

De invloed van een beslissingson- dersteuning en ervaring op de beoordeling van

DR. E.H.J. VAASSEN RA EN DRS. V.K. BRÖCHELER¹

Bij het afnemen van examens Administratieve Organisatie wordt ervan uitgegaan dat het oordeel van de beoordelaar, al dan niet in samenspraak met een tweede beoordelaar, een juiste indicator is van de kwaliteit van een uitwerking. Consensus tussen beoordelaren wordt in het algemeen beschouwd als een goede maatstaf voor beslissingskwaliteit. Dit artikel onderzoekt in hoeverre onderwijs- en examenervaring enerzijds en het gebruik van een checklist anderzijds kunnen bijdragen aan het bereiken van meer consensus onder beoordelaren. Consensus die hard nodig lijkt te zijn.

De accountant maakt al in een vroeg stadium van de controle een inschatting van de kwaliteit van de administratieve organisatie en interne controle (AO/IC). Globaal kan worden gesteld dat naarmate de kwaliteit van de AO/IC beter wordt, de accountant bij zijn controle meer zal steunen op die AO/IC en meer systeemgerichte controles zal uitvoeren. Bij het inschatten van de kwaliteit van de AO/IC vergelijkt de accountant de werkelijke bij een cliënt aangetroffen AO/IC met een normatieve AO/IC-beschrijving. Accountants moeten dus over vaardigheden beschikken om cliëntspecifieke normatieve AO/IC-beschrijvingen samen te stellen.

Accountancy-studenten moeten derhalve vaardigheden ontwikkelen om cliëntspecifieke normatieve IC-beschrijvingen samen te stellen. Traditioneel worden accountancy-studenten getraind in het uitwerken van cases betreffende een groot aantal verschillende bedrijven. De tijdens deze trainingen gemaakte normatieve AO/IC-beschrijvingen worden beoordeeld door accountants met onderwijs- en examenervaring. Voor de opleiding van accountants betekent dit dat accountants die meer ervaring hebben in het toekennen van kwaliteitspredikaten aan AO/IC-beschrijvingen zoals samengesteld door accountancy-studenten waarschijnlijk over meer deskundigheid inzake het beoordelen van AO/IC-beschrijvingen beschikken. Een grotere deskundigheid leidt in het algemeen tot snellere en betere beslissingen. In de opleiding tot accountant zal

superieur beslissingsgedrag blijken uit het feit dat docenten beter in staat zijn de goede van de slechte AO/IC-beschrijvingen zoals samengesteld door accountancy-studenten te onderscheiden en op basis daarvan betere feedback naar deze studenten te verzorgen. In de praktijk blijken docenten en examinatoren (!) er echter vaak geheel eigen ideeën op na te houden over wat goede en wat slechte AO/IC-beschrijvingen zijn. De verhitte discussies tussen studenten over identieke uitwerkingen die bij docent X met een ruime voldoende zijn beoordeeld en bij docent Y met een onvoldoende, spreken boekdelen.

De accountantsopleiding moet een gezonde basis vormen voor accountancy-studenten om goede beoefenaren van het accountantsvak te worden. Als accountantskantoren de concurrentie in de markt voor zakelijke dienstverlening het hoofd willen blijven bieden, moeten zij er voor zorgen dat zij mensen aannemen die beschikken over die gezonde basis. Dit betekent dat accountantsopleidingen voortdurend moeten uitkijken naar manieren om het onderwijsleerproces te verbeteren in termen van efficiency en effectiviteit. Beslissingsondersteunende hulpmiddelen kunnen bijdragen aan een grotere efficiency en effectiviteit van de accountantsopleiding omdat docenten, bij gebruikmaking van die hulpmiddelen, bijvoorbeeld beter onderscheid kunnen maken tussen goede en slechte AO/IC-beschrijvingen zoals samengesteld door studenten en derhalve be-

tere feedback kunnen geven naar die studenten. Dit betekent dat de ontwikkeling van beslissingsondersteunende hulpmiddelen indirect leidt tot een superieure deskundigheidsontwikkeling onder accountancy-studenten. Verder kan de ontwikkeling van beslissingsondersteunende hulpmiddelen ertoe bijdragen dat de deskundigheid van enkelen op grote schaal kan worden ingezet, waardoor ook door minder ervaren en dientengevolge minder deskundige docenten snellere en betere beslissingen kunnen worden genomen dan mogelijk zou zijn zonder gebruikmaking van de desbetreffende beslissingsondersteunende hulpmiddelen.

Checklists zijn te beschouwen als beslissingsondersteunende hulpmiddelen. De ontwikkeling van een checklist, bestaande uit elementen die bijdragen aan de kwaliteit van een AO/IC-systeem, zou dan ook kunnen bijdragen aan de efficiency en effectiviteit van het vak AO in de accountantsopleidingen.

In dit artikel wordt onderzocht of de kwaliteit van beoordelingen van normatieve AO/IC-beschrijvingen kan worden verbeterd als docenten met een verschillend aantal jaren beoordelingservaring bij het toekennen van kwaliteitspredikaten gebruikmaken van een checklist¹.

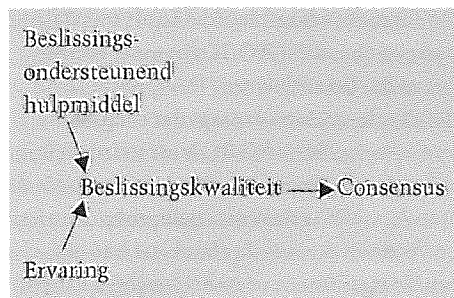
Onderzoeksmethode

Model

Op basis van voorgaand onderzoek (Bedard 1991, Weber 1978, Tversky 1977) kan een model worden ontwikkeld waarin de effecten van beslissingsondersteunende hulpmiddelen en ervaring op beslissingskwaliteit zijn gerepresenteerd. Omdat uit voorgaand onderzoek blijkt dat het gebruik van beslissingsondersteunende hulpmiddelen (Huber, 1989) en ervaring (Bonner & Lewis, 1990) tot superieur beslissingsgedrag leiden, is de verwachting dat onervaren beslissers hun besluitvorming kunnen verbeteren door toepassing van een beslissingsondersteunend hulpmiddel.

Ersteunend hulpmiddel O/IC-beschrijvingen

Verder is de verwachting dat consensus zal toenemen als de beslissingskwaliteit toeneemt (Ashton, 1985). Het model dat wordt getest, is weergegeven in figuur 1.



Figuur 1 De relatie tussen het gebruik van beslissingsondersteunend hulpmiddel, ervaring en consensus.

Onderzoeksvragen

Er zijn twee dichotome variabelen die, geïsoleerd of in combinatie, de mate van consensus en de hoogte van de toegekende kwaliteitspredikaten kunnen verklaren: de behandeling (wel/geen checklist) en beoordelingservaring (ja/nee). Om de relatie tussen behandeling, ervaring en consensus vast te stellen, worden de volgende onderzoeksvragen gesteld:

Vraag 1
Leidt toepassing van een checklist AO/IC tot meer consensus onder beoordelaren?

Vraag 2
Is er onder ervaren beoordelaren meer consensus dan onder minder ervaren beoordelaren?

Vraag 3
Zijn er interactie-effecten van ervaringen gebruikmaking van een checklist op de mate van consensus?

Om vast te stellen of er neveneffecten optreden op de hoogte van de toegekende kwaliteitspredikaten, worden de volgende onderzoeksvragen gesteld:

Vraag 4
Leidt toepassing van een checklist AO/IC tot afwijkende kwaliteitspredikaten?

Vraag 5
Geven ervaren beoordelaren andere kwaliteitspredikaten dan minder ervaren beoordelaren?

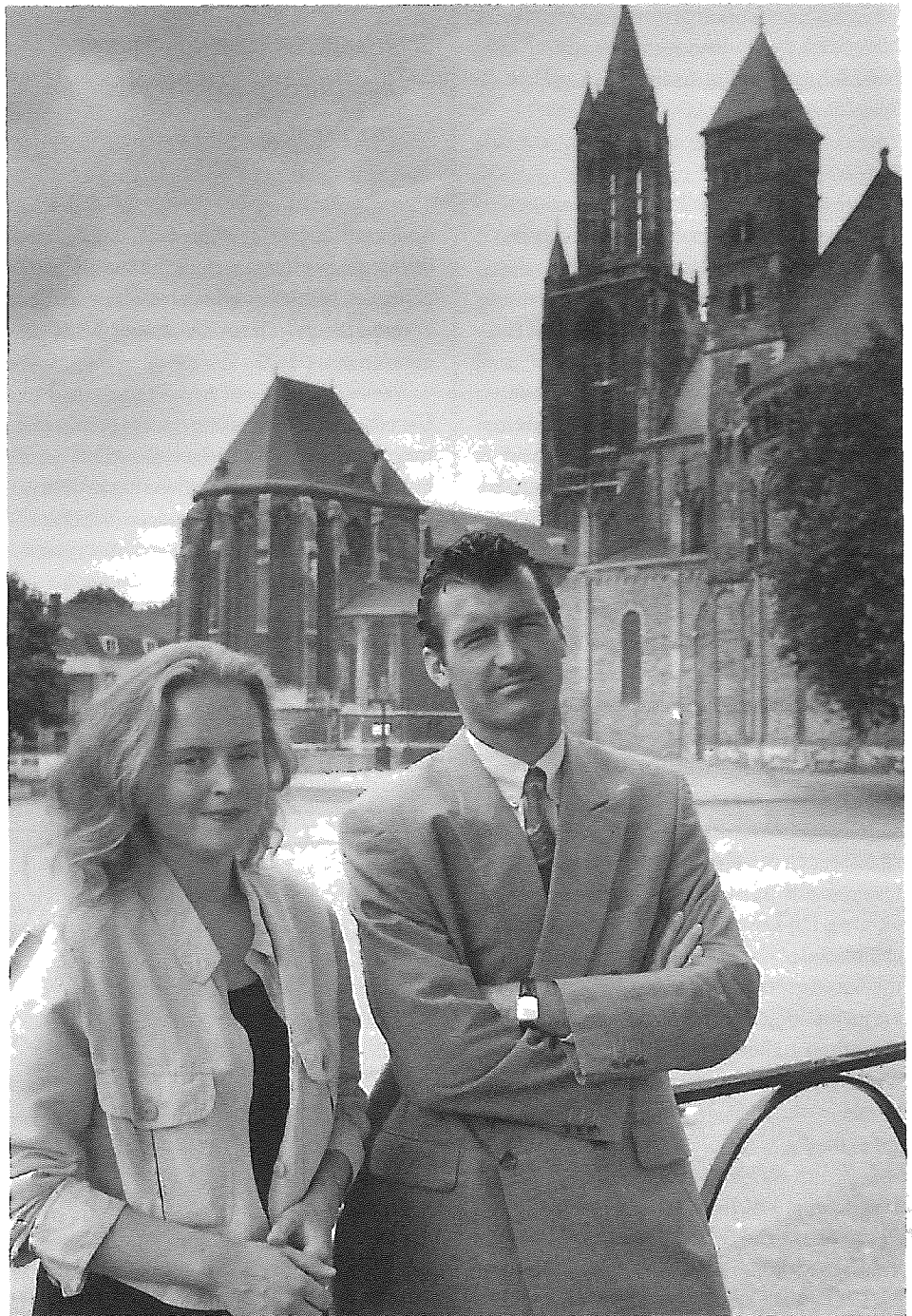
Vraag 6
Zijn er interactie-effecten van ervaring en gebruikmaking van een checklist op de hoogte van de kwaliteitspredikaten?

Taak

Er werd een experiment opgezet waarbij aan subjecten werd gevraagd de uitwerkingen van twee studenten van een AO-case te beoordelen op een schaal van één tot tien. Er werden twee groepen subjecten onderscheiden:

- groep 0; beide uitwerkingen dienden te worden beoordeeld aan de hand van een normatieve oplossing in de vorm van een traditioneel schema van uitwerking;
- groep 1; beide uitwerkingen dienden te worden beoordeeld aan de hand van een norma-

Vaassen (rechts) en Bröcheler: Een verdergaande studie naar mogelijkheden tot vergroting van consensus tussen beoordelaren is meer dan alleen nastrevenswaardig.



tieve oplossing in de vorm van een traditioneel schema van uitwerking aangevuld met een checklist bestaande uit 97 elementen verdeeld over negen categorieën²⁾.

In de checklist moest per uitwerking worden aangegeven of elk element op correcte wijze was verwerkt (code = 1) dan wel afwezig of incorrect was verwerkt (code = 0).

Subjecten

De onderzoekspopulatie bestond uit 53 registeraccountants en 22 postdoctorale accountancy-studenten. De register-accountants hadden een gemiddeld aantal jaren beoordelingservaring van 12,82. Verder waren zij allen verbonden aan het NIVRA dan wel aan één van de postdoctorale opleidingen tot registeraccountant of registercontroller in Nederland. De studenten hadden geen beoordelingservaring maar hadden wel het landelijk schriftelijk AO-examen behaald.

Case

De case die werd gebruikt was een examen-case van het NIVRA zoals die in het voorjaar van 1987 aan examenkandidaten werd aangeboden. Het betrof een productie-onderneming in de fijnmechanische industrie. Het bijbehorende schema van uitwerking was samengesteld door de redactiecommissie die ook de case-beschrijving had gemaakt. Gezien de zwaarwegende belangen, de deskundigheid van de redactiecommissieleden en het feit dat er in de normeringsvergadering na afloop van het examen consensus over de inhoud werd bereikt, zijn er géén redenen om aan te nemen dat het aldus samengestelde schema van uitwerking ernstige tekortkomingen vertoont. Aangezien er géén checklist bij het schema van uitwerking was bijgevoegd, diende deze door de onderzoekers te worden ontwikkeld. Om ervoor zorg te dragen dat de samen te stellen checklist en het schema van uitwerking volledig onderling vergelijkbaar waren, diende er een één op één relatie te bestaan tussen de checklist en het schema van uitwerking. Dit kwam er op neer dat in de checklist slechts die elementen mochten worden opgenomen die ook in het schema van uitwerking waren opgenomen en dat alle elementen uit het schema van uitwerking in de checklist werden opgenomen.

Resultaten³⁾

Toegekende kwaliteitspredikaten

De aan de twee uitwerkingen toegekende kwaliteitspredikaten zijn weergegeven in tabel 1. Het blijkt dat er conform verwach-

ting zowel onder ervaren als onder minder ervaren beoordelaren relatief grote verschillen in kwaliteitsperceptie bestaan. Voor de normatieve AO/IC-beschrijving van student 1 (NICB 1) was het laagste cijfer 4,00 en het hoogste 7,60 (variatiebreedte: 3,60). Als een differentiatie naar ervaringscategorie werd gemaakt, bleek de variatiebreedte voor de ervaren groep 3,60 te zijn en voor de minder ervaren groep 3,00. Voor de normatieve AO/IC-beschrijving van student 2 (NICB 2) was het laagste cijfer 3,00 en het

Het blijkt dat er zowel onder ervaren als onder minder ervaren beoordelaren relatief grote verschillen in kwaliteitsperceptie bestaan.

hoogste 8,00 (variatiebreedte 5,00). Als een differentiatie naar ervaringscategorie werd gemaakt, bleek de variatiebreedte voor de ervaren groep 3,70 te zijn en voor de minder ervaren groep 5,00.

Uit tabel 1 blijkt dat er een grote spreiding in toegekende cijfers aan identieke AO/IC-beschrijvingen is. Verschillen tussen de drie en vier punten kunnen in onderwijsleersituaties wellicht nog tot interessante discussies leiden, mits daartoe de juiste didactische werkvormen worden gekozen. Bijvoorbeeld:

de inschakeling van twee docenten die onafhankelijk van elkaar corrigeren en in elkaars bijzijn feedback naar de studenten verzorgen. In examensituaties zijn dergelijke verschillen echter te allen tijde verwerpelijk. Verschillen van deze omvang kunnen onzes inziens *niet* "gladgestreken" worden in een afstemmingsgesprek tussen twee examinatoren. Dit is een reden temeer om te onderzoeken wat de effecten van ervaring en het gebruik van beslissingsondersteunende hulpmiddelen kunnen zijn.

Onderzoeksvraag 1

Onderzoeksvraag 1 luidde: Leidt toepassing van een checklist AO/IC tot meer consensus onder beoordelaren? De hypothese bij deze onderzoeksvraag was dat toepassing van een checklist AO/IC leidt tot meer consensus onder beoordelaren. Het bleek dat de standaarddeviaties tussen de groep die niet en de groep die wél de checklist had gebruikt significant verschillen (bij $\alpha = 0,10$) in de verwachte richting voor NICB 1, maar niet significant verschillen voor NICB 2.

Onderzoeksvraag 2

Onderzoeksvraag 2 luidde: Is er onder ervaren beoordelaren meer consensus dan onder minder ervaren beoordelaren? De hypothese bij deze onderzoeksvraag was dat ervaren beoordelaren meer consensus zullen vertonen dan minder ervaren beoordelaren. Het bleek dat de standaarddeviaties van de ervaren en de onervaren groep niet significant verschillen (bij $\alpha = 0,10$) voor NICB 1, maar wel significant verschillen in de verwachte richting voor NICB 2.

Tabel 1 Frequentie van de toegekende cijfers, gedifferentieerd naar ervaringscategorie, in cijferklassen gerubriceerd

Cijferklassen	Beoordelingservaring		Totaal	Beoordelingservaring		Totaal
	Nee	Ja		Nee	Ja	
	NICB 1			NICB 2		
3,00 -< 3,50	0	0	0	1	3	4
3,50 -< 4,00	0	0	0	0	3	3
4,00 -< 4,50	0	7	7	3	9	12
4,50 -< 5,00	3	6	9	5	10	15
5,00 -< 5,50	5	10	15	2	17	19
5,50 -< 6,00	6	8	14	4	3	7
6,00 -< 6,50	2	11	13	3	6	9
6,50 -< 7,00	2	3	5	2	2	4
7,00 -< 7,50	2	5	7	1	0	1
7,50 -< 8,00	2	3	5	0	0	0
8,00 -< 8,50	0	0	0	1	0	1
Totaal	22	53	75	22	53	75

Onderzoeksvraag 3

Onderzoeksvraag 3 luidde: Zijn er interactie-effecten van ervaring en gebruikmaking van een checklist op de mate van consensus? De hypothese bij deze onderzoeksvraag was dat er een interactie-effect is tussen ervaring en gebruikmaking van een checklist op de mate van consensus tussen de beoordelaren. Voor NICB 1 werd gevonden dat het niet-significante ervaringseffect (p-waarde = 0,438) ook niet significant was als naar behandeling werd gedifferentieerd: als géén checklist werd gebruikt dan was de p-waarde 0,602, als wél een checklist werd gebruikt dan was de p-waarde 0,535. Verder was het significante behandelingseffect (p-waarde = 0,002) ook significant als naar ervaringscategorie werd gedifferentieerd: de onervaren groep had een p-waarde van 0,094, de ervaren groep had een p-waarde van 0,009. Alhoewel beide verschillen significant zijn, is het effect van een checklist dus groter in de ervaren groep van beoordelaren. Voor NICB 2 werd gevonden dat het significante ervaringseffect (p-waarde = 0,089) niet significant was als naar behandeling werd gedifferentieerd: als géén checklist werd gebruikt dan was de p-waarde 0,591, als wél een checklist werd gebruikt dan was de p-waarde 0,937. Verder was het niet-significante behandelingseffect (p-waarde = 0,141) ook niet significant als naar ervaringscategorie werd gedifferentieerd: bij de onervaren groep was de p-waarde 0,571, bij de ervaren groep was de p-waarde 0,878. In tabel 2 zijn de resultaten betreffende onderzoeksvragen 1, 2 en 3 samengevat.

Onderzoeksvraag 4

Onderzoeksvraag 4 luidde: Leidt toepassing van een checklist AO/IC tot afwijkende kwaliteitspredikaten? De hypothese bij deze onderzoeksvraag was dat toepassing van de checklist leidt tot afwijkende kwaliteits-

predikaten. Het bleek dat de gemiddelde kwaliteitspredikaten van de groep mét en de groep zonder checklist niet significant verschilden (bij $\alpha = 0,10$) voor NICB 1, maar wél significant verschilden voor NICB 2: toepassing van de checklist leidde tot lagere kwaliteitspredikaten.

Onderzoeksvraag 5

Onderzoeksvraag 5 luidde: Geven ervaren beoordelaren andere kwaliteitspredikaten dan minder ervaren beoordelaren? De hypothese bij deze onderzoeksvraag was dat ervaren beoordelaren een ander kwaliteitspredikaat geven dan minder ervaren beoordelaren. Het bleek dat de gemiddelde kwaliteitspredikaten gegeven door de ervaren en de onervaren groep niet significant verschilden (bij $\alpha = 0,10$) voor beide NICB-en.

Onderzoeksvraag 6

Onderzoeksvraag 6 luidde: Zijn er interactie-effecten van ervaring en gebruikmaking van een checklist op de hoogte van de kwaliteitspredikaten? De hypothese bij deze onderzoeksvraag was dat een combinatie van ervaring en de behandeling invloed heeft op de hoogte van de kwaliteitspredikaten. Voor NICB 1 bleek dat het niet-significante ervaringseffect (p-waarde = 0,410) ook niet significant was als naar behandeling werd gedifferentieerd: als géén checklist werd gebruikt was de p-waarde 0,429, als wél een checklist werd gebruikt, was de p-waarde 0,773. Verder was het niet-significante behandelingseffect (p-waarde = 0,629) ook niet significant als naar ervaringscategorie werd gedifferentieerd: bij de onervaren groep was de p-waarde 0,457, bij de ervaren groep was de p-waarde 0,911. Voor NICB 2 bleek dat het niet-significante ervaringseffect (p-waarde = 0,106) wél significant was als géén checklist werd

gebruikt (p-waarde = 0,013), en niet significant als wél een checklist werd gebruikt (p-waarde = 0,637). Verder was het significante behandelingseffect (p-waarde = 0,000) ook significant als naar ervaringscategorie werd gedifferentieerd (p-waarde in beide gevallen = 0,001). In tabel 3 op pagina 44 zijn de resultaten betreffende onderzoeksvragen 4, 5 en 6 samengevat.

Conclusie

Discussie

De resultaten van deze studie verschaffen geen eenduidig bewijs omtrent de relatie tussen het gebruik van een beslissingsondersteunend hulpmiddel en ervaring aan de ene kant en consensus en de hoogte van het toegekende kwaliteitspredikaat aan de andere kant. De belangrijkste bevinding voor NICB 1 was dat consensus toenam als een checklist werd gebruikt. De belangrijkste bevindingen voor NICB 2 waren dat de door beoordelaren mét een checklist toegekende kwaliteitspredikaten lager waren dan die door beoordelaren zonder een checklist. Verder veroorzaakte ervaring alleen verschillen als géén gebruik werd gemaakt van de checklist: meer ervaring leidde tot lagere kwaliteitspredikaten voor NICB 2. Dit impliceert dat onervaren beoordelaren die wél een checklist gebruikten kwaliteitspredikaten toekenden die de kwaliteitspredikaten van de ervaren beoordelaren benaderden.

Gebaseerd op voorgaande bevindingen kan worden geconcludeerd dat er verschillen moeten zijn in de gehanteerde NICB-en, welke verschillen het gebrek aan eenduidig bewijs mogelijk verklaren. Het meest in het oog springende verschil tussen de NICB-en is dat het gemiddelde toegekende kwaliteitspredikaat voor NICB 1: 5,709 is, terwijl dit voor NICB 2: 4,972 is. Het verschil tussen deze twee gemiddelden is significant bij $\alpha = 0,10$ (p-waarde = 0,000). Omdat NICB 1 waarschijnlijk superieur is aan NICB 2, kunnen de onverwachte bevindingen wellicht worden verklaard op basis van dit verschil in kwaliteit tussen de NICB-en. De verklaring zou als volgt kunnen zijn. Van ervaren beoordelaren wordt verwacht dat zij een beoordelingsgedrag ten toon spreiden, waarbij een groot verschil wordt gemaakt tussen superieure NICB-en en inferieure NICB-en. Dit, omdat zij de procedurele en de declaratieve kennis bezitten die het voor hen mogelijk maakt meer vertrouwen te hebben in hun beoordelingen. Verder wordt verwacht dat onder ervaren beoordelaren consensus bestaat over de kwaliteit van de NICB-en,

Tabel 2 Resultaten van de toetsen op de standaarddeviaties per behandeling en per ervaringscategorie voor beide NICB-en

NICB 1		Totaal	Checklist		p-waarde
			Nee	Ja	
Beoordelingservaring	Totaal	0.965	1.165	0.773	0.002
	Nee	0.910	1.052	0.792	0.094
	Ja	0.989	1.215	0.778	0.009
	p-waarde	0.438	0.602	0.535	nvt
NICB 2		Totaal	Checklist		p-waarde
			Nee	Ja	
Beoordelingservaring	Totaal	0.977	0.936	0.761	0.141
	Nee	1.165	1.032	0.766	0.571
	Ja	0.865	0.786	0.770	0.878
	p-waarde	0.089	0.591	0.937	nvt

NICB 1		Totaal	Checklist		p-waarde
			Nee	Ja	
Beoordelingservaring	Totaal	5.709	5.771	5.657	0.629
	Nee	5.852	6.020	5.713	0.457
	Ja	5.649	5.667	5.635	0.911
	p-waarde	0.410	0.429	0.773	nvt

NICB 2		Totaal	Checklist		p-waarde
			Nee	Ja	
Beoordelingservaring	Totaal	4.972	5.518	4.520	0.000
	Nee	5.300	6.120	4.608	0.001
	Ja	4.838	5.267	4.483	0.001
	p-waarde	0.106	0.013	0.637	nvt

Tabel 3 Resultaten van de toetsen op de gemiddelde kwaliteitspredikaten per behandeling en per ervaringscategorie voor beide NICB-en.

omdat hun beslissingen de (onbekende) juiste beslissingen benaderen. Het verschil tussen het laagste toegekende kwaliteitspredikaat en het hoogste toegekende kwaliteitspredikaat voor dezelfde NICB door verschillende ervaren beoordelaren is dus klein. Daarentegen is het verschil tussen het hoogste toegekende kwaliteitspredikaat en het laagste toegekende kwaliteitspredikaat aan verschillende NICB-en juist groot, ervan uitgaande dat er werkelijk een verschil is in de kwaliteit van de NICB-en. Van onervaren beoordelaren wordt juist verwacht dat zij een klein onderscheid maken tussen superieure en inferieure NICB-en. Dit omdat zij waarschijnlijk niet de procedurele en declaratieve kennis bezitten die nodig is om vertrouwen te hebben in hun eigen beslissingen. Meer specifiek kan worden gezegd dat onervaren beoordelaren kwaliteitspredikaten zullen toekennen die dicht liggen bij de omslag van adequate naar niet adequate beschrijvingen. Deze kritieke grens ligt normaliter bij het cijfer 5,5. Dus, tussen onervaren beoordelaren is het verschil tussen het hoogste toegekende kwaliteitspredikaat en het laagste toegekende kwaliteitspredikaat aan verschillende NICB-en klein, zelfs als er een groot verschil in kwaliteit is tussen de NICB-en. Hierdoor is het verschil tussen het hoogste en het laagste toegekende kwaliteitspredikaat voor dezelfde NICB ook klein.

Concluderend kan worden gesteld dat de ervaren en de onervaren beoordelaren verschillen op twee manieren: ten eerste kan de variabiliteit binnen de groep ervaren beoordelaren groter zijn voor verschillende NICB-en, afhankelijk van een werkelijk verschil in kwaliteit, dan de variabiliteit van de onervaren beoordelaren door een besluitvormingsvertrouwen effect. Ten tweede kan de variabiliteit van de ervaren beoordelaren voor de

zelfde NICB kleiner zijn dan de variabiliteit van de onervaren beoordelaren door een expertise effect.

Implicaties voor theorie en praktijk

Deze studie draagt bij aan audit judgment-theorie in die zin dat zij een beter inzicht verschaft in het concept van ervaring en het gebruik van een beslissingsondersteunend hulpmiddel in relatie tot consensus en besluitvormingsnauwkeurigheid. Verder verschaft deze studie een basis voor het starten van een onderzoekprogramma gericht op het verifiëren van de voorgestelde theorie van consensus en besluitvormingsnauwkeurigheid. De bijdrage aan de controlepraktijk is van een

Deze studie draagt bij aan audit judgment-theorie in die zin dat zij een beter inzicht verschaft in het concept van ervaring en het gebruik van een beslissingsondersteunend hulpmiddel in relatie tot consensus en besluitvormingsnauwkeurigheid.

indirecte aard omdat de ontwikkeling van beslissingsondersteunende hulpmiddelen ten behoeve van onderwijs-leerprocessen leidt tot een verhoogde effectiviteit (betere feedback aan studenten) en efficiency (meer gestandaardiseerde beoordelingen) binnen de opleiding tot accountant. Blijkens de onderzoeksresultaten is

een verdergaande studie naar mogelijkheden tot vergroting van consensus tussen beoordelaren meer dan alleen nastrevenswaardig. Immers, verschillen van meer dan drie punten (op een tienpuntsschaal) voor volkomen identieke uitwerkingen mogen zowel in de onderwijsfeer als in de examensfeer eigenlijk niet voorkomen.

Literatuur

Ashton A.H., *Does Consensus Imply Accuracy in Accounting Studies of Decision Making?*, *The Accounting Review*, Volume 60, 1985.
 Bedard J., *Expertise and its Relation to Audit Decision Quality*, *Contemporary Accounting Research*, Volume 8, no 1, Fall, 1991.
 Bonner S.E., B.L. Lewis, *Determinants of Auditor Expertise*, *Journal of Accounting Research*, Volume 28, Supplement, 1990.
 Huber V.L., *Comparison of the Effects of Specific and General Performance Standards on Performance Appraisal Decisions*, *Decision Sciences*, 1989.
 Tversky A., *Features of Similarity*, *Psychological Review*, 1977.
 Weber R., *Auditor Decision Making on Overall System Reliability: Accuracy, Consensus and the Usefulness of a Simulation Decision Aid*, *Journal of Accounting Research*, Autumn, 1978. ■

Noten

*) Dr. E.H.J. Vaassen RA is projectmanager onderwijsvernieuwing bij het Koninklijk NIVRA, universitair hoofddocent aan de Rijksuniversiteit Limburg en HRM-functionaris bij Deloitte & Touche. Drs. V.K. Bröcheler is als assistent-in-opleiding verbonden aan de sectie Organisatie en het MARC (Maastricht Accounting en Auditing Research Center) van de Rijksuniversiteit Limburg.
 1) Een uitgebreidere versie van dit artikel komt beschikbaar als Research Memorandum aan de Rijksuniversiteit Limburg.
 2) Dank is verschuldigd aan drs. M. Franssen RA en drs. M. Zinken RA wegens hun hulp bij het concipiëren van de checklist.
 3) Standaarddeviaties werden vergeleken door middel van een F-toets, gemiddelden werden vergeleken middels een t-toets.