaten.....	31
4.1.1.	Conclusie.....	36
4.2.	Wetenschappelijke Eisen Vis-à-vis Eisen uit de beleidspraktijk	37
4.2.1.	Standaard Statistische Verkenning	37
4.2.2.	De onafhankelijke steekproef T-Test.....	39
4.3.	De sets gemeentes	40
4.3.1.	Gematchde paren	41
4.3.2.	Grootte van de gemeenten	42
5.	Conclusie.....	44
5.1.	Beantwoording Onderzoeksvraag	44
5.2.	Oorzaken en gevolgtrekkingen.....	44
5.3.	Aanbevelingen voor vervolgonderzoek	46
5.4.	Eindconclusie	47
	Literatuurlijst.....	48
	Bijlagen	50

1. Inleiding en onderzoeksvraag

In dit onderzoek richten we ons op de vraag of er een invloed is van de aanwezigheid van wetenschappers in lokale rekenkamers op de kwaliteit van het onderzoek dat zij leveren. Sinds 1 januari 2006 is in de gemeentewet geregeld dat elke gemeente een eigen rekenkamer of rekenkamerfunctie moet hebben. Het staat de gemeentes tot op zekere hoogte vrij om de rekenkamers zelf samen te stellen. Hierbij wordt nog een onderscheid gemaakt tussen de rekenkamerfunctie en de zelfstandige rekenkamer. Het belangrijkste verschil tussen de rekenkamer en de rekenkamerfunctie is dat raadsleden geen zitting mogen hebben in een rekenkamer en wel binnen een rekenkamerfunctie.

Het kan ook voor komen dat er binnen een rekenkamer wetenschappers zitting hebben. Binnen dit onderzoek word ervan uitgegaan dat wetenschappers in staat zijn te helpen met de problematiek die Hoogerwerf en Zoutendijk (1990) aanstippen als zij Patton (1978) citeren, die stelt dat er een verschil is tussen de cultuur van onderzoekers aan de ene kant en de cultuur van politici en ambtenaren aan de andere kant.

In lijn met deze stelling conceptualiseren en operationaliseren Hoogerwerf en Zoutendijk (1990) de kwaliteit van beleidsevaluatieonderzoek via twee soorten eisen, wetenschappelijk methodologische eisen, en eisen uit de beleidspraktijk. Zij stellen ook; *"In een optimale situatie wordt aan beiderlei eisen voldaan. Dit neemt niet weg dat in feite tussen die tweeërlei eisen veelal een zekere spanning bestaat"* (Hoogerwerf & Zoutendijk, 1990, p. 7). Deze spanning bestaat uit een aantal tegenstrijdige belangen, als bijvoorbeeld de tijd: daar waar vanuit een wetenschappelijk standpunt vaak een onderzoek meer tijd nodig heeft om grondig uitgevoerd te worden is vanuit de beleidspraktijk een noodzaak tot snelle resultaten om zo het beleid vlug bij te kunnen sturen. De vraag is of de betrokkenheid van wetenschappers enige invloed uitoefent op de mate van vervulling van deze eisen. Wetenschappers zouden in staat kunnen zijn om de verschillende kanten van de wetenschappelijk methodologische eisen en de toepasbaarheid, of bruikbaarheid van een evaluatieonderzoek, te verenigen.

Hoogerwerf en Zoutendijk (1990) stellen dat veel mensen in het openbaar bestuur vinden dat de kwaliteit en de bruikbaarheid van de gedane onderzoeken tekortschieten. Als nu blijkt dat de aanwezigheid van wetenschappers in de rekenkamers tot een betere kwaliteit van evaluatieonderzoeken leidt dan zou het beter zijn voor de kwaliteit en toepasbaarheid van lokaal rekenkameronderzoek om meer wetenschappers in de rekenkamers te hebben.

De algemene vraag van dit onderzoek is, wat de invloed is van universitaire aanwezigheid op de kwaliteit van lokaal-rekenkamer onderzoek. Er zijn meerdere manieren van universitaire aanwezigheid te bedenken: de aanwezigheid van wetenschappers in de rekenkamercommissie, uitbesteding van het onderzoek aan een universiteit of het uitvoeren door een student van een universiteit. Deze verschillende manieren kunnen allemaal op hun eigen manier een invloed hebben op de kwaliteit van het onderzoek. Dit onderzoek richt zich alleen op de aanwezigheid van wetenschappers in de lokale rekenkamers en de invloed daarvan op de kwaliteit van de lokale rekenkamerrapporten.

Bij de vraag wat de beleidsevaluatie- en methodologische kwaliteit van evaluatieonderzoek is, moet getheoretiseerd worden hoe deze kwaliteit is te benoemen, en daarna verder is te operationaliseren om bruikbaar te zijn in dit onderzoek. Als we zoeken naar theorie aangaande de kwaliteit en kwaliteitseisen van beleidsevaluatieonderzoek komen we uit bij de auteurs Hoogerwerf en Zoutendijk. Zij hebben zowel in "Beleidsevaluatie" (Bressers & Hoogerwerf, 1995) alsook via een praktische toepassing in "Naar een evaluatiesysteem voor overheidbeleid" (Hoogerwerf & Zoutendijk, 1990) een duidelijke uiteenzetting gegeven wat de belangrijkste kwaliteitseisen zijn voor evaluatieonderzoek en ook specifiek voor evaluatierapporten. Zij bezien de kwaliteit van het evaluatierapport via een stelsel van eisen, waarbij deze opgedeeld zijn in eisen vanuit de wetenschap en eisen vanuit de beleidspraktijk. In dit onderzoek worden dezelfde maatstaven voor de kwaliteit gebruikt, hierbij wordt respectievelijk gebruikt gemaakt van de wetenschappelijke eisen en de eisen uit de beleidspraktijk om de methodologische- en beleidsevaluatiekwaliteit te meten.

Hoofdvraag en deelvragen

- o Wat is de invloed van de aanwezigheid van wetenschappers met een universitaire aanstelling in lokale rekenkamers op de mate waarin in rekenkameronderzoeken voldaan wordt aan de wetenschappelijke eisen en de eisen uit de beleidspraktijk voor beleidsevaluaties?
 - Wordt in lokale rekenkameronderzoeken van rekenkamers waar wetenschappers met een universitaire aanstelling in zitten meer of minder voldaan aan de wetenschappelijke eisen voor beleidsevaluaties, dan bij de onderzoeken van rekenkamers waar geen wetenschappers met een universitaire aanstelling in zitten?
 - Wordt in lokale rekenkameronderzoeken van rekenkamers waar wetenschappers met een universitaire aanstelling in zitten meer of minder voldaan aan de eisen uit de

beleidspraktijk voor beleidsevaluaties, dan bij de onderzoeken van rekenkamers waar geen wetenschappers met een universitaire aanstelling in zitten?

Het rapport is verder als volgt opgebouwd. Het tweede hoofdstuk gaat in op de theorie die in dit onderzoek wordt gebruikt. Hoofdstuk drie gaat in op de methoden en technieken, dus hoe is dit onderzoek opgebouwd en uitgevoerd. Hoofdstuk vier gaat in op de resultaten, zowel kwalitatief als kwantitatief. Hoofdstuk vijf bevat de conclusie, dus de beantwoording van de onderzoeksvragen, opvallende zaken uit het onderzoek en de aanbevelingen voor vervolgstudie.

2. Theorie

2.1. Kwaliteit beleidsevaluatie

Hoogerwerf en Zoutendijk (1990) refereren aan Hoogerwerf (1986, pp. 268-279) (1988) als zij schrijven; *"Rond de departementale beleidsevaluatie doen zich verscheidene problemen voor. Tot deze problemen behoren in het algemeen gesproken, volgens topfunctionarissen in het openbaar bestuur: een tekort aan evaluatie, onvoldoende kwaliteit van de evaluatie, te weinig gebruik van evaluatierapporten in de beleidsvorming en onvoldoende duidelijkheid van de beleidsdoeleinden waaraan evaluatiecriteria kunnen worden ontleend."*

Zij stellen dat de doelstelling van hun onderzoek is het verminderen van de problemen van departementale beleidsevaluatie. Hiertoe gaan zij over tot het vaststellen van de eigenschappen die de kwaliteit van een beleidsevaluatie bepalen en specifiek die van het evaluatieproduct, namelijk het evaluatierapport.

De wetenschap en het openbaar bestuur worden door hen gezien als twee verschillende systemen, maar de belangrijke en complexe problemen van overheidsbeleid doen hen aanbevelen om toch ook wetenschappelijke kennis te mobiliseren. Hierdoor operationaliseren zij de mate van kwaliteit in twee dimensies, te noemen de eisen vanuit de wetenschap en de eisen vanuit de beleidspraktijk. Ze stellen duidelijk dat beide typen eisen op een zekere voet van spanning met elkaar staan. Als voorbeeld wordt genoemd de korte termijn waarop de beleidspraktijk graag een evaluatie heeft, in tegenstelling tot de wetenschappelijk onderzoeker die weet dat er enige tijd moet zijn verlopen voordat de effecten van het beleid zichtbaar zijn. Ondanks deze tegenstelling zal er toch gestreefd moeten worden om beiden serieus te nemen (Hoogerwerf & Zoutendijk, 1990, p. 7).

2.1.1. *Wetenschappelijke eisen*

Hoogerwerf en Zoutendijk(1990, pp. 14-18), gaan in op drie velden van eisen met betrekking tot de wetenschappelijke eisen voor goede beleidsevaluatie:

- Ten eerste de theoretische eisen waarbij aandacht is voor de formulering van de probleemstelling, de begripsomschrijvingen, de hypothesen en de samenhang tussen begrippen en hypothesen.
- Ten tweede de methodisch technische eisen waarbij aandacht is voor de geldigheid en betrouwbaarheid van de onderzoeksresultaten. Hierbij leggen ze de geldigheid uit als dat men werkelijk onderzoekt wat men beoogt te onderzoeken. Onder de betrouwbaarheid verstaan ze dat een herhaald onderzoek onder dezelfde omstandigheden hetzelfde resultaat oplevert.
- Ten derde de onderzoekstechnische eisen. Deze behandelen in principe de evaluatie als zodanig, dus hoe is het onderzoek uitgevoerd? Voornaamste daarbij is de operationalisering van de evaluatiecriteria aan de hand waarvan kan worden vastgesteld of de situatie in het beleidsveld beantwoordt aan de maatstaf van de beoordeling.

Verder gaan Hoogerwerf en Zoutendijk (1990) nog in op criteria die op alle drie de velden van toepassing kunnen zijn. Genoemd worden continuïteit van onderzoek ten opzichte van eerder en toekomstig onderzoek, de originaliteit ten opzichte van bestaande kennis, de objectiviteit, de systematiek, de analyse, de exactheid en de taalbeheersing.

Hierna komen zij tot een specificatie van de eisen welke te vinden zijn onder **bijlage 2** (deze lijst is aangepast aan dit onderzoek, zie 3.4 Data analyse).

2.1.2. *Eisen uit de beleidspraktijk*

Hier volgt nu eerst een korte samenvatting van enkele punten die Hoogerwerf en Zoutendijk gebruiken om de achtergrond van de eisen vanuit de beleidspraktijk toe te lichten.

Hoogerwerf en Zoutendijk (1990) beginnen met de verschillende typen onderzoek, te noemen het fundamenteel en het toegepast onderzoek. Fundamenteel onderzoek is daarbij onderzoek zonder een bepaalde praktische toepassing, terwijl toegepast onderzoek daar juist wel naar streeft. Echter; *"De eisen die vanuit de beleidspraktijk worden gesteld, betreffen uiteraard vooral de bruikbaarheid van het onderzoeksrapport voor het beleid."* (Hoogerwerf & Zoutendijk, 1990, p. 22).

Van belang is ook, vanuit de beleidsdoeleinden bezien, dat er informatie wordt verschaft, of het beleidsprobleem verminderd wordt. Bovendien; *"..., is het belang van evaluatieonderzoek met name gelegen in het verschaffen van informatie over de bijdrage van afzonderlijke instrumenten tot het bereiken van de beleidsdoeleinden."* (Bressers, Klok, Kuks, & Lulofs, 1988).

Ook informatie over het tijdstip waarop doeleinden van het beleid bereikt kunnen zijn is belangrijk. En los van het beleid zelf, is ook het proces van belang, zodat bij het opzetten van een evaluatieonderzoek het wenselijk is om afstemming te hebben op het gebruik van het rapport bij het ontwerpen, bepalen en uitvoeren van beleid.

"Dit betekent dat bij iedere evaluatieconclusie en -aanbeveling uitdrukkelijk wordt aangegeven op welke argumenten deze steunen" (Hoogerwerf & Zoutendijk, 1990, p. 24). Hoogerwerf en Zoutendijk beargumenteren dat de resultaten van een evaluatieonderzoek overtuigend zouden moeten worden gepresenteerd en zodat ze beter aansluiten op de gang van zaken in een beleidsproces. Bij deze argumentatie gaan ze onder andere in op werk van Patton (1978, pp. 63-64), waar hij ingaat op de verschillende culturen tussen onderzoekers en ambtenaren/politici. Hij concludeert verder uit onderzoek dat slechts twee factoren van belang zijn bij het wel of niet gebruiken van een onderzoek, namelijk: Politieke overwegingen(of het onderzoek het beleid wel of niet steunt) en de Persoonlijke factor (of er een persoon of groep is die geïnteresseerd is in de evaluatie en de informatie daaruit). Pröpper (1987, 1989a, 1989b) , betreft de stelling dat een beleidsevaluatie expliciet als een argumentatie moet worden ingericht. In het verlengende daarvan stelt hij dat als een beleidsevaluatie, óók door de gebruiker, wordt gezien en geïnterpreteerd als een vorm van argumentatie, dat dan de mate waarin deze gebruikt zal worden, groter zal zijn.

Tevens is de relatie van onderzoek met de beleidstheorie van belang. Daarbij zal een toegepast onderzoek trachten een of meer veronderstellingen uit de beleidstheorie te toetsen. Als laatste wordt weer teruggekoppeld naar de relatie tussen evaluatieonderzoek en het ontwerpen van beleid: er wordt gekeken naar één van de methoden voor het ontwerpen van beleid. Aan de hand hiervan komen er nog

de eisen van aandacht voor het beleidsprobleem, de doeleinden, de instrumenten en de effecten van het beleid.

Als laatste moet een evaluatierapport ook doelmatig zijn. Het moet zowel nieuwe informatie verschaffen en toevoegen aan bestaande kennis, terwijl er eveneens een redelijke verhouding moet zijn tussen de kosten van de evaluatie en de baten voor het beleid.

Hoogerwerf en Zoutendijk (1990, p. 28) geven aan dat de eisen vanuit de beleidspraktijk vooral zijn ingegeven door het doel van het beleid, namelijk zo doeltreffend, doelmatig en aanvaardbaar mogelijk vorm te geven aan maatschappelijke ontwikkelingen. Als uitgangspunt voor de eisen wordt het beginsel gekozen dat evaluatieonderzoek op de beleidspraktijk moet afgestemd zijn. Om die afstemming te specificeren gebruiken zij theorie over beleidspraktijk zoals hierboven samengevat.

De uitgebreide lijst van eisen zoals Hoogerwerf en Zoutendijk (1990, pp. 28-31) die hebben geformuleerd is te vinden onder **Bijlage 3** (deze lijst is aangepast aan dit onderzoek, zie 3.4 Data analyse). Hier een korte samenvatting van de hoofdpunten:

- Het rapport verwijst naar de beleidspraktijk.
- Het rapport verwijst naar de (beleids)doeleinden.
- Het rapport toetst een beleidstheorie uit de beleidspraktijk.
- Het rapport geeft een beoordeling aan de hand van maatstaven.
- Het rapport verwijst naar beleidsinstrumenten en gaat na of deze hebben bijgedragen tot het bereiken van hun doelstelling.
- Het rapport verwijst naar de gegevensbronnen waaruit geput wordt en naar het perspectief van waaruit het onderzoek gedaan wordt.
- Het rapport besteedt aandacht aan de (neven)effecten van het beleid en waardeert ze als positief en negatief en tracht deze te vergelijken.
- Het rapport bevat conclusies waarin een beoordeling van het gevoerde beleid wordt gegeven.
- Het rapport geeft aan waarin het beleidsuitvoering en het beleidsveld te beïnvloeden is en geeft daarbij advies aan hoe dit te doen.
- Het rapport heeft een goede opbouw.
- Het rapport is duidelijk geschreven, ook voor anderen dan wetenschapsbeoefenaars.

- De financiële kosten bleven binnen de begroting, het rapport werd opgeleverd binnen de overeengekomen termijn en dit tijdstip was zodanig dat het rapport bruikbaar was binnen het beleidsproces.
- Het rapport geeft een afweging ten aanzien van de maatschappelijke gevolgen van het beschikbaar komen van het rapport.

2.1.3. *Andere theorieën*

Ter verbreding van de theorie over beleidsevaluatie wordt nu gekeken naar Beryl Radin, die in zijn boek 'Beyond Machiavelli: Policy analysis comes of age' (2000) ook ingaat op de twee culturen van politiek en (wetenschappelijke) analyse. In hoofdstuk 4 gaat hij in op de twee verschillende culturen en de weerstand van de politiek tegen wetenschappelijke beleidsanalyse. Hierbij richt hij zich meer op de bedenking dat beleidsevaluatie te klinisch is in zijn uitvoering en niet voldoet aan de eisen die het politieke proces eraan stelt. Hij behandelt verder de eisen die gesteld worden aan beleidsanalyse vanuit het macro en micro niveau van politiek. In hoofdstuk 5 gaat hij meer in op de methodologische grondslag van de beleidsevaluatie, o.a. komt de (economische) kosten/baten analyse aan bod.

Al met al heeft Radin (2000) in zijn theorie en onderbouwing van de twee culturen, veel overeenstemming met de twee velden van eisen van Hoogerwerf en Zoutendijk. Het verschil zit hem er vooral in dat hij meer uitgaat van grotere vanzelfsprekende aanwezigheid van de wetenschappelijke methodologische competentie en zich meer focust op de kant van de politieke bruikbaarheid van evaluatieonderzoek, terwijl Hoogerwerf en Zoutendijk beide velden zien als mogelijk niet voldoende uitgewerkt. Daarbij zijn de punten van Radin over de politieke eisen veel algemener en gericht op het Amerikaanse systeem. Een functionele uiteenzetting van methodologische eisen ontbreekt geheel.

Ook in "Public policy analysis: an introduction" (Dunn, 2008) wordt ingegaan op eisen van goed evaluatieonderzoek zowel vanuit de strikt methodische kant als ook de meer deliberatieve kant, zoals in Hoofdstuk 8, "Developing Policy Arguments" (Dunn, 2008, pp. 377-420). Dit zijn echter allemaal eisen die verspreid over het boek staan -daar dit ook een algemeen boek is over beleidanalyse- en vaak niet op dezelfde manier worden geëxpliciteerd zoals Hoogerwerf en Zoutendijk (1990) dat doen.

De theorie van Hoogerwerf en Zoutendijk (1990) wordt dusdanig ondersteund door voorgenoemde andere auteurs en lijkt goed toepasbaar voor dit onderzoek, dat wordt volstaan met het toepassen van hun theorie en uitwerking.

2.1.4. Positieve invloed wetenschappers

De stelling die ten grondslag ligt aan dit onderzoek, namelijk dat de aanwezigheid van wetenschappers in de rekenkamers leidt tot betere onderzoeken kan onderbouwd worden met de theorieën die de tegenstelling behandelen tussen het domein van de onderzoeker en het domein van de politiek, zoals we deze hierboven gezien hebben.

De aanwezigheid van wetenschappers zou ten eerste de focus moeten leggen op wetenschappelijke methodologische kwaliteit. Veelal wordt simpelweg als probleem van beleidsevaluatie aangeduid dat de kwaliteit niet voldoende is, dit zou door wetenschappers gewaarborgd kunnen worden.

Als we kijken naar de eisen vanuit de beleidspraktijk dan zou beargumenteerd kunnen worden dat wetenschappers met een aanstelling aan een universiteit veelal in een positie staan waar zij zich zowel bezighouden met de harde wetenschap via onderzoek, alsook de vertaalslag moeten maken naar bruikbare kennis via het onderwijs en publicaties/lezingen e.d. Hierdoor zou men over de expertise moeten beschikken om beide gebieden van aandacht, dus zowel de wetenschappelijke eisen als ook de eisen uit de beleidspraktijk, goed te kunnen ondervangen in lokaal rekenkameronderzoek. In een ideaal geval zouden de wetenschappers een brug kunnen slaan tussen verschillende eisen, om zo tot een algeheel betere vervulling van de beiden sets eisen te komen.

2.2. Intervenierende variabelen

Buiten de genoemde variabelen waar dit onderzoek primair naar kijkt (aanwezigheid van wetenschappers en de kwaliteit van de rapporten) zijn er ook nog mogelijke derde variabelen die ook een effect zouden kunnen hebben op de afhankelijke variabele, de kwaliteit van de rapporten.

2.2.1. Type onderzoek

Bressers en Hoogerwerf onderscheiden in "*Beleidsevaluatie*" (1995) drie soorten beleidsevaluatie; de evaluatie van inhoud, processen en effecten. In dit onderzoek richten we ons op de twee laatste soorten onderzoek. Ten eerste procesonderzoek, hierbij richt het onderzoek zich voornamelijk op hoe het proces

doorlopen is. Dit proces kan zijn het totstandkomingproces maar ook bijvoorbeeld het uitvoeringsproces. Hierbij wordt dus niet specifiek gekeken naar de effecten die het beleid (mogelijk) te weeg heeft gebracht. Dit wordt uiteraard wel gedaan bij effectenonderzoek, dit onderzoek is weer op te splitsen in meerdere soorten, zoals Van de Graaf en Hoppe(1996) dat doen. Zij stellen dat er 4 soorten onderzoeksoptellingen zijn voor beleidsevaluatie, in volgorde van complexiteit te noemen: situatieonderzoek, effectenonderzoek, doelbereikingsonderzoek, en effectiviteitsonderzoek.

Situatieonderzoek kijkt puur en alleen naar het veld en inventariseert welke veranderingen er opgetreden zijn na de invoering van het beleid. En er wordt niet specifiek ingegaan op de vraag of deze veranderingen toe te schrijven zijn aan het gevoerde beleid

Effectenonderzoek kijkt naar de effecten die aan de hand van het gevoerde beleid zijn opgetreden, hierbij wordt specifiek gekeken of deze effecten opgetreden zijn vanwege het gevoerde beleid. Hierbij wordt echter geen onderscheid gemaakt tussen de beoogde effecten en de eventuele neveneffecten.

Doelbereikingsonderzoek verschilt van situatieonderzoek in de te onderzoeken veranderingen, hierbij wordt specifiek gekeken naar de veranderingen die het beleid beoogt te bereiken en of deze optreden, echter er wordt weer niet gekeken of deze verandering toe te schrijven zijn aan het beleid

Effectiviteitsonderzoek is de meest uitgebreide vorm en hier wordt zowel gekeken of het beleid de gewenste verandering te weeg heeft gebracht, alsook of het dit alleen gedaan heeft, dus of er geen andere factoren meespelen. Dus niet alleen had het beleid effect, maar was het ook het effect wat van tevoren zo voorspeld was.

Deze benaderingen hebben een oplopende mate van complexiteit van onderzoek, en we gaan er hierbij vanuit dat het mogelijk is dat deze mate van complexiteit ook invloed heeft op de te meten afhankelijke variabele, de kwaliteit van het rapport.

2.2.2. Grootte van de gemeente

Een andere variabele die van belang zou kunnen zijn voor de uitkomsten van dit onderzoek is de grootte van de gemeente en de indirecte gevolgen daarvan voor de samenstelling van de rekenkamer en het budget dat beschikbaar is voor rekenkameronderzoek. Grotere gemeentes zouden een grotere groep van personen hebben die zitting kunnen hebben in een rekenkamer en/of ook meer middelen en budget om rapporten extern uit te laten voeren of juist een betere ambtelijke en/of wetenschappelijke staf te hebben ter ondersteuning van de rekenkamer.

2.2.3. Externe of Interne uitvoering rapport

In het verlengde van het voorgenoemde kan de wijze van uitvoering ook invloed hebben op de rapporten. Het is niet ongebruikelijk dat rekenkamers rapporten in opdracht laten uitvoeren door externe adviesbureaus of universiteiten. Te denken valt dat deze rapporten op een andere manier worden uitgevoerd dan een rekenkamer dat zelf zou doen. Dit zou invloed kunnen hebben op het verband tussen wetenschappers en de kwaliteit van de rapporten, vandaar wordt deze variabele wordt meegenomen.

3. Methoden en Technieken

3.1. Onderzoeksontwerp

Om de onderzoeksvragen te beantwoorden zal de volgende strategie gebruikt worden. De onafhankelijke variabele, dus de variabele waartegen we willen testen is de aanwezigheid van wetenschappers in de rekenkamers. De rekenkamers zijn echter niet de uiteindelijke eenheid van analyse, dit zijn de rekenkamerrapporten. We willen dus wat zeggen over de kwaliteit van de rapporten die de rekenkamers afleveren. Hiervoor gebruiken we de rekenkamerrapporten ook als de eenheid van observatie. Van belang is om op deze onafhankelijke variabele, variëteit te hebben, dus zowel rekenkamers met als zonder wetenschappers. De afhankelijke variabele waar we naar kijken is de kwaliteit van de evaluatierapporten. Hierop willen we zoveel mogelijk vergelijkbaarheid, zodat niet andere eigenschappen van de rapporten een invloed hebben op de te onderzoeken relatie. Hieruit volgt ook dat de evaluatierapporten onze eenheden van onderzoek zijn.

Het concept "aanwezigheid van wetenschappers in de rekenkamers" wordt meetbaar gemaakt door de leden van de rekenkamers te controleren op een wetenschappelijke aanstelling aan een universiteit.

Het concept kwaliteit van evaluatierapporten wordt meetbaar gemaakt door de in de theorie besproken wetenschappelijke eisen en eisen uit de beleidspraktijk te gebruiken om een selectie van rapporten te waarderen. De resultaten hiervoor worden genoteerd en per blok van eisen opgeteld. Hierdoor zal per rapport een score bepaald worden die wat zegt over de kwaliteit van het rapport. Via een statistische toets zal worden vastgesteld of er sprake is van covariatie tussen de aanwezigheid van wetenschappers in de rekenkamers en de kwaliteit van de rapporten, dit dan dus zowel voor de twee soorten eisen afzonderlijk, als ook tezamen.

Dit onderzoek is een niet- (quasi-)experimentele studie. Daarbij kunnen we binnen de niet-experimentele ontwerpen spreken van een kwantitatieve correlatiestudie die dan weliswaar relatief weinig cases heeft, maar toch op een kwantitatieve manier behandeld wordt (Shadish, Cook, & Campbell, 2002).

Uit Babbie (2001) volgt dat dit onderzoek een unobtrusive onderzoek is, een onderzoek waarbij de onderzoeker niet een bepaalde vorm van experiment gebruikt, dus geen inmenging is in de werkelijkheid maar slechts van een afstand gekeken en gemeten wordt. Een verdere specificering is dat dit onderzoek "inhoudsanalyse" toepast, dat wil zeggen dat uit beschikbare bronnen die ooit gepubliceerd zijn een codering (operationalisatie) wordt gemaakt van de te meten variabelen, waarna vervolgens door verdere analyse een uitspraak wordt gedaan over een eventueel causaal verband. Babbie (2001, p. 310) merkt ook op dat bij het coderen van latente inhoud, wat ook de manier is zoals het in dit onderzoek gebeurt, het risico is dat de onderzoeker voor een gedeelte zijn persoonlijke eigenschappen laat meewegen in het doen van de meting. Ter verduidelijking, bij manifeste inhoud meten we het aantal malen dat een bepaald woord of een zinsnede voorkomt, dit is een hele directe en duidelijke manier van meten. Het probleem is alleen dat daarmee niet een dieperliggende inhoud vast te stellen is, of zoals in dit onderzoek, dat het vrijwel onmogelijk is om directe woorden of zinsneden aan te wijzen waarmee de kwaliteit van een evaluatierapport is vast te stellen. Daarom is het niet anders mogelijk dan, zoals gezegd, gebruik te maken van latente inhoudsanalyse in dit onderzoek.

3.1.1. Doelgerichte Steekproef

Zoals al is vastgesteld, zal er geen experiment gedaan worden. Hierdoor is het ook niet mogelijk om een gerandomiseerde selectie te maken. Dit betekent dat we doelgerichte sampling toepassen bij de selectie. De selectie van cases zal plaatsvinden op basis van criteria van gelijkheid die we belangrijk vinden. Dit is dan zoals Shadish, et al (2002, pp. 375-376) et al het zeggen, purposive sampling of typical instances, hierbij ligt de nadruk op het vinden van cases die zoveel mogelijk gelijk zijn behalve op de onafhankelijke variabele. De methodologische valkuil ligt hier in de eventuele derde variabelen die ook een invloed zouden kunnen uitoefenen op de causale relatie die onderzocht wordt. Shadish, et al (2002, pp. 380-382) gaan hier op in als zij spreken over ruling out irrelevancies. Daar behandelen zij variabelen die als irrelevant geacht worden, en daardoor niet homogeen hoeven te zijn maar ook heterogeen mogen en kunnen zijn. In het ideale geval worden deze irrelevanties meegenomen in het onderzoek zodat vastgesteld kan worden dat zij geen meetbare invloed hebben op de te onderzoeken causale relatie. Dit is

echter niet altijd mogelijk, zeker in onderzoeken met een kleine dataset kan het zoeken naar irrelevanties en het goed testen daarvan de eigenlijke onderzoeksvraag schaden. In deze gevallen is het van belang om in ieder geval de mogelijke irrelevanties te benoemen en als aanname op te voeren in het onderzoeksmodel.

Concreet zal in dit onderzoek geprobeerd worden de variabelen; type onderzoek, grootte van de gemeente en wijze van uitvoering mee te nemen in de als mogelijke interveniërende variabelen.

3.1.2. Bedreigingen

De grootste bedreiging voor dit onderzoeksontwerp ligt in de interne validiteit: is het effect wat we gaan meten wel toe te wijzen aan de onafhankelijke variabele. Door de toepassing van niet gerandomiseerde case selectie is het risico voor selectie fouten heel groot. Door de toepassing van het principe van matching te gebruiken bij de case-selectie is het risico te verminderen. Dit is ook het verschil tussen ruling out irrelevanties en matching, bij ruling out irrelevanties worden de mogelijke derde variabelen zoveel mogelijk meegenomen om ze uit te kunnen sluiten van covariatie, dus wordt concreet ook getest of ze covariëren met de afhankelijke variabele. Bij matching worden de derde variabelen gebruikt om uit te sluiten dat ze in de invloed hebben op de uitkomst. Er wordt niets vastgesteld of ze eventueel ook covariëren met de afhankelijke variabele.

3.2. Data verzameling

De dataverzameling is tweeledig, in de eerste plaats is er voor dit onderzoek in mei 2010 een database gecreëerd op basis van informatie van de NVRR website aangaande gepubliceerde rapporten en ook op de website vermelde rekenkamers.

Het tweede deel zal zijn uit deze dataset via de hierboven genoemde eisen van case-selectie rapporten te selecteren van rekenkamers met wetenschappers en zonder wetenschappers overeenstemmend met de eigenschappen van de rekenkamers en rapporten die in de case selectie besproken zijn.

De database die al gecreëerd is zal nu qua opzet eerst verder besproken worden. De vraag die in voorbereiding van dit onderzoek gesteld werd was een inventarisatie te maken van de samenstelling van de lokale rekenkamers op het criterium wel of geen zitting hebbende wetenschappers. Dit is concreet gebeurd door aan de hand van de genoemde rekenkamers op de website via internet na te gaan wie daar allemaal zitting in hebben en van elk individueel lid via Google na te zoeken of zij werkten op, of een

aanstelling hadden op een universiteit. In de database werden per rekenkamer de volgende data opgenomen:

- Naam/Plaats van de rekenkamer
- Contactpersoon
- Provincie
- Lidmaatschap NVRR
- Aantal gepubliceerde rapporten op de website van de NVRR
- Aantal Praktiserend wetenschappers
- Aantal Niet wetenschappers

Dit heeft uiteindelijk geleidt tot een verzameling van 246 lokale rekenkamers. Verder is voor de rapporten gekeken naar de publicaties vanaf 1 januari 2009 tot 23 maart 2010. Deze rapporten zijn geanalyseerd door ze door te lezen en te zoeken op de volgende punten, ten eerste het soort onderzoek, uitgesplitst naar intern, extern en gemixt. Daarbij werd in de gevallen extern en gemixt, de (mede-)uitvoerder ook benoemd. Dit leverde de mogelijkheden: student, universiteit, extern bureau/onderzoek, hoge school, andere rekenkamer of ambtelijk op.

Bij ambtelijk is het criterium gebruikt dat het gehele onderzoek uitgevoerd is door ambtenaren, waarbij ook een groot gedeelte van de onderzoeksopzet door hen is bedacht.

In de database is per rapport opgenomen:

- Naam rapport
- Datum
- Soort onderzoek
- (Mede-)Uitvoerders
- Rekenkamer

Uit deze database zal een selectie van te vergelijken cases gemaakt moeten worden. Hierbij moeten dan de relevant punten uit de theorie meegenomen worden.

Concreet valt op als we alvast kijken naar de database dat er relatief weinig rekenkamers zijn met wetenschappers erin, namelijk 29 van de totaal 246 onderzochte. Dit levert het probleem op dat de selectie op de eigenschappen van de rekenkamers en de rapporten beperkt wordt op aantal. Daarbij zijn het dus het aantal rekenkamers met wetenschappers en logischerwijs ook hun rapporten in de minderheid, namelijk 61 van de 310 rapporten. De selectie van de rapporten zal in hoge mate geleid

worden door de mate van aanwezigheid van onze selectiecriteria onder de rapporten van rekenkamers met wetenschappers. Dit zou op zich geen methodologisch probleem moeten zijn, zolang de te vergelijken rapporten van rekenkamers zonder wetenschappers maar zoveel mogelijk gelijk zijn op de mogelijk criteria genoemd bij de case-selectie.

3.3. *Case selectie*

Voordat er werkelijk gemeten en geanalyseerd kan worden moet een selectie gemaakt worden van welke cases bekeken worden. Zoals al genoemd bij het onderzoeksontwerp moet er een variatie zijn op de onafhankelijke variabele, we willen dus zowel evaluatieonderzoeken bekijken van rekenkamers met wetenschappers alsook zonder wetenschappers.

Daarbij zijn er, zoals we bij de theorie al gezien hebben, meerdere mogelijk interveniërende variabelen. Te noemen het type onderzoek, de grootte van de gemeente en de wijze van uitvoering (intern of extern) Voor de case selectie wordt vanwege deze mogelijk interveniërende variabelen gebruik maken van het principe van matching (Shadish, et al., 2002, pp. 118-122,304-306), we gaan dus van de te onderzoeken rapporten cases zoeken die zoveel mogelijk op elkaar lijken en alleen verschillen op de te onderzoeken variabele. De aanwezigheid van wetenschappers is steeds de variabele zijn die verschillend is, terwijl de andere variabelen zoals genoemd zoveel mogelijk gelijk gehouden worden. Idealiter zou daarbij ook, zoals genoemd bij purposive sampling de andere variabelen ook apart gevarieerd moeten worden. Het voordeel hiervan is dat we ze dan mogelijk uit zouden kunnen sluiten van mogelijke covariatie waardoor de causale relatie sterker zou worden. Dit is echter niet altijd mogelijk zijn door gebrek aan cases.

Aangaande de grootte van de gemeente wordt er een matching gemaakt tussen een selectie van 4 gemeentes met rekenkamers met wetenschappers van variërende grootte met 4 andere gemeentes met rekenkamers zonder wetenschappers van gelijke grootte als de 4 eerste gemeentes. Dit levert concreet 4 paren op van gemeentes van zoveel mogelijk vergelijkbare grootte (zover deze in de dataset te vinden waren) met een variantie op de variabele aanwezigheid van wetenschappers in de rekenkamer. De paren zijn verder ook nog gematched op afstand ten op zichte van elkaar. Dit is gedaan met in het achterhoofd de gedachte dat er eventuele effecten zouden kunnen zijn van bijvoorbeeld de landsstreek en dergelijken.

Dit leverde de volgende paren op (zoals genoemd met de beperkingen van de dataset):

- Apeldoorn en Deventer (157.000 en 99.000 inwoners)
- Den Bosch en Venlo (141.000 en 100.000)

- Hardenberg en Wijk bij Duurstede (60.000 en 23.000)
- Veenendaal en Ede (63.000 en 109.000)

De beperkingen van de relatief kleine dataset leveren helaas wel op dat het type onderzoek en de wijze van uitvoering niet meegenomen kunnen worden in enige verdere vorm van matching. Dit is overigens niet enkel te wijten aan de kleine dataset. Een van conclusies die later nog verder toegelicht wordt is dat er bijna enkel alleen "simpel" procesonderzoek wordt gedaan door de lokale rekenkamers en ook dat er bijna altijd wordt uitbesteed naar externe bureaus. (Zelfs daar waar ogenschijnlijk het onderzoek intern is uitgevoerd, wordt vaak nog verwezen dat de rekenkamercommissie wel is bijgestaan door een extern bureau op bepaalde onderdelen.)

De complete case-selectie is te vinden in bijlage 1.

3.4. *Data analyse*

Hoogerwerf en Zoutendijk (1990, pp. 35-69) behandelen per rapport een kort geschreven analyse, per punt de respectievelijke eisenlijsten, te noemen de wetenschappelijke eisen te stellen aan een evaluatierapport en de eisen uit de beleidspraktijk.

Hoogerwerf en Zoutendijk (1990) noemen specifiek dat de waardering van de rapporten bij hun onderzoek slechts door één van de auteurs is gedaan dus dat het niet uitgesloten is dat een andere onderzoeker tot (iets) andere waarderingen was gekomen. Dit in lijn met wat Babbie (2001) ons zegt over het meten van latente inhoud.

Voor dit onderzoek wordt de uitgebreide behandeling van de rapporten achterwege gelaten omwille van de hoeveelheid tijd die er is voor uitvoering van dit onderzoek. Wel wordt een algemene toelichting gegeven van de waardering van de rapporten. Verder is de opsplitsing in de begrippen: expliciet, beargumenteerd, duidelijk en exact zoveel mogelijk achterwege gelaten. Om deze begrippen goed toe te passen bij de eisen is een intensievere bestudering van de rapporten nodig dan het tijdsbestek van dit onderzoek toelaat. De opdeling in eisen is al redelijk uitputtend en er wordt dus per eis alleen naar de algemene eis gekeken en niet naar de expliciete, beargumenteerdheid etc. Zie onder **bijlage 2 en 3** de aangepaste eisen naar aanleiding van Hoogerwerf en Zoutendijk (1990).

In dit onderzoek wordt een statistische toets uitgevoerd om te kijken of er een significante relatie is. Vanwege de voorgenoemde opdeling bij de eisen gebruikten Hoogerwerf en Zoutendijk 2 verschillende schalen, te weten een dichotome + - schaal en een vijfpuntschaal. Omdat de opdeling in de verscheidene begrippen achterwege wordt gelaten zullen we bij de waardering slechts een vijfpuntschaal

gebruiken. Hierbij zal zoals gebruikelijk een waardering van 1 slecht zijn, een waardering van 3 voldoende nog onvoldoende en een waardering van 5 goed zijn.

Belangrijk om steeds mee te nemen bij het berekenen van de totaalscore van een rapport is dat het gemiddeld aantal punten goed berekend moet worden. Sommige eisen zullen niet toepasselijk zijn omdat bepaalde eisen een type onderzoek veronderstellen waar niet in alle rapporten sprake van is. Daarom zullen deze eisen niet meetellen voor het gemiddelde. Dit betekent concreet dat voor de meest aanwezige rapporten in de steekproef, namelijk rapporten van procesonderzoek, de meest ingewikkelde eisen uit de eisenkaders wegvallen. Als voorbeeld zijn te noemen de eisen die voor verklarend onderzoek bedoeld zijn, zoals eis 7 van de wetenschappelijke eisen: *De afzonderlijke veronderstellingen of hypothesen die worden getoetst of zijn ontwikkeld, zijn in het rapport omschreven. Aangegeven is of het om causale verbanden gaat.* Omdat er bij procesonderzoek geen specifieke hypothese worden getoetst of ontwikkeld vervalt deze eis.

Aan het eind is dan een lijst over van rapporten met een puntenaantal waarna enkel nog de statistische toetsing overblijft om te bepalen of er een significant verband is tussen de aanwezigheid van wetenschappers in de rekenkamers en de kwaliteit van de rapporten die zij produceren.

3.5. Operationalisatie

Bij het latent coderen van rapporten is een strakke operationalisatie vooraf moeilijk, zo niet onmogelijk te maken. Dat er uiteraard wel een proces is geweest van vertaling van het kader van eisen in waarderingen van de rapporten moge duidelijk zijn. Hier volgt een beschrijving van hoe de scores van de rapporten tot stand zijn gekomen. Dit is een proces waarbij zoals bij elke vorm van latent coderen, de onderzoeker veel invloed heeft in de uiteindelijke uitkomst, echter zolang de manier van waarderen maar consequent is toegepast zou dit geen problemen moeten opleveren. Belangrijk om te noemen is dat de waardering een lopend proces is geweest waarbij zoals genoemd, er van tevoren geen vaste operationalisering is vastgesteld omdat niet duidelijk was wat er in de rapporten gevonden zou worden. Na het waarderen van alle rapporten is teruggegaan en zijn alle rapporten nogmaals bekeken om zeker te zijn dat er op een gelijke manier is beoordeeld. De scores die hierna genoemd worden en de condities waaronder ze gegeven zijn omvatten alleen de scores zoals ze werkelijk gewaardeerd zijn, er is dus geen uitgewerkte argumentatie voor de scores die niet zijn toegekend. Bij het waarderen is wel steeds het complete bereik van de schaal van 1 tot 5 in gedachten meegenomen. Het was dus bij elke eis wel mogelijk dat een rapport gewaardeerd zou kunnen worden over de hele schaal.

Er zijn echter geen scores toegekend van 1, dit omdat er simpelweg niet zo dusdanig slecht werd gescoord op één van de eisen dat een echte score van slecht voor de hele eis toegekend moest worden. Daar waar een score van 2 ontbreekt kan in het algemeen gezegd worden dat die toegekend zou zijn wanneer duidelijk geen ambigue 3 meer gegeven zou kunnen worden maar dat er echt sprake is van onvoldoende voldoen aan één of meer aspecten van de eis. De eisen die op geen enkel rapport van toepassing waren zijn vervallen.

Wetenschappelijke eisen (A. ontbreekt hier omdat daar geen eisen voor zijn.)

B. Type onderzoek

1. Als het op te maken valt uit de context of het onderzoek beschrijvend dan wel verklarend is wordt een score van 3 toegekend, uitzonderingen naar boven toe, waar explicieter wordt ingegaan op het type onderzoek krijgen een 4 toegekend. Daar waar expliciet de termen beschrijvend en verklarend worden gebruikt wordt een 5 gegeven.
2. Vervalt.
3. Bij procesonderzoek wordt het onderzochte proces ook geacht het beleid te zijn. Daar waar ingegaan wordt op doelbereiking, effectiviteit, etc. wordt een 4 toegekend, als er minder expliciet op in wordt gegaan een 3. Daar waar het rapport dusdanig anders is vormgegeven dat het niet redelijkerwijs te verwachten is dat er wordt ingegaan op effecten van het beleid of proces wordt de geen score gegeven. Als er werkelijk effecten onderzoek van beleid is uitgevoerd wordt een punt extra gegeven bovenop de bestaande score van 3 of 4.

C. Doelstelling en probleemstelling

4. Een verwijzing naar een doelstelling levert een score van 3 op, des te explicieter en duidelijker toegelicht levert een hogere score van 4 of 5 op.
5. Een vermelding van een probleemstelling met onderzoeksvragen levert een score van 5 op, een 4 wordt gegeven wanneer of de probleemstelling of de onderzoeksvragen niet geheel helder zijn. Een score van 2 wordt gegeven wanneer een probleemstelling en/of onderzoeksvragen compleet missen.

D. Theorie en begrippen.

6. Vervalt.
7. Vervalt.
8. Als er in het rapport impliciet ingegaan wordt op de begrippen uit de probleemstelling levert dit een 3 op, een verwijzing naar de begrippen in bijvoorbeeld een normenkader levert een 4 op. Als er nog explicieter op ingegaan wordt in een aparte paragraaf of hoofdstuk levert dit een 5 op.

E. Methoden van Onderzoek

9. Een genoemd en toegelichte methodische benadering levert een 5 op, als er minder uitgebreid op in wordt gegaan levert dit een 4 op. Een methodische benadering die wel impliciet is op te maken maar niet genoemd wordt levert een 3 op.
10. Een expliciete operationalisering in de vorm van begrippen vertaald in te meten variabelen, maar ook bijvoorbeeld een normenkader waaraan een proces te toetsen is, leveren een score van een 4 of 5 op, naar gelang de mate van expliciteit en beargumenteerdheid. Een score van 3 wordt gegeven wanneer niet expliciet is ingegaan op een vorm van operationalisering, maar het wel uit bijvoorbeeld de analyse op te maken is.
11. Een uitgewerkt en beargumenteerd normenkader of eisenlijst levert een 5 of een 4 op naar mate van de duidelijkheid. Een minder expliciete vermelding, maar wel op te maken uit bijvoorbeeld de verdere analyse levert een 3 op.
12. Een expliciete en duidelijke beoordeling levert een 5 op, een minder duidelijke beoordeling een 4, het ontbreken van een expliciete beoordeling, maar wel op te maken uit de interpretatie van de data levert een 3 op.
13. Er wordt verondersteld dat er op de geldigheid van het onderzoek wordt ingegaan als er een verband is tussen de vraagstelling en de conclusies, dus als er een beantwoording van onderzoeksvragen en dergelijke wordt gegeven. Is dit niet het geval dan wordt een score van 2 toegekend. Is dit wel het geval dan een score van 3. Een duidelijk reflectie bij de conclusies op de onderzoeksvragen levert een score van 4 op.
14. Bij het niet expliciet of impliciet ingaan op de betrouwbaarheid van het onderzoek, maar waar uit de zorgvuldigheid omtrent de opzet van de dataverzameling en analyse wel blijkt dat er zorg is gedragen voor de betrouwbaarheid wordt een score van 2 toegekend.

15. De mate van objectiviteit wordt vastgesteld door te kijken naar de verwijzingen naar interviews en bijvoorbeeld wederhoor aan de hand van tussenliggende resultaten. Daar waar dit onduidelijk wordt gelaten maar wel de indruk is gewekt dat er enige vorm van overleg en interactie is geweest die de objectiviteit zou kunnen waarborgen wordt een 3 gegeven. Daar waar expliciet wordt verwezen naar interviews om de data te onderbouwen een 4 en daar waar ook nog expliciet wederhoor wordt toegepast ter verificatie van de data een 5.

F. Technieken van onderzoek.

16. Een vermelding van de wijze van verzameling van data levert een 4 op, een 5 is gegeven daar waar het meest uitgebreid hierop werd ingegaan.
17. Expliciete en duidelijk vermelding van empirische gegevens, compleet met getallen etc. levert een 5 op, een vier is voor minder duidelijk vermelding, bijvoorbeeld alleen geaggregeerde gegevens, een 3 is voor niet expliciete vermelding, maar enkel korte verwijzing bij de analyse en een 2 is gegeven waarbij niet op te maken wat er uit de data-verzameling is gekomen.
18. Duidelijk onderscheid en uitleg van de technieken, zoals interviews en case onderzoek levert een 5 op. Als er minder expliciet wordt ingegaan op de technieken een 4 en een slecht impliciete verwijzing, bijvoorbeeld op te maken uit de analyse levert een 3 op.

G. Analyse van de data en conclusies.

19. Als (een deel) van de wijze van data-analyse te herleiden is uit de conclusies dan wordt een 3 gegeven, als er wel expliciet wordt ingegaan op de data-analyse wordt een 4 gegeven, als dit ook nog duidelijk en uitgebreid is wordt een 5 gegeven.
20. Als er ingegaan wordt op de vertaalslag van data naar conclusies dan wordt een 4 gegeven, als dit heel duidelijk gebeurd een 5 en zo dit niet goed gebeurd een 3.
21. Een expliciete en uitgebreide behandeling van de conclusies met daarin een beantwoording van de onderzoeksvragen levert een 5 op. Een minder duidelijke beantwoording levert een 4 op. Daar waar niet alle of niet een direct beantwoording van de onderzoeksvragen wordt gedaan, levert een 3 op.

H. Opbouw

22. Bij een heldere en gestructureerde opbouw, bijvoorbeeld ondersteund door een leeswijzer en een duidelijke hoofdstukindeling wordt een 5 gegeven. Als één van deze delen niet voldoende scoort wordt een 4 gegeven. Een 3 is gewaardeerd daar waar bijvoorbeeld een leeswijzer helemaal ontbreekt en de opbouw van het stuk ook dusdanig is dat zonder leeswijzer het verloop niet helder is.

I. Overige Eisen.

23. Aanwezigheid van voetnoten, citaten en een heldere literatuurlijst leidt tot een score van 5. Een ontbreken van één van deze delen of een mindere uitwerking levert een 4 op. Daar waar niet goedduidelijk is waar theorieën en methoden en technieken vandaan komen wordt een 3 gegeven.
24. Rapporten die doorbouwen op bestaande theorieën, zoals door de toepassing van een normenkader op de praktijk en de daaropvolgende conclusietrekking levert een 4 op. Daar waar het de vraag is wat van de resultaten nuttig is om acht op te slaan in het kader van kennisopbouw wordt een 3 gegeven. Een twee is toegekend daar waar niet vast te stellen is of het rapport iets toevoegt aan bestaand kennis.
25. Een 5 is gegeven daar waar geen twijfel bestond over de mate van exactheid van het rapport. Een score van 4 is gegeven voor lichte twijfel aangaande de exactheid van de woorden.
26. Een goede beheersing van de taal levert een 5 op, daar waar hier en daar een enkele fout te vinden is levert dit een 4 op.

Eisen vanuit de beleidspraktijk (B. ontbreekt hier omdat daar geen eisen voor zijn.)

A. Aanleiding en onderwerp van evaluatie

1. Daar waar duidelijk aan een vraag vanuit de raad of andere belanghebbende partijen wordt gerefereerd wordt een 5 gegeven. Daar waar op eigen initiatief onderzoek gedaan is, maar een probleem wordt geïllustreerd vanuit de beleidspraktijk wordt gewaardeerd met een 4. Als het niet expliciet duidelijk is vanwaar het onderzoek wordt gedaan maar wel uit het onderzoek in zijn geheel is op te maken dat er een behoefte is uit de beleidspraktijk wordt met een 3 gewaardeerd.
2. Als er wordt verwezen naar oorzaken en gevolgen van een probleem dan wordt gewaardeerd met een 5, wordt er slechts naar één van beiden verwezen dan wordt een 4 gegeven, als niet expliciet

duidelijk wordt verwezen naar oorzaken en gevolgen, maar wel een probleem wordt genoemd wordt een 3 gegeven.

3. Daar waar specifiek verwezen wordt naar een vraag vanuit de raad om een onderzoek te doen wordt dit gewaardeerd met een vijf. Daar waar door (verscheidene) belanghebbende een verzoek is ingediend tot onderzoek wordt dit gewaardeerd met een 4. Als het echter niet specifiek duidelijk is dat er een opdracht is gegeven, maar wel dat er noodzaak is vanuit de beleidspraktijk wordt gewaardeerd met een 3
4. Als er specifiek en duidelijk wordt verwezen naar een beleid waarop de evaluatie betrekking heeft dan wordt gewaardeerd met een 5. Als er meer algemeen maar nog wel duidelijk wordt aangegeven dan wordt een 4 gegeven. Als slechts in het algemeen een beleidsgebied wordt aangeduid, wordt met een 3 gewaardeerd.
5. Daar waar expliciet en duidelijk wordt verwezen naar het onderwerp van de evaluatie wordt gewaardeerd met een 5, daar waar impliciet wordt verwezen wordt met een 4 gewaardeerd. Daar waar het onderwerp uit de context van de onderzoeksvragen of probleemstelling moet worden afgeleid wordt een 3 gegeven.
6. Als er specifieke personen binnen de instanties worden genoemd waar contact mee is geweest dan wordt met een 5 gewaardeerd, als er functies binnen de instanties worden genoemd dan wordt met een 4 gewaardeerd. Als er slechts in het algemeen wordt verwezen naar een instantie dan wordt met een 3 gewaardeerd.
7. Als er direct wordt verwezen naar de raad of een belanghebbende instantie die bij het onderzoek is betrokken en dat in het onderzoek aanbevelingen zitten voor deze instanties, dan wordt gewaardeerd met een 5. Als er meer indirect wordt verwezen naar de instanties en niet specifiek dat het onderzoek op hen gericht is wordt een 4 gegeven. Daar waar uit de conclusie moet blijken voor wie het onderzoek mogelijk profijt kan hebben wordt een 3 gegeven.

C. Doelstelling en probleemstelling.

8. Een expliciete en duidelijke verwijzing naar doeleinden van het onderzoek die verband houden met de beleidspraktijk wordt gewaardeerd met een 5, daar waar minder expliciet en/of duidelijk wordt verwezen wordt gewaardeerd met een 4.

9. Een expliciete en duidelijke verwijzing naar doeleinden van bepaalde personen of organisaties wordt gewaardeerd met een 5, daar waar minder expliciet en/of duidelijk wordt verwezen wordt gewaardeerd met een 4.
10. Als er een duidelijk verband wordt gelegd met het beleid/proces en het probleem uit de beleidspraktijk en in welke mate dit beleid proces heeft meegewerkt tot een oplossing wordt een 5 gegeven. Daar waar dit minder duidelijk is maar wel direct af te leiden valt wordt met een 4 gewaardeerd. Als er slechts uit conclusies en analyse blijkt dat er toch een verband wordt gelegd word gewaardeerd met een 3.

D. Theorie en begrippen

11. Als er bij bijvoorbeeld de operationalisering of bij de theoretisering wordt ingegaan op specifieke en afzonderlijk veronderstelling uit de beleidspraktijk wordt met een 5 gewaardeerd. Wordt dit gedaan door een bijvoorbeeld een normenkader op te stellen wordt dit gewaardeerd met een 4. Wordt slechts theorie uit de literatuur gebruikt wordt gewaardeerd met een 3.

E. Methoden van onderzoek

12. Een expliciete en duidelijke vermelding van maatstaven, bijvoorbeeld in de vorm van een normenkader wordt gewaardeerd met een 5. Een minder expliciete en/of duidelijke vermelding van maatstaven levert een 4 op. Daar waar uit de analyse van data geconcludeerd moet worden dat er wel een beoordeling was aan de hand van maatstaven maar geen expliciete of duidelijke vermelding word gewaardeerd met een 3. Als het nergens uit blijkt dat er een vorm van maatstaven betrokken was bij het onderzoek wordt een 2 gegeven.
13. Daar waar een rapport outputbudgettering behandeld en er op een voldoende mate expliciet en duidelijk ook wordt ingegaan op de beleidsprestaties van de organisatie wordt met een 4 gewaardeerd.
14. Daar waar redelijkerwijs te spreken is van effectiviteitsonderzoek, of procesonderzoek waar ook ingegaan wordt op de effectiviteit van het proces ten aanzien van de doeleinden van het beleid, word met een 5 gewaardeerd als duidelijk vermeld is in hoeverre de beleidsinstrumenten hebben bijgedragen aan het bereiken van hun doelstelling. Als dit minder expliciet wordt vermeld word met een 4 gewaardeerd en een 3 word gegeven voor een niet expliciete vermelding maar waar er

wel in de conclusie op word ingegaan. Als er niet te spreken was van effectiviteitsonderzoek en ook niet word ingegaan op enige effecten van het proces word geen score gegeven.

F. Technieken van onderzoek.

15. Rapporten die zowel een duidelijk bronvermelding hebben aangaande de onderzoeksdata als ook een vermelding vanuit welk perspectief (bijvoorbeeld de raad of juist de beleidsinstantie) geredeneerd wordt met een 5 gewaardeerd, daar waar één van deze twee punten mist wordt met een 4 gewaardeerd. Als zowel de bronvermelding als het perspectief niet expliciet vermeld word maar wel (ten dele) is op te maken uit het rapport word met een 3 gewaardeerd.

G. Analyse van de data en conclusies.

16. Als er expliciet en duidelijk aandacht wordt geschonken aan de neveneffecten van het beleid en het rapport ook echt effectenanalyse toepast, dan word met een 5 gewaardeerd. Als er in procesonderzoek aandacht word besteedt aan de neveneffecten van het proces, wordt gewaardeerd met een 4. Als de neveneffecten niet direct worden gekoppeld aan zowel beleid als proces maar wel bij de conclusies of aanbevelingen genoemd worden wordt met een 3 gewaardeerd.
17. Daar waar een werkelijke afweging is van positieve en negatieve effecten en dit ook expliciet en duidelijk zo vermeld is, word met een 5 gewaardeerd. Een minder expliciete en duidelijke vermelding levert een 4 op. Daar waar uit het rapport blijkt dat er wel een afweging is geweest maar niet hoe en wat die afweging was wordt met een 3 gewaardeerd. Een 2 wordt gegeven als het niet duidelijk is te herleiden dat er ingegaan is op de positieve en negatieve effecten en dat ook geen afweging daarvan is geweest.
18. Een specifieke verwijzing naar tijdskeuzen en een confrontatie daarmee met de praktijk levert een 5 op. Een verwijzing naar enkel tijdskeuzen van het beleid of hoe in de praktijk het beleid ten aanzien van tijd en timing is uitgepakt levert een 4 op. Geen duidelijk verwijzing naar tijdskeuzen of praktische tijdsverlopen maar als er wel bij de conclusie op in wordt gegaan levert een 3 op. Als in het geheel niet naar tijdverloop of keuzen wordt gekeken wordt een 2 gegeven.
19. Als de conclusies van het rapport zowel expliciet, duidelijk als ook beargumenteerd zijn wordt met een 5 gewaardeerd. Als één van deze punten mist wordt met een 4 gewaardeerd.

20. Als er expliciet en duidelijk een terugkoppeling is in de conclusie op de bijvoorbeeld het normenkader of andere maatstaven en in welke mate het beleid of proces daaraan voldoet wordt met een 5 gewaardeerd, als dit minder expliciet wordt gedaan of als het rapport minder duidelijke of heldere maatstaven heeft wordt met een 4 gewaardeerd.
21. Als er expliciet en helder wordt ingegaan op de aanbevelingen ten aanzien van het beleid of proces en ook in duidelijk wordt dat deze praktisch uitvoerbaar zijn wordt met een 5 gewaardeerd. Er wordt met een 4 gewaardeerd wanneer niet duidelijk wordt of de aanbevelingen meer idealistischer zijn of dat ze ook praktisch uitvoerbaar zijn.
22. Wanneer zowel een aanbeveling wordt gedaan als ook ingegaan wordt op de uitvoering daarvan wordt met een 5 gewaardeerd, een 4 wordt gegeven wanneer niet op de uitvoering wordt ingegaan en een 3 is gegeven wanneer niet helder is of er een werkelijke advies is gegeven of dat het meer algemene punten van kritiek zijn.
23. Als heel helder is dat uit de conclusie alle vragen zijn beantwoord die voortkwamen uit de opdracht voor het onderzoek wordt met een 5 gewaardeerd. Als in grote lijnen wel voldaan is aan de opdracht maar als bepaalde elementen bijvoorbeeld weggefallen of er bijgekomen zijn wordt een 4 gegeven.
24. Een heldere en duidelijke indeling met hoofdstukken en paragrafen levert een 5 op, een minder heldere maar wel voldoende een 4, een onvoldoende indeling, maar nog niet acuut storend wordt met een 3 gewaardeerd
25. Een duidelijke samenvatting en conclusies levert een 5 op, een minder duidelijke samenvatting levert een 4 op.
26. De rapporten die op een voldoende wijze geschreven zijn voor ook anderen dan wetenschapsbeoefenaars worden gewaardeerd met een 4, de rapporten die daar overheen gaan en speciale moeite doen in het bijvoorbeeld extra uitleggen van ingewikkeldere concepten worden gewaardeerd met een 5.
27. Vervalt.
28. De rapporten waarbij het zowel duidelijk is dat er een moment is waarop het rapport bruikbaar is voor het beleidsproces als ook waar het rapport op dat moment opgeleverd wordt worden gewaardeerd met een 5. De rapporten waar dit moment niet expliciet genoemd wordt, maar waar uit de context van het hele rapport wel blijkt dat het op een bruikbaar moment is opgeleverd worden gewaardeerd met een 4. Als er geen blijk is van enige tijdsmoment van bruikbaarheid

wordt een 3 gegeven, omdat ook niet duidelijk kan vastgesteld worden dat het rapport niet op een onbruikbaar moment is opgeleverd.

29. Vervalt.

30. Bij rapporten waar niet expliciet of impliciet wordt ingegaan op enige maatschappelijke gevolgen van het uitbrengen van het rapport wordt gewaardeerd met een 2.

4. Resultaten

4.1. *Kwalitatieve resultaten*

Wetenschappelijke eisen (A. ontbreekt hier omdat daar geen eisen voor zijn.)

B. Type onderzoek

Hierbij valt vooral op dat er in de meeste rapporten niet expliciet wordt ingegaan op de vraag of het onderzoek beschrijvend of verklarend is. Dit is echter meestal wel op te maken uit de context en de vraagstelling. Echt verklarend onderzoek komt overigens niet voor binnen de onderzochte rapporten. Bij de vraag of een rapport een beoordeling is van effecten moet genoemd worden dat aangezien er vrij weinig echt effectenonderzoek wordt uitgevoerd, hier het proces dat bij procesonderzoek wordt bestudeerd ook gezien wordt als een vorm van beleid. In die zin wordt er dan bij de meeste onderzoeken wel gekeken naar effecten, juist bij procesonderzoek ligt de focus vaak op de gevolgen die een proces hebben gehad op de totstandkoming of uitvoering van een beleid.

C. Doelstelling en probleemstelling

Alle onderzochte rapporten hebben een beschreven doelstelling. Bij de probleemstelling en de onderzoeksvragen waren er echter een tweetal rapporten waar dit niet naar behoren was uitgewerkt. Bij deze twee rapporten viel op dat de vorm en stijl van afweek van andere rapporten, er was duidelijk geen gebruik gemaakt van een bepaald "format".

D. Theorie en begrippen.

Geen van de onderzoeken zijn toetsend en dus worden er ook geen hypothesen getest. Deze criteria zijn dan ook niet meegenomen in de waardering. Verder zijn de centrale begrippen uit de probleemstelling worden meestal goed omschreven. Dit gebeurt veelal in de vorm van een normenkader wat opgesteld wordt aan de hand van de onderzoeksvragen. Soms worden zelfs specifiek de begrippen uitgelegd in een losse paragraaf.

E. Methoden van onderzoek

Eigenlijk bij alle rapporten is de methodische benadering goed of zeker ruim voldoende omschreven. De onderzoeksaanpak is in de meeste gevallen voldoende of goed verwoord, twee uitzonderingen daargelaten, deze rapporten maken gebruik van een heel andere opzet en volgen niet een vast "format".

De operationalisatie van de verschillende begrippen wordt ook bij de meeste rapporten goed gedaan. Er wordt vaak ingegaan op een definitie van de genoemde begrippen in de vraagstelling, waarna ook vaak verwezen wordt in de werkwijze hoe dit gemeten wordt, via vragenlijsten of interviews of casestudie. In het verlengde daarvan is er vaak sprake van een normenkader waaraan getoetst wordt, dit is in ook gelijk de operationalisatie van de meeste begrippen. Concreet wordt er met de onderzoeksvragen een normenkader opgesteld aan de hand van literatuur en de beleidspraktijk. Dit normenkader wordt bij het meeste procesonderzoek toegepast door casestudies aangevuld met interviews en vragenlijsten. Dit is tevens vaak de confrontatie van de normen/maatstaven met de empirie, zoals genoemd casestudies en interviews/vragenlijsten.

De geldigheid van het onderzoek wordt bij geen enkel rapport expliciet op ingegaan. Desondanks wordt aangenomen dat door de beantwoording van de onderzoeksvragen we kunnen aannemen dat de opstellers van het rapporten impliciet aangeven dat daarmee voldaan wordt aan de eisen van geldigheid. Immers, de vraag of er onderzocht is wat men wil onderzoeken is te zien als de onderzoeksvragen(wat men beoogde te onderzoeken) en de beantwoording daarvan(wat men onderzocht heeft. De betrouwbaarheid wordt verder helaas in geen enkel rapport op ingegaan. Het streven naar objectiviteit wordt in elk rapport wel in meerdere of mindere mate gevonden, veelal ligt dit gewaarborgd in interviews en wederhoor over de resultaten voor publicatie.

F. Technieken van Onderzoek.

De dataverzameling wordt meestal wel goed toegelicht. De meeste rapporten bevatten ook in meerdere of mindere mate de empirische gegevens, enkele uitzonderingen daargelaten met geen tot weinig gegevens.

De onderzoekstechnieken zijn bijna altijd wel redelijk goed beschreven.

G. Analyse van de data en conclusies

Bij de wijze van analyse van de data en de daaropvolgende interpretatie was het vaak lastig om daar binnen de rapporten een onderscheid in te vinden. Dit leidde bij de meeste rapporten tot hoge scores en slechts subtiele verschillen. De laagste scores zijn gegeven daar waar echt duidelijk slecht wordt ingegaan op de wijze van omgang met de data. Bijvoorbeeld wanneer er direct conclusies worden verbonden aan de data zonder dat er een vertaalslag wordt gemaakt, of dat men juist wel ingaat op de interpretatie maar dat niet helder is hoe men uit de ruwe data tot een geaggregeerde waardering komt.

Echter daar waar dit zo was leverde dit geen zo grote problemen op dat de leesbaarheid of begrijpelijkheid van het rapport in het geding kwam

De beantwoording van de probleemstelling en de onderzoeksvragen wordt voor alle rapporten goed gedaan. Enige uitzondering daarop met een beduidend lagere score waren de rapporten waar in eerste instantie bij de doelstelling en probleemstelling ook al steken waren gevallen. Toch voor zover er sprake was van een probleemstelling en vragen werd gepoogd hier een antwoord op te geven.

H. Opbouw

Ook hier over het geheel genomen alleen maar hoge scores, waarbij ook hier de uitzonderingen zaten in de rapporten waarbij al eerder dingen als doelstelling en probleemstelling niet goed uitgewerkt waren.

Over het algemeen lijkt het duidelijk dat een aantal standaard "formats" is gebruikt voor het opstellen van de rapporten.

I. Overige eisen

Continuïteit in de vorm van verwijzingen naar eerder verricht onderzoek vormt ook meestal geen probleem enkele uitzonderingen daargelaten. Er wordt vaak een in goed uitgeruste literatuurlijst voorzien en daar waar nodig worden voetnoten gebruikt. Ook hierbij wordt sterk de indruk gewekt dat vaak volgens een vast "format" wordt gewerkt, de lage scores zaten hier ook weer in de rapporten die afweken daarvan en algeheel wat rommeliger waren. De mate van originaliteit blijkt meestal uit het feit dat bij de beantwoording van de vragen en aanbevelingen voor de toekomst nieuwe kennis ten toon gespreid wordt die voor het onderzoek nog niet bekend was. Dit zijn echter slechts meestal aanbevelingen omtrent het proces dat bestudeerd is. Echte diepe toevoegingen aangaande theorie, methodologie of onderzoekstechnologie worden meestal niet gedaan. Dit ligt voornamelijk aan het feit dat meestal alleen simpel procesonderzoek wordt gedaan. Vandaar toch over het algemeen hoge scores met het voornaamste verschil in de wijze hoe de resultaten gepresenteerd worden en of daar impliciet wordt verwezen naar het feit dat het nieuw opgedane kennis is. De exactheid van de rapporten en de beheersing van de taal is altijd goed te noemen en er doen zich daar geen opmerkelijke dingen voor.

Eisen vanuit de beleidspraktijk (B. ontbreekt hier omdat daar geen eisen voor zijn)

A. Aanleiding en onderwerp van evaluatie

De meeste rapporten verwijzen voldoende naar de beleidspraktijk, het probleem en de opdracht uit de beleidspraktijk. Uitzonderingen die daar minder goed op scoren verwijzen in het algemeen niet naar een probleem maar stellen gewoon dat er iets onderzocht gaat worden. Het beleid waar het onderzoek betrekking op heeft en het onderwerp van de evaluatie zijn ook meestal goed verwoord, hier ook de uitzondering voor rapporten die een ander "format" gebruiken dan dat gebruikelijk blijkt te zijn.

Verwijzingen naar personen en/of instanties zijn meestal wel goed aanwezig.

C. Doelstelling en probleemstelling.

Zowel de verwijzing naar de doeleinden die verband houden met de beleidspraktijk alsook naar doeleinden van bepaalde personen of instanties zijn bij alle rapporten aanwezig. Dit is ook niet zo vreemd aangezien dit allemaal rekenkameronderzoek is wat in essentie altijd voor de gemeenteraad bedoeld is en in het algemeen natuurlijk beleid behandeld van het college. Omdat het meeste onderzoek procesonderzoek is, is het probleem niet vaak een concreet probleem waar het beleid betrekking op heeft maar meer het proces en het verloop daarvan. In dat opzicht proberen de meeste rapporten wel na te gaan of er is bijgedragen tot een oplossing. Echter zoals de vraag strikt bedoeld is, voldoen slechts enkele rapporten volledig. Dit zijn de rapporten die (soms naast procesonderzoek) concreet ingaan op de werkelijke effecten van een beleid in het beleidsveld.

D. Theorie en begrippen.

Ook hier komt weer om de hoek kijken dat de meeste onderzoeken procesonderzoek zijn. Toch wordt er vaak een normenkader gebruikt, waarbij teruggevallen wordt op theorie en begrippen.

E. Methoden van onderzoek

Veelal wordt gebruikt gemaakt van normenkaders om bij procesonderzoek te gebruiken. Verder wordt vaak goed ingegaan op de verschillende delen van het proces en hoe dit zijn weerslag had op het uiteindelijke resultaat en betrokken partijen. Daar waar in het rapport ook meer echt gekeken wordt naar effectiviteit worden de afzonderlijke beleidsinstrumenten geanalyseerd.

F. Technieken van onderzoek

Meestal wordt duidelijk genoemd welke bronnen in de vorm van literatuur en case studie zijn gebruikt, alsook interviews en vragenlijsten. In welk perspectief deze gegevens staan wordt minder vaak duidelijk aangemerkt. Vooral de rapporten die een duidelijk vast "format" gebruiken scoren het hoogste.

G. Analyse van de data en conclusies.

Bij procesonderzoek wordt primair gekeken naar hoe een beleid tot stand gekomen is of hoe de communicatie of de concrete uitvoering is gebeurd. Dit proces wordt ook als een vorm van "beleid" gezien voor de beantwoording van de vragen aangaande beleid. In dat licht wordt in de meeste rapporten ook wel aandacht besteed aan de neveneffecten van beleid. Vaak zijn dat dan de negatieve gevolgen die een verkeerd gelopen proces heeft gehad op het de beleidsuitvoering of op betrokken partijen. De concretere vraag of er een afweging wordt gemaakt tussen positieve en negatieve effecten en of daar een afweging in wordt gemaakt moet helaas veelal negatief beantwoord worden. Dit wordt bijna altijd onvoldoende gemaakt. Er is slechts één rapport waar een voldoende beoordeling wordt gegeven, en dat is omdat er concreet wordt gesproken van zowel negatieve alsook positieve effecten van het echte achterliggende beleid.

Ten aanzien van de tijdskeuzen van beleid wordt in procesonderzoek daar altijd wel in meerdere en mindere mate aandacht besteed. Onderdeel van een goed verlopen proces is immers ook dat de tijdskeuzen goed in elkaar haken en op elkaar aansluiten.

Alle conclusies van de rapporten waren voldoende expliciet, duidelijk en beargumenteerd. Alleen maar voldoende en goede scores. Via de aanname dat concrete aanbevelingen voor de toekomst een indicatie zijn dat beleid manipuleerbaar is voldoen de meeste rapporten aan de vraag of zij aangeven of de onderzochte "variabelen" manipuleerbaar zijn.

H. Opbouw.

Bijna alle rapporten hebben een duidelijke indeling en samenvatting alsmede conclusies. Ook hier weer de uitzondering de rapporten die afwijken van een vast "format". Vooral de duidelijke indeling had hieronder te leiden.

I. Overige eisen.

Bij alle rapporten blijkt dat ze redelijkerwijs goed te lezen zijn voor niet wetenschappers. Het tijdstip van oplevering en de daaruit volgende nuttigheid voor het beleidsproces wordt veelal niet expliciet op ingegaan. Dit moet afgeleid worden uit het rapport zelf, maar lijkt meestal wel voldoende te zijn. Alle rapporten, zonder uitzondering scoren onvoldoende op de vraag of het rapport een afweging geeft over de positieve en negatieve maatschappelijke gevolgen van het beschikbaar komen van de onderzoeksresultaten. Hier zou meer focus op moeten liggen. Enkele onderzochte rapporten hadden wel verregaande conclusies aangaande falen vanuit de gemeente alsook maatschappelijke organisaties. Dit zou tenminste leiden tot enige politieke onrust en niet ondenkbaar is dat belanghebbenden ook consequenties zullen ondervinden.

4.1.1. Conclusie

Het eerste punt wat opvalt na het bestuderen en waarderen van de rapporten is dat er veelal slechts procesonderzoek wordt gedaan. Strikt genomen wordt hier dan niet naar het beleid en de beleidseffecten gekeken maar enkel naar delen van het proces rondom het beleid. Dit is helaas niet altijd even zwart wit, vaak komen toch delen van het concrete beleid en de effecten voor het voetlicht. Voor de waardering van de eisen levert dit problemen op zodat begrippen als beleid en beleidseffecten soms wat ruimer opgevat moeten worden. Verder valt op dat veel rapporten volgens enkele vaste "formats" opgebouwd blijken te zijn. Mogelijk is dit niet verwonderlijk gezien het aantal rapporten waarbij de concrete onderzoeken uitbesteed zijn aan derde partijen zoals onderzoeksbureaus.

De hoge gemiddelde scores die hierna zullen blijken uit de statistische bewerking zijn waarschijnlijk het resultaat van relatief goed uitgevoerde rapporten op een lichte moeilijkheidsschaal, maar ook een meetinstrument in de vorm van de eisen van Hoogerwerf en Zoutendijk dat mogelijk niet goed toepasbaar is voor dit soort rapporten. De meest ingewikkelde eisen vallen weg omdat ze simpelweg bedoeld zijn voor echt effectenonderzoek, dus waar men ingaat op beleid, de beoogde effecten en de werkelijke effecten. Tevens zijn sommige eisen die eigenlijk bedoeld zijn voor effectenonderzoek wat breder opgevat om ook toepasbaar te zijn op procesonderzoek, om zo wel voldoende meetpunten over te houden. Echter door een eis minder scherp te hanteren als hij eigenlijk bedoeld was verliest hij wel zijn echte waarde als meetpunt.

4.2. Wetenschappelijke Eisen Vis-à-vis Eisen uit de beleidspraktijk

De indeling volgens Hoogerwerf en Zoutendijk (1990) van de twee sets eisen naast elkaar wordt het eerste punt waar we naar gaan kijken. Vanuit de scores voor de losse eisen worden geaggregeerde scores samengesteld voor de twee deelgebieden. Hierbij worden scores die ontbreken omdat ze niet relevant zijn, niet meegenomen voor het gemiddelde. Zie **bijlage 4** voor de uitgewerkte resultaten.

4.2.1. Standaard Statistische Verkenning

De standaard verkenning voordat er verdere statistische berekening wordt gedaan volgt hier:

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Wetenschappelijke eisen	24	100,0%	0	0,0%	24	100,0%
Eisen uit de Beleidspraktijk	24	100,0%	0	0,0%	24	100,0%

Zoals genoemd geen ontbrekende waarden.

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean
Wetenschappelijke eisen	24	1,30	3,35	4,65	4,2267
Eisen uit de Beleidspraktijk	24	1,02	3,57	4,59	4,1646
Valid N (listwise)	24				

Bij de algemene beschrijvingen vallen een aantal dingen op.

De gemiddelden zijn respectievelijk 4,2 en 4,2 op een schaal van 1 tot 5 voor de wetenschappelijk eisen en eisen uit de beleidspraktijk.

Ten tweede zijn de minimum en maximum scores ook relatief hoog. Dit betekent dan ook dat we in een smalle bandbreedte zitten aan het hoge eind van de schaal, met een spreiding van 1,3 en 1,0 op een schaal van 1 tot 5. Simpel gezegd, alle rapporten scoren vrij hoog en met relatief weinig onderling verschil.

Tests of Normality

	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Wetenschappelijke eisen	,243	24	,001	,785	24	,000
Eisen uit de Beleidspraktijk	,144	24	,200*	,963	24	,502

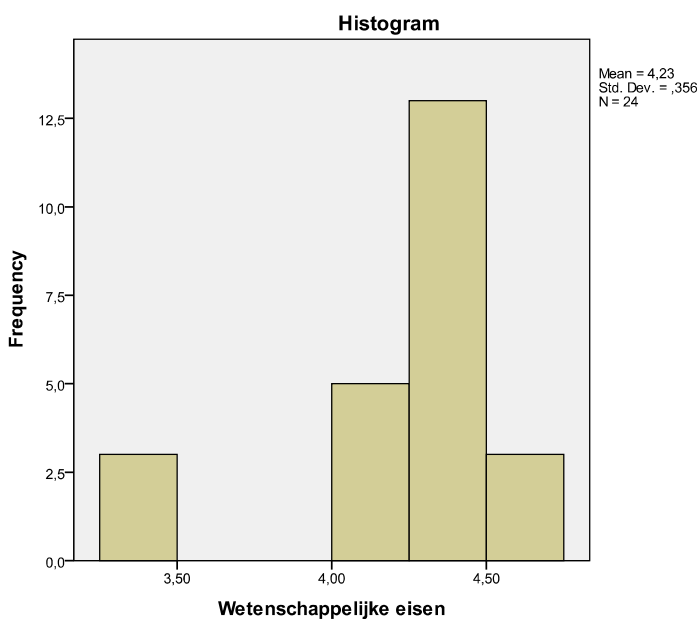
*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

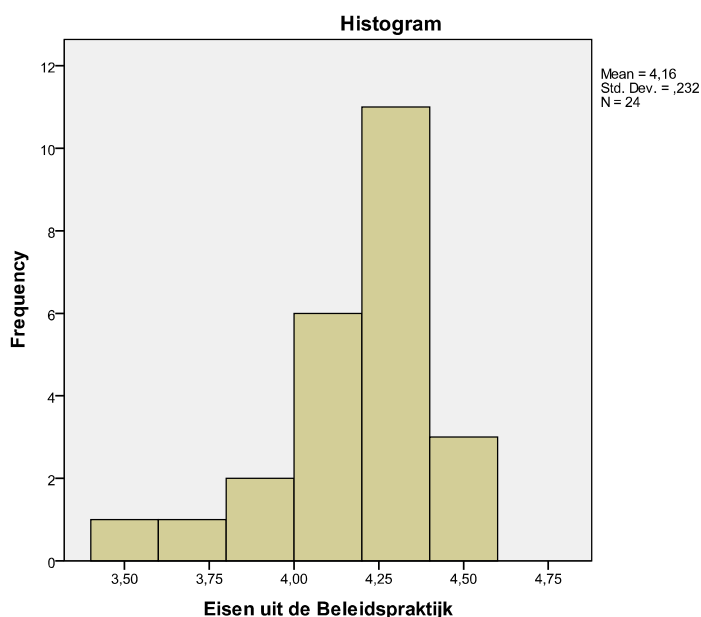
Bij de testen voor een normale verdeling zouden we in eerste instantie gezien de kleine N de Shapiro-Wilk moeten volgen. Hier komt voor de wetenschappelijke eisen een W van 0,76 uit met een p van 0,000. De W is lager dan 0,90 wat duidt op een niet normale verdeling en de significantie is ook kleiner dan 0,05 wat de niet normale verdeling onderschrijft. Ook de Kolmogorov-Smirnov, die over het algemeen alleen gebruikt wordt voor een veel grotere steekproef onderschrijft een niet normale verdeling met een van 0,001.

Bij de Eisen vanuit de beleidspraktijk volgt uit de Shapiro-Wilk een W van 0,96, dit is hoger dan 0,90 dus duidt het op een normale verdeling. Ook de significantie is met een p van 0,50 hoger dan 0,05 wat de conclusie van een normale verdeling ondersteunt. De Kolmogorov-Smirnov bevestigt hier ook een normale verdeling met een significantie hoger dan 0,05, namelijk $p=0,20$

Als we verder kijken naar het histogram, dan zien we waarom de wetenschappelijke eisen niet normaal verdeeld zijn:



Er zijn blijkbaar enkele lage afwijkende scores, dit strookt ook met de scores voor de rapporten, er waren enkele rapporten bij die opvallend slecht scoorden voor de wetenschappelijke eisen.



Het histogram voor de eisen uit de beleidspraktijk geeft inderdaad een mooier beeld weer, richting een normale verdeling.

Ondanks de niet normale verdeling bij de wetenschappelijke eisen lijkt het toch zinnig om door te gaan met een Independent Samples T-Test. Het is niet praktisch mogelijk om de afwijkend rapporten te schrappen of andere rapporten te zoeken vanwege de kleine dataset, want hiervoor zou ook een selectie van andere gemeentes nodig zijn, wat lastig is vanwege het gematchde karakter van de gemeentes. Het kunnen testen van de gepaarde gemeentes onderling wordt hier als belang hoger geacht dan een strikt correcte normale verdeling. Hierdoor blijven alle 24 rapporten zoals ze nu zijn in de selectie en gaan we door met de T-Test.

4.2.2. De onafhankelijke steekproef T-Test

Bij een onafhankelijke steekproef T-Test worden gekeken of er een effect is tussen de onafhankelijke en de afhankelijke variabelen, de test geef ook weer of dit effect statistisch significant is (dus of er wetenschappelijk te spreken valt van een significant verschil)

Concreet zullen we hier de volgende hypothesen testen:

De H_0 is dat de scores voor de rapporten van rekenkamers met wetenschappers hetzelfde zijn als de scores van de rapporten van rekenkamers zonder wetenschappers.

De H_1 is dat de scores voor de rapporten van rekenkamers met wetenschappers hoger zijn dan de scores van de rapporten van rekenkamers zonder wetenschappers.

H_0 : Score voor rapporten met wetenschappers = Score voor rapporten zonder wetenschappers

H_1 : Score voor rapporten met wetenschappers > Score voor rapporten zonder wetenschappers

Dit wordt nu eerst gedaan voor de twee sets eisen in z'n totaliteit voor alle rekenkamers.

Group Statistics

		N	Mean	Std. Deviation	Std. Error Mean
Wetenschappelijke eisen	Wel Wetenschappers	12	4,3392	,17906	,05169
	Geen Wetenschappers	12	4,1142	,45304	,13078
Eisen uit de Beleidspraktijk	Wel Wetenschappers	12	4,2183	,12157	,03509
	Geen Wetenschappers	12	4,1108	,30312	,08750

Uit de groepstatistieken valt af te lezen dat er een verschil te zien is. De gemiddelde scores bij de wetenschappelijke eisen zijn 4,34 voor de rapporten van rekenkamers met wetenschappers en 4,11 voor de rapporten van rekenkamers zonder wetenschappers. Dit verschil is ook nog eens in de richting die verwacht werd. Voor de eisen vanuit de beleidspraktijk zijn de gemiddelde scores respectievelijk 4,22 en 4,11. Hier is de richting wederom zoals verwacht was.

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Wetenschappelijke eisen	Equal variances assumed	8,211	,009	1,600	22	,124	,22500	,14063	-,06664	,51664
	Equal variances not assumed			1,600	14,355	,131	,22500	,14063	-,07591	,52591
Eisen uit de Beleidspraktijk	Equal variances assumed	7,146	,014	1,140	22	,266	,10750	,09428	-,08802	,30302
	Equal variances not assumed			1,140	14,449	,273	,10750	,09428	-,09412	,30912

Bij de uitslag van de T-Test zien we echter dat de uitkomst voor beide sets eisen niet significant is. Bij beiden geen gelijke varianties, met voor de wetenschappelijke eisen een tweezijdige significantie van 0,13 wat voor een enkelzijdige test nog steeds 0,07 oplevert wat bij een $\alpha=0,05$ niet significant is. Dus de H_0 kan niet verworpen worden. Voor de eisen uit de beleidspraktijk volgt dubbelzijdige significantie van 0,27 wat wederom bij een enkelzijdige test nog slechts 0,14 oplevert wat ook bij een $\alpha=0,05$ geen significantie oplevert. De H_0 kan hier ook niet verworpen worden.

Conclusie is dus bij beiden sets eisen dat er geen statistisch significant effect optreedt, ondanks dat het bij de wetenschappelijk eisen bij een enkelzijdige test wel heel dichtbij komt.

4.3. De sets gemeentes

Dan rest nog resultaten voor de testen van de sets van gemeentes. In de eerste plaats worden de paren gemeentes onderling vergeleken. Dit wordt gedaan om zo te kijken of bij een bepaalde set gemeentes,

die zoals in bij de case-selectie genoemd, gematched zijn op grootte en locatie, wel een significant verschil optreedt.

Ten tweede zullen de variabele grootte van de gemeente direct testen tegen de kwaliteit van de rapporten. Dit om zo te vast te stellen of er een verband is tussen de grootte van de gemeentes en de kwaliteit van de rapporten. Hierbij zal dus de variabele aanwezigheid van wetenschappers niet meegenomen worden.

4.3.1. Gematchde paren

Apeldoorn en Deventer

Hier is het gemiddelde voor de wetenschappelijke eisen voor de rekenkamer met wetenschappers(Apeldoorn) 4,3 en zonder wetenschappers(Deventer) 4,3. Voor de eisen vanuit de beleidspraktijk zijn de gemiddelden respectievelijk 4,2 en 4,1. Praktisch geen verschil dus. Dit wordt ook bevestigd door de uitkomst van de T-test, met een tweezijdige significantie van 0,95 voor de wetenschappelijke eisen 0,92 voor de eisen uit de beleidspraktijk

Den Bosch en Venlo

Hier is het gemiddelde voor de wetenschappelijke eisen voor de rekenkamer met wetenschappers(Den Bosch) 4,3 en zonder wetenschappers(Venlo) 4,1. Voor de eisen vanuit de beleidspraktijk zijn de gemiddelden respectievelijk 4,3 en 4,3. Ook hier weinig verschil dus. Dit wordt ook bevestigd door de uitkomst van de T-test, met een tweezijdige significantie van 0,67 voor de wetenschappelijke eisen en 0,78 voor de eisen uit de beleidspraktijk

Hardenberg en Wijk bij Duurstede

Hier is het gemiddelde voor de wetenschappelijke eisen voor de rekenkamer met wetenschappers(Hardenberg) 4,4 en zonder wetenschappers(Wijk bij Duurstede) 3,7. Voor de eisen vanuit de beleidspraktijk zijn de gemiddelden respectievelijk 4,2 en 3,8. Hier duidelijk wel verschil. Dit wordt ook bevestigd door de uitkomst van de T-test, met een tweezijdige significantie van 0,18 voor de wetenschappelijke eisen en 0,04 voor de eisen uit de beleidspraktijk. Voor de eisen uit de beleidspraktijk is er dus een significant verschil op $\alpha=0,05$ met een p van 0,04.

Veenendaal en Ede

Hier is het gemiddelde voor de wetenschappelijke eisen voor de rekenkamer met wetenschappers(Veenendaal) 4,3 en zonder wetenschappers(Ede) 4,3. Voor de eisen vanuit de beleidspraktijk zijn de gemiddelden respectievelijk 4,2 en 4,2. Door afronding met significante getallen hier zelfs helemaal geen verschil. Dit wordt ook weer bevestigd door de uitkomst van de T-test, met een tweezijdige significantie van 0,47 voor de wetenschappelijke eisen en 0,57 voor de eisen uit de beleidspraktijk

De complete statistische uitkomsten zijn bijgevoegd in **bijlage 5**.

4.3.2. Grootte van de gemeenten

Voor de test op grootte van de gemeenten is een grens gesteld op 100.000, alle gemeenten met inwonersaantallen boven de 100.00 zijn gegroepeerd onder groot en alle gemeenten met minder dan 100.00 inwoners zijn gegroepeerd onder klein. In de praktijk betekent dit dat Apeldoorn, Den Bosch, Venlo en Ede vallen onder de grote gemeenten en Deventer, Hardenberg, Wijk bij Duurstede en Veenendaal vallen onder de kleine.

Group Statistics

	Groote	N	Mean	Std. Deviation	Std. Error Mean
Wetenschappelijke eisen	Groot	12	4,2725	,30374	,08768
	Klein	12	4,1808	,40998	,11835
Eisen uit de Beleidspraktijk	Groot	12	4,2533	,16008	,04621
	Klein	12	4,0758	,26487	,07646

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Wetenschappelijke eisen	Equal variances assumed	1,486	,236	,622	22	,540	,09167	,14729	-,21380	,39713
	Equal variances not assumed			,622	20,280	,541	,09167	,14729	-,21531	,39864
Eisen uit de Beleidspraktijk	Equal variances assumed	2,664	,117	1,987	22	,060	,17750	,08934	-,00778	,36278
	Equal variances not assumed			1,987	18,090	,062	,17750	,08934	-,01013	,36513

Voor de wetenschappelijke eisen is er een gemiddelde voor de grote gemeenten van 4,3 en voor de kleine 4,2. Dit blijkt geen significant verschil te zijn met een tweezijdige p van 0,54

Voor de eisen uit de beleidspraktijk is er een gemiddelde voor de grote gemeente van 4,3 en voor de kleine 4,1. Dit blijkt wel een significant verschil te zijn met een enkelzijdig p van 0,03.

De conclusie is dus dat voor de eisen uit de beleidspraktijk er een significant verschil is voor de scores van de rapporten van de grote gemeentes en de kleine gemeentes. Daarbij is dit verschil dat grote gemeentes significant beter scoren.

5. Conclusie

De aanwezigheid van lokale rekenkamers wordt gezien als een positief middel om beleid op lokale schaal beter en onafhankelijker te evalueren, door substantieel onderzoek te doen naar de doeltreffendheid, doelmatigheid en rechtmatigheid van beleid. Als de aanwezigheid van wetenschappers in de rekenkamers een verbetering van het gedane onderzoek en een betere bruikbaarheid in de beleidspraktijk zou opleveren zou dat maatschappelijk gezien relevant zijn, zo wordt de functie van de rekenkamer ten volste benut.

5.1. *Beantwoording Onderzoeksvraag*

Hoofdvraag:

Er wordt niet significant meer voldaan aan de wetenschappelijke eisen voor beleidsevaluaties in rekenkameronderzoeken van rekenkamers waar wetenschappers met een universitaire aanstelling in zitten.

Deelvragen:

- Er wordt niet significant meer voldaan aan de eisen uit de beleidspraktijk voor beleidsevaluaties in rekenkameronderzoeken van rekenkamers waar wetenschappers met een universitaire aanstelling in zitten.
- Er is geen significante invloed van de aanwezigheid van wetenschappers met een universitaire aanstelling in lokale rekenkamers op de mate waaraan in rekenkameronderzoeken voldaan wordt aan de wetenschappelijke eisen en de eisen uit de beleidspraktijk voor beleidsevaluaties.

5.2. *Oorzaken en gevolgtrekkingen*

Bij het doen van dit onderzoek kwamen nogal wat aanwijzingen naar boven die richting een redenering zouden kunnen wijzen waarom er geen effect is. Ten eerste, zoals genoemd bij de statistische resultaten, zijn alle rapporten vrij hoog gewaardeerd en is de totale spreiding ook klein. Een meetinstrument waarbij alle resultaten relatief hoog uitkomen en er weinig spreiding is wijst op een mogelijk onbetrouwbaar of zelfs niet valide meetinstrument. Als de volle spreiding van het instrument niet gebruikt wordt zou het heel goed niet toepasbaar kunnen zijn voor wat er concreet geprobeerd wordt te

meten. In dit geval zou de set eisen van Hoogerwerf en Zoutendijk (1990) mogelijk niet toepasselijk kunnen zijn.

Mogelijke redenen van deze slechte fit is de grote hoeveelheid procesonderzoek. Dit onderzoek is zoals genoemd een simpele soort onderzoek en gaat eigenlijk enkel en alleen het verloop van een proces na en doet ten hoogste aanbevelingen voor een verbetering in de toekomst. Door dit relatief simpele onderzoek scoren de meeste rapporten vrij hoog. Dit ligt ten dele aan het feit dat het eisenkader van Hoogerwerf en Zoutendijk niet speciaal is toegespitst op dit type onderzoek, maar meer uitgaat van effectiviteitsonderzoek. Hierdoor vervallen sommige van de zwaardere eisen omdat ze niet van toepassing zijn of zijn sommige eisen vrijer geïnterpreteerd dan strikt de bedoeling was van Hoogerwerf en Zoutendijk, dit dan om toch een werkbare hoeveelheid relevante eisen over te houden. Ten tweede is van belang om op te merken dat procesonderzoek zoals gezegd een vrij simpele vorm van onderzoek is die daarbij ook goed uitgekristalliseerd is en mits redelijk stringent uitgevoerd, makkelijk goede scores kan halen, zeker op het gebied van de wetenschappelijke eisen. Dit is ook te zien in het feit dat veel van dit procesonderzoek wordt uitbesteed aan externe onderzoeksbureaus en dan opvalt dat er onder hen slechts een beperkt aantal "formats" schijnt te wezen in de opbouw van de rapporten. Hierdoor lijkt het een beetje standaardwerk te worden, waar dus weinig mee mis kan gaan.

Dit blijkt maar des te meer uit het feit dat de het enige significante verschil opdook bij de rekenkamer van Wijk bij Duurstede die blijkens hun onderzoeken veelal intern werken. Hierdoor bleken er toch wel enkele dingen in het rapport anders, dit leverde een lagere waardering op, ten dele omdat sommige delen gewoon minder goed waren uitgewerkt en ook omdat de wijze van latent coderen een manier van waarderen had opgeleverd waarbij een afwijkend rapport mogelijk al gauw een lagere score kreeg.

Interessant is verder te zien dat de gemeente Wijk bij Duurstede veruit de kleinste gemeente is die is geanalyseerd, het zou zomaar kunnen dat kleinere gemeentes minder budget hebben voor uitbesteding van lokaal rekenkameronderzoek en daardoor meer intern werken. Dit wordt ook gestaafd door het testen op basis van de grootte van de gemeentes, waar uitkwam dat de grootte van de gemeente wel degelijk een positieve invloed heeft op de eisen uit de beleidspraktijk.

Concluderend kan gezegd worden dat het instrument in de vorm van het kader van eisen van Hoogerwerf en Zoutendijk niet geschikt is voor dit onderzoek, óf dat simpelweg de rapporten gewoon allemaal hoog scoorden omdat ze gewoon goed zijn. Beide punten zijn waarschijnlijk ten dele waar. Het gros van de rapporten scoort inderdaad heel hoog omdat het gewoon goed uitgevoerde procesonderzoeken zijn. Echter deze hoge score komt uiteraard tot stand via het eisen kader, dat, zoals

genoemd, vooral bedoeld was om ingewikkelder effectiviteitsonderzoek te analyseren. Hierover bij de aanbevelingen meer

Naast de grootte van de gemeente is ook nog steeds te stellen dat de in de theorie genoemd variabelen; type onderzoek en wijze van uitvoering, een veel grotere invloed hebben op de kwaliteit van de rapporten dan de aanwezigheid van Wetenschappers in de rekenkamer. Hierbij val te denken dat het meten van de positieve invloed van wetenschappers misschien beter gedaan kan worden door te kijken naar de mate van "ingewikkeld" onderzoek dat uitgevoerd wordt door de rekenkamer. De set eisen die Hoogerwerf en Zoutendijk (1990) opgesteld hebben lijken meer van toepassing op onderzoek waarbij een hogere mate van complexiteit is.

Door de tussenkomst van externe bureaus lijkt het moeilijk de invloed vast te stellen tussen de rekenkamer en de kwaliteit van het geleverde onderzoek. De stelling dat de aanwezigheid van wetenschappers invloed heeft op de aanbesteding van het onderzoek bij de externe bureaus en daarmee ook een verschil in kwaliteit bewerkstelligd is vanwege de vele simpele procesonderzoek niet vast te stellen. Bij echt doelmatigheidsonderzoek zou nog wel eens verschil op kunnen treden in de competentie van de verschillende externe bureaus.

Saillant detail is verder dat in veel onderzoeken die in de praktijk slechts procesonderzoek zijn, toch wordt gehint of soms zelfs expliciet gemeld dat er onderzoek gedaan wordt naar de doelmatigheid en doeltreffendheid van beleid. Dit is niet verwonderlijk omdat dit de doelstelling zou moeten zijn van lokale rekenkameronderzoek. Echter dat hiermee vaak de plank compleet wordt misgeslagen geeft aan hoe makkelijk wordt omgegaan met de distincties van de soorten onderzoek en is het dus ook niet verwonderlijk dat vaak het simpele pad van een procesonderzoek wordt gekozen. Dat hier vaak ook nog andere factoren een rol spelen zoals tijd voor uitvoering en budgetten laten we even buiten beschouwing, echter idealiter zouden wetenschappers meer effectenonderzoek moeten (laten) doen. Het instrument van de rekenkamer wordt anders bot gemaakt door de vele bijna "ambtelijk" simpele procesonderzoeken die soms niet meer is dan een financiële rapportage over het lokale onrendabele zwembad.

5.3. *Aanbevelingen voor vervolgonderzoek*

Het "probleem" waar het meest prominent tegenaan werd gelopen in dit onderzoek is de relatief kleine dataset aan onderzoeken. Om de selectiecriteria die volgen uit de theorie, zoals type onderzoek en wijze van uitvoering in combinatie met de grootte van de gemeente goed te kunnen verwerken in de case selectie, is een grotere dataset vereist. Daarbij wordt relatief veel procesonderzoek uitgevoerd en dan

ook vaak door externe bureaus. Als we de eisen voor kwalitatief onderzoek van Hoogerwerf en Zoutendijk (1990) willen gebruiken, is het noodzakelijk om vooral ingewikkelder onderzoek te hebben waarbij de ook de effecten van beleid worden meegenomen. Hierdoor zou de focus dan wel meer komen te liggen op het type rapport, waarbij het gericht vergelijken van ingewikkeldere rapporten het risico met zich meeneemt dat alleen rekenkamers in de selectie meegenomen worden die deze rapporten ook publiceren.

Als het niet mogelijk blijkt om een grotere data-set te creëren waarbij dus ook voldoende ingewikkeldere onderzoeken aanwezig zijn, bijvoorbeeld omdat ze gewoon niet voldoende voorhanden zijn, is een andere onderzoeksopzet misschien een oplossing. Hierbij zou dan de mate van ingewikkeldere onderzoeken die uitgevoerd worden een indicator kunnen zijn van een vorm van kwaliteit van onderzoek. Als zou blijken dat de rekenkamers met wetenschappers significant meer ingewikkelde onderzoeken (laten) doen dan rekenkamers zonder wetenschappers zou dat misschien een handvat bieden voor verder onderzoek.

5.4. *Eindconclusie*

Het effect zoals vanuit de theorie verwacht wordt treedt in het algemeen niet op. De reden daarvoor zou kunnen zijn het vrij simpele karakter van het onderzoek zoals dit meestal wordt gedaan bij de lokale rekenkamers. De mogelijke interveniërende variabele grootte van de gemeente blijkt wel een effect te hebben, zij het dan alleen ten aanzien van de eisen uit de beleidspraktijk. De variabelen type onderzoek en wijze van uitvoering, hoewel niet direct getest in dit onderzoek, doen aan de hand van de conclusies van hetgeen wel onderzocht is, vermoeden dat zij een mate van invloed hebben op de kwaliteit van de rapporten. Naar alle waarschijnlijkheid is de reden dat bijna alle rapporten consistent hoog scoren, omdat zij gaan over simpel procesonderzoek wat door externe bureaus is uitgevoerd die veelal volgens vaste "formats" werken.

Een laatste conclusie en algemene aanbeveling is dat het zinvol lijkt voor lokale rekenkamers om meer onderzoek te doen naar doelmatigheid en doeltreffendheid en dat dit niet alleen in de kantlijn van het rapport genoemd wordt zonder dat dit type onderzoek werkelijk wordt gedaan.

Literatuurlijst

- Babbie, E. (2001). *The practice of social research*: Wadsworth/Thomson Learning.
- Bressers, J. T. A., & Hoogerwerf, A. (1995). *Beleidsvaluatie*. Alphen aan de Rijn: Samsom HD Tjeenk Willink.
- Bressers, J. T. A., Klok, P. J., Kuks, S. M. M., & Lulofs, K. R. D. (1988). *Handleiding instrumententheorie*: Ministerie van Volkshuisvesting, Ruimtelijke Ordening en Milieubeheer : DOP [distr.].
- Dunn, W. N. (2008). *Public policy analysis: an introduction*: Pearson Prentice Hall.
- Hoogerwerf, A. (1986). *Vanaf de top gezien: visies van de politieke elite*: Sijthoff.
- Hoogerwerf, A., m.m.v. H.N.J. Smits (Ed.). (1988). *Departementale beleidsvaluaties*. 's-Gravenhage.
- Hoogerwerf, A., & Zoutendijk, D. C. (1990). *Naar een evaluatiesysteem voor overheidsbeleid*.
- Patton, M. Q. (1978). Utilization-focussed Evaluation.
- Pröpper, I. M. A. M. (1987). Beleidsvaluatie als argumentatie: Beleidswetenschap.
- Pröpper, I. M. A. M. (1989a). Argumentatie en machtsuitoefening in onderzoek en beleid. *Evaluatie-onderzoek naar de Wet Investeringsrekening en het gebruik ervan in het beleidsproces*.
- Pröpper, I. M. A. M. (1989b). Het Gebruik van Beleidsvaluatie: Terugkoppeling. *HOOGERWERF, A.(red.), Overheidsbeleid. Een Inleiding in de Beleidswetenschap. Alphen aan den Rijn: Samsom Uitgeverij, 185-203.*
- Radin, B. A. (2000). *Beyond Machiavelli: Policy analysis comes of age*: Georgetown Univ Pr.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*: Houghton Mifflin College Div.

Snijders, P., Zantinge, A. (1989). *Een omgevallen boekenkast: Inventarisatie van beleidsevaluaties binnen de sociale zekerheid*. Enschede.

van de Graaf, H., & Hoppe, R. (1996). *Beleid en politiek. Een inleiding tot de beleidswetenschap en de beleidskunde*.

Bijlagen

Bijlage 1. Overzicht rapporten.

Rekenkamer	Weten-Schappers	Naam Rapport	Nummer rapport
Apeldoorn	1	Tussen ambitie en realiteit	1
Apeldoorn	1	Touwtjes aan elkaar knopen	2
Apeldoorn	1	Collegeonderzoek in de spiegel- Gemeente Apeldoorn	3
Deventer	0	Gebiedsontwikkeling in de mist	4
Deventer	0	Jong-leren met de jeugd (smeden aan een betere zorgketen)	5
Deventer	0	Wonen boven Winkels Deventer	6
Den Bosch	2	Onderzoek Evaluatie Stedelijke Museum 's-Hertogenbosch (SM's)	7
Den Bosch	2	Onderzoek afvalkosten gemeente 's-Hertogenbosch	8
Den Bosch	2	Leefbaarheid: de volgende stap?	9
Venlo	0	Onderzoek grote projecten Venlo	10
Venlo	0	Voortbouwen op solide basis, zelfevaluatie	11
Venlo	0	Onderzoek voortgang collegeprogramma 2006-2010	12
Hardenberg	1	Rioolrechten	13
Hardenberg	1	Onderzoek onderzocht - Doe mee onderzoek	14
Hardenberg	1	een rechter en een linkerschoen - Een onderzoek naar de sturingsrelaties	15
Wijk bij Duurstede	0	Op de valreep geprivatiseerd: met enige fortuin gescoord	16
Wijk bij Duurstede	0	Op Hoop Van Zegen, rapportage onderzoek herinrichting Lekhaven	17
Wijk bij Duurstede	0	Juist samenwerken verdient een eigen kader!	18
Veenendaal	1	Van rapport naar resultaat	19
Veenendaal	1	Bouwsteen voor de toekomst	20
Veenendaal	1	brouwerspoort getoetst	21
Ede	0	Besluitwaardig, informatievoorziening grote projecten	22
Ede	0	Kunst en Vliegwerk	23
Ede	0	Verbind en Beheers - De gemeente Ede en haar verbonden partijen	24

Bijlage 2. Wetenschappelijke eisen ten aanzien van een evaluatierapport over beleidseffecten.(Hoogerwerf & Zoutendijk, 1990, pp. 19-21)(Aangepast)

A. Aanleiding en onderwerp van evaluatie

Voor deze categorie zijn geen specifieke wetenschappelijke eisen geformuleerd.

B. Type onderzoek

1. In het rapport is vermeld of het onderzoek beschrijvend dan wel verklarend is.
2. Indien het onderzoek verklarend is, moet in het rapport zijn vermeld of het onderzoek toetsend dan wel explorerend is, dat wil zeggen of gepoogd wordt hypothesen te falsifiëren dan wel te ontwikkelen.
3. Het rapport is een beoordeling van effecten van een beleid, dat wil zeggen: de beoordeling richt zich op de doelbereiking, effectiviteit, neveneffecten en/of de kosten en baten van een beleid.

C. Doelstelling en probleemstelling

4. Het rapport bevat een doelstelling.
5. Het rapport heeft een probleemstelling (eventueel aangevuld met onderzoeksvragen) in de vorm van een of meer vragen waarop in het evaluatieonderzoek een antwoord wordt gezocht.

D. Theorie en begrippen

6. Indien het onderzoek toetsend is, moet in het rapport worden vermeld welke theorie (hetzij uit de beleidspraktijk en/of uit een wetenschappelijke publicatie) wordt getoetst. Indien het onderzoek explorerend is, moet in het rapport worden vermeld welke theorie is ontwikkeld.
7. De afzonderlijke veronderstellingen of hypothesen die worden getoetst of zijn ontwikkeld, zijn in het rapport omschreven. Aangegeven is of het om causale verbanden gaat.
8. De centrale begrippen uit de probleemstelling van het onderzoek zijn in het rapport omschreven.

E. Methoden van onderzoek

9. De gekozen methodische benadering is in het rapport omschreven.
10. De wijze van operationalisatie van de begrippen (d.w.z. een omschrijving aan de hand waarvan de aanwezigheid empirisch kan worden vastgesteld) is in het rapport omschreven.
11. Het rapport bevat één of meer maatstaven (doeleinden, waarden of normen).
12. Het rapport bevat beoordeling, dat wil zeggen een confrontatie van de te beoordelen empirische gegeven met de maatstaven.

13. De geldigheid van het onderzoek (dat wil zeggen: de vraag of is onderzocht wat men beoogt te onderzoeken) is in het rapport genoemd.
14. De betrouwbaarheid van het onderzoek (dat wil zeggen de vraag of een herhaling van het onderzoek onder gelijke omstandigheden hetzelfde resultaat zal opleveren) is in het rapport genoemd.
15. Het rapport streeft naar objectiviteit, dat wil zeggen naar het onderscheiden van feiten en oordelen en naar een onbevooroordeelde en nauwkeurige weergaven van feiten.

F. Technieken van onderzoek

16. De wijze van verzameling van de data (gegevens) is in het rapport omschreven.
17. Het rapport bevat vermelde empirische gegevens.
18. De afzonderlijke onderzoekstechnieken (voor de verzameling en analyse van de data) zijn in het rapport omschreven.

G. Analyse van de data en conclusies

19. De wijze waarop de data zijn geanalyseerd (dat wil zeggen volgens bepaalde gezichtspunten onderscheiden en ontleed) is in het rapport omschreven.
20. De wijze waarop de geanalyseerde data worden geïnterpreteerd is in het rapport omschreven.
21. De vragen uit de probleemstelling en de onderzoeksvragen worden in het rapport beantwoord.

H. Opbouw

22. Het rapport heeft een duidelijke en beargumenteerde systematiek, dat wil zeggen een opbouw van het geheel en de delen.

I. Overige eisen

23. Het rapport geef blijk van continuïteit, dat wil zeggen het bouwt -onder meer blijkens voetnoten en/of literatuurlijst- voort op reeds eerder verricht onderzoek met een vergelijkbare theoretische, methodische en/of onderzoekstechnische benadering.
24. Het rapport geeft blijk van originaliteit, dat wil zeggen wordt in het rapport uiteengezet wat het onderzoek in theoretisch, methodisch en/of onderzoekstechnisch opzicht aan de bestaande kennis toevoegt dan wel hoe bestaande kennis op een nieuwe wijze wordt geordend.
25. Het rapport is exact, dat wil zeggen precies, niet alleen in het gebruik van getallen, maar ook in het gebruik van woorden (omschrijvingen, citaten, verwijzingen)
26. Het rapport geeft blijk van een goede beheersing van de taal waarin het gesteld is.

Bijlage 3. Eisen vanuit de beleidspraktijk ten aanzien van een rapport over beleidseffecten

(Hoogerwerf & Zoutendijk, 1990, pp. 28-31)(Aangepast)

A. Aanleiding en onderwerp van evaluatie

1. Het rapport verwijst naar een in de beleidspraktijk bestaande behoefte aan informatie waarin het onderzoek kan voorzien
2. Het rapport verwijst duidelijk naar een probleem uit de beleidspraktijk (hetzij het beleidsveld of het openbaar bestuur), zijn oorzaken en gevolgen.
3. Het rapport verwijst naar een eventuele opdracht vanuit de beleidspraktijk tot het verrichten van het onderzoek.
4. Het rapport verwijst naar het beleid of beleidsproject waarop de evaluatie betrekking heeft.
5. Het rapport verwijst naar het onderwerp van de evaluatie. Dit onderwerp kan in het algemeen gesproken betrekking hebben op het intern of extern gerichte beleid, de beleidstheorie, het beleidsproces (en zijn onderdelen: agendavorming, voorbereiding, bepaling, uitvoering, evaluatie, terugkoppeling, beleidsbeëindiging), de beleidseffecten, de beleidsorganisatie en het beleidsveld.
6. Het rapport verwijst naar contacten met (personen van) de opdrachtgevende instantie(s).
7. Het rapport verwijst naar beleidsinstantie(s) die direct of indirect bij het geëvalueerde beleid zijn betrokken en die van de onderzoeksresultaten kunnen profiteren.

B. Type onderzoek

Voor deze categorie leken geen specifieke eisen vanuit de beleidspraktijk naast de eerder vermelde wetenschappelijke eisen ten aanzien van het type onderzoek nodig te zijn. Men kan zeggen dat de beleidspraktijk voor belangstelling heeft voor toegepast onderzoek, dat wil zeggen voor onderzoek waarbij men een bepaalde praktische toepassing voor het oog heeft. Maar ook fundamenteel onderzoek, waarbij met geen praktische toepassing beoogt, blijkt voor de beleidspraktijk relevant te kunnen zijn. Iets dergelijks geldt ook voor onderzoek dat niet in opdracht van een beleidsinstantie wordt verricht.

C. Doelstelling en probleemstelling

8. Het rapport verwijst naar een of meer doeleinden van het onderzoek, die verband houden met de beleidspraktijk.
9. Het rapport verwijst naar een of meer bepaalde beleidsdoeleinden van bepaalde personen of organisaties en gaan na of deze -al dan niet dan zij het beleid- zijn bereikt of kunnen worden bereikt.
10. Het rapport gaat na wat het beleid tot de oplossing van het probleem heeft bijgedragen of kan bijdragen.

D. Theorie en begrippen

11. Het rapport geef een toetsing van een of meer veronderstelling (de beleidstheorie) uit de beleidspraktijk waar op het beleid berust.

E. Methoden van onderzoek

12. Het rapport geen een beoordeling aan de hand van een of meer vermelde maatstaven (beginselen, normen of doeleinden) uit de beleidspraktijk
13. Bij outputbudgettering van gesubsidieerde particuliere organisaties worden ook de beleidsprestaties (primaire effecten, toegepaste beleidsmiddelen) van de gesubsidieerde organisaties en hun resultaten geëvalueerd.
14. Het rapport verwijst, indien het effectiviteitonderzoek betreft, naar één of meer bepaalde beleidsinstrumenten (middelen) en gaat na in hoeverre deze hebben bijgedragen of kunnen bijdragen tot het bereiken van het doel waarvoor zij worden ingezet.

F. Technieken van onderzoek

15. Het rapport verwijst naar de bronnen waaruit de gegevens voor het onderzoek (en met name voor de beoordeling van het beleid) zijn geput. Het rapport verwijst tevens naar het perspectief van het onderzoek: dat van de betrokken beleidsvoorbereiders, beleidsbepalers of uitvoerders dan wel anderen zoals burgers ("cliënten") belangengroepen, adviesorganen, politieke partijen media etcetera. (Snijders, 1989, pp. 70-72)

G. Analyse van de data en conclusies

16. Het rapport besteedt aandacht aan neveneffecten (niet beoogde effecten) van het beleid.
17. Het rapport maakt ten aanzien van de effecten van het beleid onderscheid tussen positief gewaardeerde effecten (baten of voordelen) en negatief gewaardeerde effecten (kosten of nadelen) en tracht deze met elkaar te vergelijken.
18. Het rapport verwijst naar eventuele tijdskeuzen (ten aanzien van tijdstippen, tijdsindeling, volgorde, tijdsduur of tempo) van het beleid en gaat na in hoeverre het beleidsproces en de beleidseffecten in overeenstemming zijn met de gedane tijdskeuzen.
19. De conclusies van het rapport zijn expliciet, duidelijk en beargumenteerd.
20. Het rapport geeft ook in zijn conclusies een beoordeling van (de effecten van) een gevoerd of te voeren beleid (in termen van waarden, normen, doeleinden)
21. Het rapport geeft ten aanzien van verklarende variabelen -waaronder de beleidsuitvoering en bepaalde kenmerken van het beleidsveld- aan in hoeverre deze vatbaar zijn voor beïnvloeding door de beleidsinstantie, dus manipuleerbaar zijn.
22. Het rapport geeft in zijn conclusies naast de evaluatie een advies ten aanzien van het te voeren beleid, waarbij ook aandacht wordt besteed aan de uitvoering van het beleid.
23. De conclusies zijn in overeenstemming met de eventuele opdracht tot het onderzoek.

H. Opbouw

- 24. Het rapport heeft een duidelijke indeling
- 25. Het rapport heeft een samenvatting, alsmede conclusies(Snijders, 1989, pp. 83-84)

I. Overige eisen

- 26. Het rapport is ook voor anderen dan de wetenschapsbeoefenaars duidelijk geschreven
- 27. De financiële kosten van het onderzoek blijven binnen de begroting.
- 28. Het onderzoeksrapport wordt op een zodanig tijdstip opgeleverd dat er in het beleidsproces gebruik van kan worden gemaakt.
- 29. Het onderzoeksrapport wordt opgeleverd binnen de eventueel overeengekomen termijn
- 30. Het onderzoeksrapport geef een afweging van de positieve en negatieve maatschappelijke gevolgen van het beschikbaar komen van de onderzoeksresultaten.

Bijlage 4. Geaggregeerde scores van de beide sets eisen voor alle rapporten:

Dit zijn gemiddelde scores voor alle rapporten opgesplitst naar de twee sets eisen.

	Wetenschappelijk	Beleidspraktijk
1	4,48	4,23
2	4,26	4,26
3	4,13	4,00
4	4,43	4,56
5	4,35	3,89
6	4,17	3,96
7	4,43	4,42
8	4,35	4,30
9	4,26	4,26
10	4,65	4,59
11	3,43	4,04
12	4,35	4,20
13	4,00	4,07
14	4,61	4,30
15	4,48	4,22
16	3,35	3,57
17	3,43	3,78
18	4,43	4,00
19	4,22	4,04
20	4,39	4,26
21	4,35	4,26
22	4,30	4,37
23	4,30	4,15
24	4,22	4,22

Bijlage 5. Uitkomsten gematchde paren gemeentes

Apeldoorn <> Deventer

Group Statistics

Wetenschappers		N	Mean	Std. Deviation	Std. Error Mean
Wetenschappelijke eisen	Wel Wetenschappers	3	4,3267	,23714	,13691
	Geen Wetenschappers	3	4,3167	,13317	,07688
Eisen uit de Beleidspraktijk	Wel Wetenschappers	3	4,1633	,14224	,08212
	Geen Wetenschappers	3	4,1367	,36828	,21263

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Wetenschappelijke eisen	Equal variances assumed	1,355	,309	,064	4	,952	,01000	,15702	-,42596	,44596
	Equal variances not assumed			,064	3,147	,953	,01000	,15702	-,47674	,49674
Eisen uit de Beleidspraktijk	Equal variances assumed	4,844	,093	,117	4	,913	,02667	,22794	-,60619	,65952
	Equal variances not assumed			,117	2,584	,915	,02667	,22794	-,76955	,82288

Den Bosch <> Venlo

Group Statistics

Wetenschappers		N	Mean	Std. Deviation	Std. Error Mean
Wetenschappelijke eisen	Wel Wetenschappers	3	4,3467	,08505	,04910
	Geen Wetenschappers	3	4,1433	,63571	,36703
Eisen uit de Beleidspraktijk	Wel Wetenschappers	3	4,3267	,08327	,04807
	Geen Wetenschappers	3	4,2767	,28290	,16333

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Wetenschappelijke eisen	Equal variances assumed	7,800	,049	,549	4	,612	,20333	,37030	-,82478	1,23145
	Equal variances not assumed			,549	2,072	,636	,20333	,37030	-1,33833	1,74500
Eisen uit de Beleidspraktijk	Equal variances assumed	4,108	,113	,294	4	,784	,05000	,17026	-,42272	,52272
	Equal variances not assumed			,294	2,344	,793	,05000	,17026	-,58862	,68862

Bacheloropdracht Wybe Kloppenburg s0026360

Hardenberg <> Wijk bij Duurstede

Group Statistics

	Wetenschappers	N	Mean	Std. Deviation	Std. Error Mean
Wetenschappelijke eisen	Wel Wetenschappers	3	4,3633	,32130	,18550
	Geen Wetenschappers	3	3,7233	,61232	,35352
Eisen uit de Beleidspraktijk	Wel Wetenschappers	3	4,1967	,11676	,06741
	Geen Wetenschappers	3	3,7833	,21502	,12414

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Wetenschappelijke eisen	Equal variances assumed	2,746	,173	1,603	4	,184	,64000	,39924	-,46846	1,74846
	Equal variances not assumed			1,603	3,024	,207	,64000	,39924	-,62492	1,90492
Eisen uit de Beleidspraktijk	Equal variances assumed	,604	,480	2,926	4	,043	,41333	,14126	,02112	,80555
	Equal variances not assumed			2,926	3,085	,059	,41333	,14126	-,02929	,85596

Veenendaal <> Ede

Group Statistics

	Wetenschappers	N	Mean	Std. Deviation	Std. Error Mean
Wetenschappelijke eisen	Wel Wetenschappers	3	4,3200	,08888	,05132
	Geen Wetenschappers	3	4,2733	,04619	,02667
Eisen uit de Beleidspraktijk	Wel Wetenschappers	3	4,1867	,12702	,07333
	Geen Wetenschappers	3	4,2467	,11240	,06489

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Wetenschappelijke eisen	Equal variances assumed	1,975	,233	,807	4	,465	,04667	,05783	-,11390	,20723
	Equal variances not assumed			,807	3,007	,479	,04667	,05783	-,13714	,23048
Eisen uit de Beleidspraktijk	Equal variances assumed	,169	,702	-,613	4	,573	-,06000	,09792	-,33188	,21188
	Equal variances not assumed			-,613	3,942	,574	-,06000	,09792	-,33347	,21347