

TIJD IS GELD: De e-learning paradox. Schrijven van managementletters via e-learning

Kees Maat & Bert Meuffels

Al jarenlang verzorgt het bureau Vergouwen Overduin trainingen in het schrijven van managementletters voor accountants van verschillende grote accountantsmaatschappijen. Om redenen van efficiëntie zijn de trainingen enige tijd geleden omgezet naar een 'blended' vorm: deels e-learning en deels classroom learning.

Om de deelnemers in staat te stellen het e-learningtraject zo snel mogelijk te doorlopen, is een diagnostische voortoets ontwikkeld die uitmondt in specifieke adviezen over het optimale leerpad door de e-learning. In dit e-learning programma worden 20 verschillende subvaardigheden onderscheiden die verondersteld worden een rol te spelen bij het schrijven van managementletters. De voortoets sluit hierbij aan: elke subvaardigheid wordt aan de hand van een aparte subtoets gemeten met vier items. Op basis van het resultaat op deze subtoetsen volgt voor elke deelnemer een op maat gesneden advies: welke van de 20 onderscheiden subvaardigheden moeten grondig geoefend worden, welke niet?

De deelnemers vinden echter - gezien de werkdruk van accountants - dat die voortoets een te groot beslag op hun kostbare tijd legt. De centrale vraag van dit onderzoek luidt dan ook: is het mogelijk elk van de 20 subtoetsen in te korten van vier naar drie items, zonder al te groot verlies aan betrouwbaarheid en validiteit, zodat de deelnemers minder tijd kwijt zijn aan het afleggen van de (totale) toets?

De toets, afgenomen aan 147 accountants, is onderworpen aan een psychometrische analyse: de betrouwbaarheid en constructvaliditeit van de 20 4-item subtoetsen is vergeleken met die van de 3-item subtoetsen. Qua constructvaliditeit onderscheiden de 3-item subtoetsen zich niet van de 4-item toetsen; evenmin is er sprake van een forse reductie in betrouwbaarheid.

In de discussie worden twee (tentatieve) verklaringen voorgesteld voor de resultaten van dit psychometrisch onderzoek.

1 Een paradox

In een tekst van een accountantsorganisatie aan de directie van een bank was enige tijd geleden het volgende fragment te lezen.

De geldende richtlijnen voor acceptatie van hypotheekleningen worden in vele gevallen niet nageleefd. Er worden leningen verstrekt die de executiewaarde van het onderpand overschrijden, taxatierapporten worden vaak niet opgevraagd en de kredietwaardigheid van de debiteur wordt niet gecontroleerd.

Het risico dat verliezen ontstaan doordat bij insolventie van de debiteur de waarde van het onderpand ontoereikend is om de hypotheekschuld te voldoen is hoog.

Advies Wij adviseren u erop toe te zien dat de richtlijnen strikt worden nageleefd.

Dit fragment is illustratief voor het tekstgenre dat in dit artikel centraal staat: de *managementletter*. Dat is een tekst waarin de accountant problemen signaleert in een organisatie. Het probleem dat hier aangestipt wordt, kennen we sinds de hypotheekcrisis allemaal. Opvallend is dat de accountant dit toch vrij ernstige probleem behandelt als

ware het louter een procedurele kwestie. Verder valt op dat de accountant de lezer expliciet van advies dient en dat die lezer wordt aangesproken met het formele ‘u’ (die lezer is de leiding, het bestuur van de organisatie, de directie). Verder illustreert het fragment dat de schrijver in kwestie bepaald geen gemakkelijk leesbare, voor iedereen (ook voor buitenstaanders) toegankelijke stijl hanteert – en dat, terwijl de managementletter in accountancy-kringen juist het visitekaartje van de accountant genoemd wordt.

Om de kwaliteit van de managementletter te verbeteren, verplichten alle grote en middelgrote accountantsmaatschappen (aankomende) accountants die doorgaans de controle bij een organisatie uitvoeren (de zogenoemde controleleiders) deel te nemen aan trainingen in het schrijven van managementletters. Al jarenlang verzorgt het bureau Vergouwen Overduin zulke trainingen; om redenen van efficiëntie (tijdwinst) zijn de trainingen enige tijd geleden op verzoek van de maatschappen omgezet naar een ‘blended’ vorm: deels e-learning en deels classroom-training. Een van de voordelen van e-learning die consequent in de literatuur naar voren worden gebracht, is *tijdwinst* (Rosenberg, 2001; Kirkpatrick, 1998): elke deelnemer kan zich op elk willekeurig tijdstip of plaats de oefenstof eigen maken; daar komt bij dat de deelnemers uitsluitend die onderdelen hoeven te oefenen die voor hen relevant zijn.¹

Om de deelnemers aan dit trainingsprogramma in staat te stellen in zo kort mogelijke tijd het e-learning traject af te ronden, is een diagnostische voortoets ontwikkeld (bestaande uit 80 items) die uitmondt in specifieke adviezen over het optimale leerpad: welke vaardigheden moeten grondig geoefend worden, welke vaardigheden kan de accountant-schrijver met een gerust hart overslaan? De deelnemers echter vinden, blijkens de evaluaties, dat deze toets een *te groot beslag* legt op hun kostbare tijd: een ware e-learning paradox. Hoe beter een schrijfvaardigheid qua inhoud in het e-learning traject wordt gedekt, hoe meer items in de toets aangeboden moeten worden, hoe groter de gevergd tijdsinvestering – waarmee juist een van de veronderstelde voordelen van e-learning verloren gaat. Is het mogelijk aan dit dilemma te ontkomen en de toets in te korten van 80 naar 60 items zonder verlies aan betrouwbaarheid en validiteit, zodat het afleggen van de toets de deelnemers minder tijd kost?

2 De accountant en de managementletter

Er bestaan vele misverstanden over de functie van accountant², bijvoorbeeld dat de accountant de boekhouding zou doen voor een organisatie. Maar dat is niet het geval: de accountant controleert ‘slechts’ of de cijfers die een organisatie jaarlijks publiceert in de vorm van de jaarrekening, een ‘getrouw’ beeld geven van de financiën van die organisatie, en hij geeft daarover een publiek oordeel. In feite controleert de accountant of de boekhouders hun werk goed hebben gedaan. Bij grote en middelgrote ondernemingen is de accountantscontrole zelfs wettelijk verplicht. De accountant legt zijn oordeel over de getrouwheid van de jaarrekening vast in de zogenoemde accountantsverklaring, een tekst die samen met de jaarrekening gepubliceerd wordt en die dient om het maatschappelijk verkeer zekerheid te geven over die cijfers. Accountantsverklaringen (er zijn verschillende typen) zijn standaardteksten met een voorgeschreven formulering; ze zijn bestemd voor verschillende ‘stake-holders’ van de organisatie, zoals aandeelhouders, overheid of banken.

Naast de accountantsverklaring produceert de accountant één of twee maal per jaar nog een tekst die echter niet-standaard is: de zogenoemde managementletter. Dat is een tekst aan de leiding van de organisatie waarvan de jaarrekening wordt ge-

controleerd. In de managementletter schrijft de accountant over tekortkomingen of ‘verbeterpunten’ in de gecontroleerde organisatie.

Zeker bij grote organisaties, zoals multinationals, is de controle van de jaarrekening een complex en omvangrijk werk, waarbij de accountant niet alleen de cijfers, maar ook allerlei andere aspecten van de bedrijfsvoering analyseert. Zo zal de accountant niet alleen getallen moeten controleren, maar ook dossiers moeten lezen, na moeten gaan of wettelijke regelingen terecht zijn toegepast, interviews moeten houden met functionarissen, enzovoort. Al deze werkzaamheden worden uitgevoerd door het controleteam: dat is een team van (meestal aankomende) accountants dat de controle bij een klant uitvoert.

In de loop van de werkzaamheden komt het controleteam soms zaken tegen die niet in orde zijn of die verbetering behoeven, of die de controle van de jaarrekening bemoeilijken. Vanuit zijn (zoals dat in de accountancy heet) ‘natuurlijke adviesfunctie’ heeft de accountant de plicht om deze ‘onvolkomenheden’ te melden. Dat gebeurt in de managementletter.

Wat is nu de inhoud van zo’n managementletter? In de managementletter ligt het accent vaak op de administratieve organisatie (zijn de werkprocessen zo ingericht dat de juiste cijfers op tijd beschikbaar zijn, worden begrotingen gemaakt en worden nacalculaties gedaan, baseert de boekhouding zich op de juiste documenten?) en op de interne controle (wordt zó gewerkt dat fouten en fraudes aan het licht kunnen komen, worden betalingen door een kassier bijvoorbeeld gecontroleerd door een medewerker van de boekhouding?). Maar het kan ook gaan om zaken van meer algemene aard (is het bestuur van de organisatie in voldoende mate ‘in control’, wordt voldaan aan wettelijke eisen?). Op al deze gebieden kunnen zich situaties voordoen die niet volgens de regels zijn of die risico’s bevatten voor de continuïteit van de organisatie. Dergelijke zaken worden dan gemeld in de managementletter.

De inhoud van de managementletter bevat dus vooral onvolkomenheden: het is een slechtnieuwsbrief. De managementletter kan door de klant al snel worden opgevat als kritiek – iets wat in de praktijk dan ook regelmatig gebeurt. Duidelijk mag zijn dat de managementletter bijzonder hoge eisen aan de schrijfvaardigheid van de schrijvers stelt.

Hoewel je zou verwachten dat de managementletter, gezien zijn specifieke aard en functie, in de praktijk wordt geschreven door senior-accountants met veel schrijfervaring, is dat vaak niet zo. Veel managementletters worden juist geschreven door de managers van de controleteams: de controleleiders, meestal aankomende of net afgestudeerde accountants met amper enige schrijfervaring. Vanuit dit perspectief is het niet verwonderlijk dat de grote en middelgrote maatschappen alle nieuwe controleleiders verplichten deel te nemen aan een training in het schrijven van managementletters. Een dergelijke training in het schrijven van managementletters wordt in ‘blended’ vorm aangeboden door bureau Vergouwen Overduin, waaraan jaarlijks enkele honderden (aankomende) accountants deelnemen. Dat is onze doelgroep die de voorloper bij het e-learning traject heeft afgelegd.

3 Van classroom naar blended training

De toenemende noodzaak bij de maatschappen tot efficiënter werken heeft er enkele jaren geleden toe geleid dat de door het bureau verzorgde classroom training in het schrijven van managementletters is vervangen door een variant die minder beslag legt op de (kostbare) tijd van de deelnemers: blended training. Deze term duidt op een

combinatie van e-learning en classroom training. De fysieke (groeps)bijeenkomst wordt aldus bekort en de overblijvende leerstof wordt via de computer ‘op afstand’ gedoceerd. Overigens, het is niet zo vreemd dat de vraag om een blended opzet uit de accountancy kwam, want dit is één van de beroepsgroepen waar e-learning het eerst werd toegepast; accountants zijn dus al jarenlang gewend aan e-learning.

De oorspronkelijke training (classroom) duurde drie dagen met tussenliggende avonden en was als volgt opgezet (zie het schema in Figuur 1).



Figuur 1 Classroom learning

- De deelnemer schrijft voorafgaand aan de bijeenkomst een managementletter aan de hand van een casus.
- De docent beoordeelt de managementletters en schrijft voor iedere deelnemer een sterkte/zwakteanalyse.
- De deelnemer volgt de bijeenkomst. Alle deelnemers ontvangen hun sterkte/zwakteanalyse. In de training komt de theorie aan bod en doen de deelnemers de oefeningen, meestal in subgroepjes, soms individueel. Tijdens de bijeenkomst schrijven de deelnemers in subgroepjes bovendien een tweede managementletter die wordt beoordeeld door medecursisten en de docenten.
- Tijdens de bijeenkomst komen nog enkele vaktechnische accountancyonderwerpen aan bod.

In de nieuwe opzet is de trainingsbijeenkomst teruggebracht tot twee dagen met een tussenliggende avond en uitgebreid met een e-learningprogramma (zie Figuur 2).



Figuur 2 Blended learning

- De deelnemer doet de voortoets. Die toets leidt tot een score, een sterkte/zwakteanalyse en een advies over de te volgen onderdelen van het e-learningprogramma.
- De deelnemer volgt de relevante delen van het e-learningprogramma. Dat programma bevat alle leerstof in vijf modules die elk zijn onderverdeeld in vier inhoudelijke episodes. Per episode bevat het programma tussen de 10 en 20 oefeningen, in totaal ongeveer 300. Per oefening krijgt de deelnemer uitgebreid feedback en een link naar de – voor die oefening - relevante pagina van een elektronisch leerboek.
- De deelnemer sluit het e-learningprogramma af en evalueert dit.
- De deelnemer doet de natoets. Die is inhoudelijk min of meer identiek aan de voortoets, zodat de scores vergeleken kunnen worden. Voor de natoets moet een minimum score behaald worden (de hoogte wordt vastgesteld in

overleg met de opdrachtgever; verder geldt dat het vervolgtraject uitsluitend toegankelijk is voor deelnemers die een voldoende score hebben behaald).

- De deelnemer schrijft voorafgaand aan de bijeenkomst een managementletter aan de hand van een casus.
- De docent leest de managementletters door, voorafgaand aan de training.
- Deelnemer volgt de bijeenkomst. Tijdens de bijeenkomst komen nog uitsluitend korte terugblikken op de stof aan bod, waarna de deelnemers in rondes elkaars managementletters bespreken.
- Tijdens de bijeenkomst komen nog enkele schrijfprocesoefeningen aan bod en enkele vaktechnische accountancyonderwerpen.

Bij zowel de oorspronkelijke training *Managementletters schrijven* als bij de blended opzet is een multidisciplinair team van docenten betrokken: enerzijds schrijftainers (vaak met een taalbeheersingsachtergrond) met jarenlange ervaring in het trainen van accountants, anderzijds ervaren (senior) accountants die als co-docent optreden.

4 De inhoud van de voortoets en van de e-learning

De e-learning – en dus ook de voortoets - bevat taalbeheersingsonderdelen die ook in andere schrijfopleidingen voorkomen, zoals structureren, argumenteren, formuleren, stijl, toon en presentatie. Daarnaast bevat de het programma ook vakinhoudelijke aspecten uit de accountancy. Zo wordt bijvoorbeeld de selectie van de inhoud gebaseerd op de regelgeving over de controleaanpak van de accountant (zie de episode: *1.1 Van COS³ naar tekstopzet*). Een voorbeeld staat in Figuur 3.

<p>Hiernaast staat een situatieschets opgenomen, afgeleid uit een controledossier. Ook zie je drie mogelijke manieren om hiermee om te gaan in de managementletter.</p> <hr/> <p>Geef aan wat volgens jou in dit geval de beste aanpak is.</p> <ul style="list-style-type: none"> • De bevinding opnemen, als volgt afgerond: Ons advies was om de interne controle door middel van oogtoezicht aan te scherpen. Wij raden u dringend aan om dit advies alsnog op te volgen. • De bevinding kort samenvatten, met een verwijzing naar de rapportage van vorig jaar ('follow up'). Opnieuw aandacht vragen voor dit punt. • De bevinding niet apart opnemen. In de bijlage 'Follow up bevindingen vorig jaar' melding maken van het feit dat de interne controle op de baromzet nog niet is verbeterd. 	<p>Vorig jaar hebben wij gerapporteerd over het gebrek aan interne controle op de baromzet. Dit jaar constateerden wij bij onze controle (door middel van waarneming ter plaatse) dat de barkeppers nog steeds op verzoek van klanten cocktails schenken die niet op de drankkaart vermeld staan. Deze consumpties kunnen niet via het POS-systeem worden afgerekend. De gemiddelde prijs van een cocktail bedraagt volgens de drankkaart euro 12,25. Het verschil tussen de afgedragen kassa-omzet en de baromzet volgens POS bedraagt ongeveer euro 1.100 (positief), dit is 0,7% van de totale baromzet. Of dit verschil geheel betrekking heeft op de cocktailverkoop is niet bekend.</p>
--	---

Figuur 3 Voorbeeld van een opgave uit de episode *Van COS naar tekstopzet*

In de de module *Lezergerichte Bevindingen* gaat het bijvoorbeeld over de opbouw en de argumentatie in de 'bevindingen' (zie paragraaf 1 voor een voorbeeld van een fragment van een bevinding).

De tabel hieronder geeft een overzicht van de zeer diverse inhoud van de e-learning en de daaraan voorafgaande toets.

Tabel 1 Inhoud van de e-learning en de toets

Modules	Episodes
1. Efficiënte tekstopzet	1.1 Van COS naar tekstopzet
	1.2 Bevindingen logisch indelen
	1.3 Samenvatten, inleiden, afsluiten
	1.4 Presentatievorm kiezen
2. Lezegerichte bevindingen	2.1 Van dossier naar managementletter
	2.2 Ingrediënten herkennen
	2.3 Relevante informatie geven
	2.4 Overtuigend adviseren
3. Toegankelijk betoog	3.1 Duidelijk verwijzen
	3.2 Adequaat opsommen
	3.3 Zinnen aan elkaar knopen
	3.4 Overzichtelijke alinea's maken
4. Aantrekkelijke formulering	4.1 Concrete zinnen schrijven
	4.2 Taal van de lezer gebruiken
	4.3 Kiezen voor actief of passief
	4.4 Informatie doseren
5. Uitstraling en toon	5.1 Formuleren met rechte rug
	5.2 Professioneel schrijven
	5.3 Formuleren met respect
	5.4 De lezer aanspreken

Zowel het e-learningprogramma als de beide toetsen zijn ontwikkeld door de schrijfdocenten van de training *Managementletters schrijven*, in samenwerking met de accountant-docenten.

In de voortoets zijn alle 20 episodes vertegenwoordigd door vier items; op grond van de scores op deze 20 subtoetsen volgt voor elke deelnemer een op maat gesneden advies over het te volgen e-learningtraject. De oorspronkelijke vraag of de 80-item toets ingekort kan worden tot een 60-item toets moet dus in feite geherformuleerd worden tot: kunnen de 20 episodes elk verkort worden van vier tot drie items, zonder verlies aan betrouwbaarheid en validiteit?

5 Betrouwbaarheid

5.1 Opzet

In totaal 148 RA accountants-in-spe hebben in de periode 2006-2007 de diagnostische voortoets afgelegd en het daarop aansluitende e-learningtraject doorlopen. In de toets krijgt niet elke deelnemer per episode precies dezelfde items aangeboden: de vier items worden aselekt getrokken uit een itembank die in totaal uit 140 items bestaat. In deze itembank zijn per episode zeven items geconstrueerd waaruit per deelnemer een aselechte trekking (zonder teruglegging) plaatsvindt.⁴ De datamatrix bestaat dus structureel uit missing values die geschat moeten worden.

De betrouwbaarheid van elk van de episodes (uitgedrukt in Cronbach's alpha, de homogeniteitsindex) is berekend voor een variërend aantal items per episode, aflopend van zeven tot twee.⁵ De betrouwbaarheid is eerst berekend voor alle beschikbare zeven scores; vervolgens is aselekt een item uit de pool van zeven items verwijderd, waarna de betrouwbaarheid voor de resterende zes items werd berekend. Uit de resterende zes items is vervolgens weer aselekt een item verwijderd, enzovoort, totdat er slechts twee items resteren voor de samenstelling van de schaal.

5.2 Resultaten

In Tabel 2 zijn in exemplarische zin de resultaten weergegeven voor 2 van de 20 episodes: *1.1 Van COS naar tekstopzet* en *3.1 Duidelijk verwijzen*. Links in de tabel staat het (aflopende) aantal items waaruit de episode is samengesteld, rechts daarvan is de corresponderende betrouwbaarheid weergegeven.

Tabel 2 *Betrouwbaarheid van twee episodes bij een variërend aantal items*

Voorbeeld van betrouwbaarheid van twee episodes (1.1: Van COS naar tekstopzet; 3.1: Duidelijk verwijzen)		
Aantal items	Betrouwbaarheid	Betrouwbaarheid
	Episode 1.1	Episode 3.1
7	.82	.84
6	.80	.82
5	.79	.78
4	.80	.76
3	.76	.75
2	.71	.65

De betrouwbaarheden van deze twee episodes zijn niet alleen vrij hoog bij zeven items, maar ook bij vier items; verder valt op dat het verschil in betrouwbaarheid tussen vier en drie items per episode gering is. Dat beeld geldt overigens voor alle episodes (met uitzondering van de episode *5.1 Formuleren met rechte rug* waar de betrouwbaarheid .63 bedraagt). Concluderend kan gesteld worden dat uit het oogpunt van betrouwbaarheid een subtoets bestaande uit vier items weliswaar superieur is, maar dat er geen reden is om een subtoets die uit drie items bestaat op psychometrische gronden als inadequaaf van de hand te wijzen.

6 Constructvaliditeit

6.1 Opzet

Betrouwbaarheid is niet het enige en evenmin het belangrijkste criterium aan de hand waarvan de kwaliteit van de 20 oorspronkelijke en verkorte voortoetsen moet worden afgemeten. Van groter gewicht voor die kwaliteit is de constructvaliditeit: meten de 20 instrumenten (i.c. de vragen, de oefeningen) die voor elke episode geconstrueerd zijn, ook specifiek datgene wat ze pretenderen te meten? Of meten die instrumenten wellicht een algemene vaardigheid zoals 'handigheid in vragen beantwoorden' of 'tekstgevoel'?

De constructvaliditeit is onderzocht via de convergente en de discriminante validiteit: hangen in conceptueel opzicht verwante items (vragen, oefeningen) met elkaar samen? Anders uitgedrukt: convergeren zij? En hangen zij onderling meer met elkaar samen dan met niet-verwante of minder verwante items (anders uitgedrukt: discrimineren zij)? Toegepast op onze toets: hangen bijvoorbeeld de items die de episode *Van COS naar tekstopzet* instrumentaliseren onderling samen en wel zodanig dat deze items onderscheiden kunnen worden van bijvoorbeeld de items die de episode *Presentatievorm kiezen* (pretenderen te) markeren?

Met het oog op de convergente en discriminante validiteit zijn via covariantiestructuur-analyse twee concurrerende modellen⁶ getoetst:

Model 1 Een model waarin vier latente, gecorrleerde factoren worden onderscheiden die corresponderen met de vier episodes binnen een

module; de vier respectievelijk drie items per episode correleren uitsluitend met die latente factor waarop ze geacht worden te ‘laden’.

Model 2 Een model met slechts één algemene latente factor waarop alle 16 (of 12) items een lading vertonen.

Model 2 is strijdig met de convergente en discriminante validiteit: immers, mocht dit model bij de gegevens passen, dan zou een empirisch en dus conceptueel onderscheid tussen de vier episodes per module noch mogelijk noch zinvol zijn. Model 1 echter is juist indicatief voor de validiteit van de toets.

6.2 Resultaten

Bij vier items per episode krijgt de deelnemer per module (4 x 4) 16 opgaven voorgesteld in de voortoets. Hieronder in de tabel een exemplarisch voorbeeld van de meest aannemelijke schatters in model 1 voor de module die het zwakst uit de betrouwbaarheidsmeting kwam: *Uitstraling en toon* (nulladingen zijn a priori op 0 gefixeerd). Links in de tabel de 16 items en horizontaal achtereenvolgens de vier latente factoren:

5.1 Formuleren met rechte rug (RUG)

5.2 Professioneel schrijven (PROF)

5.3 Formuleren met respect (RESP)

5.4 De lezer aanspreken (AANS)

Tabel 3 Parameterschattingen voor model 1 (module: *Uitstraling en toon*)

	RUG	PROF	RESP	AANS
Item 1	.386	.000	.000	.000
Item 2	.550	.000	.000	.000
Item 3	.473	.000	.000	.000
Item 4	.781	.000	.000	.000
Item 5	.000	.697	.000	.000
Item 6	.000	.603	.000	.000
Item 7	.000	.802	.000	.000
Item 8	.000	.655	.000	.000
Item 9	.000	.000	.458	.000
Item 10	.000	.000	.584	.000
Item 11	.000	.000	.695	.000
Item 12	.000	.000	.751	.000
Item 13	.000	.000	.000	.742
Item 14	.000	.000	.000	.614
Item 15	.000	.000	.000	.657
Item 16	.000	.000	.000	.675

De items 1 t/m 4 (*Formuleren met rechte rug*) vertonen hoge, significant van 0 afwijkende ladingen op de latente variabele die getypeerd kan worden als *Formuleren met rechte rug* terwijl ze niet correleren met de overige drie latente factoren die indicatief zijn voor de overige drie episodes. En dat geldt ook voor de resterende items: in model 1 zijn precies vier clusters te identificeren. In model 2 echter (waarvan we hier de parameterschattingen vanwege ruimtegebrek niet afdrukken) is dat niet het geval.

Eenzelfde beeld is te zien bij de modellen 1 en 2 voor episodes die gemarkeerd worden door drie items: ook daar past model 1 naadloos bij ons inhoudelijke voortoetsmodel.

Zou dit patroon toeval zijn en uitsluitend optreden bij de module *Uitstraling en toon*? De passingsstatistieken voor de modellen 1 en 2 bij drie en bij vier items zijn in Tabel 4 weergegeven voor de module *Efficiënte tekstopzet*. Blijkens de chi-kwadraatwaarden voor model 1 bij drie en bij vier items past model 1 goed, wat ook tot uiting komt in de corresponderende p-waarden. Als we dat vergelijken met model 2 voor drie en vier items, zien we een verre van rooskleurig beeld. Hetzelfde geldt voor de GFI- (Goodness of Fit) index.

Tabel 4 *Passingsstatistieken voor model 1 en 2 bij 3, resp. 4 items per episode (module: Efficiënte tekstopzet)*

<i>Passingsstatistieken voor model 1 en 2</i>			
<i>Voorbeeld: Efficiënte tekstopzet</i>			
Model/items	Chi-kw / df	P-waarde	GFI
Model 1 met 3 items	46.79 / 48	.522	.953
Model 1 met 4 items	105.45 / 98	.295	.921
Model 2 met 3 items	236.50 / 54	.000	.769
Model 2 met 4 items	508.72 / 104	.000	.642

Het beeld dat we zien bij de module *Efficiënte tekstopzet* blijkt gegeneraliseerd te kunnen worden naar alle modules: telkens past model 1 goed, zowel bij vier als bij drie items (GFI's tussen de .911 en .994) terwijl model 2 consequent verworpen moet worden (GFI's tussen .639 en .767). Concluderend kan gesteld worden dat de constructvaliditeit van de 20 episodes, onderzocht in termen van convergente en discriminante validiteit, bevredigend is, om het even of de episodes uit vier dan wel drie items bestaan.

7 Conclusie en discussie

De vraag of het mogelijk is de 20 subtoetsen uit de diagnostische voortoets bij de e-learning in te korten van vier naar drie items, en dat zonder al te groot verlies aan betrouwbaarheid en validiteit, kan op grond van de uitgevoerde psychometrische analyses positief beantwoord worden; daarmee kan - gedeeltelijk althans - ontsnapt worden aan de e-learningparadox. Reductie van het aantal items per subtoets van vier naar drie levert de 'modale' deelnemer een besparing op van ongeveer een uur: 80% van de deelnemers weet de 80-items toets in drie uur (of minder) af te ronden, terwijl eenzelfde percentage de 60-item toets in twee uur (of minder) afrondt. Gezien de jaarlijkse hoeveelheid deelnemers aan het trainingsprogramma *Managementletters schrijven* een relevante besparing voor de accountancy-maatschappen in Nederland.

Het opzetten van een e-learningtraining en de constructie van de bijbehorende voor- en natoets is arbeidsintensief, zeker als dat programma zo'n 300 oefeningen bevat met verschillende oefenvormen, met per oefening uitgebreide feedback en links vanuit iedere oefening naar de relevante pagina van een elektronisch cursusboek van ruim 100 pagina's A4.

Uit de psychometrische resultaten van dit onderzoek blijkt dat (sub)vaardigheden in het schrijven van managementletters met drie items zowel betrouwbaar als valide gemeten kunnen worden. Hoe zijn dergelijke resultaten mogelijk? En kunnen dergelijke resultaten ook gegeneraliseerd worden naar een willekeurig ander e-learningsschrijfprogramma?

Twee mogelijke verklaringen willen we noemen. De eerste, meer algemene verklaring is dat de taalbeheersing inmiddels zo veel heeft bijgedragen aan de profes-

sionalisering van schrijfopleidingen dat bij de ontwikkeling van onderwijsmateriaal en toetsen ‘als vanzelf’ een e-learningsschrijfprogramma ontstaat met optimale samenhang en coherentie. De tweede, meer specifieke verklaring luidt dat de ontwerpers van de onderhavige e-learning over een enorme hoeveelheid kennis beschikken op het gebied van het specifieke tekstgenre ‘managementletter’ en van de accountants die dit genre produceren. Generalisatie van onze resultaten is dus niet zonder meer gegarandeerd.

Noten

1. In de jaren 80 van de vorige eeuw bestond e-learning al, zij het onder andere namen (zoals CAI - computer aided instruction, of CBL - computer based learning; veel accountants noemen een e-learningprogramma daarom nog steeds: ‘een CBL-letje). Door internet heeft e-learning een stormachtige ontwikkeling doorgemaakt (Dobbs, 2002), juist ook omdat distributie en administratie via internet eenvoudig zijn geworden. E-learning - ook wel ‘webbased training’, ‘tele-leren’ of ‘on line’ leren genoemd (Rubens & Admiraal, 2003) - stelt de deelnemer via de computer in staat zich waar dan ook (thuis, op de werkplek of elders) en wanneer dan ook (overdag, in de avonduren of anderszins) de aangeboden leerstof eigen te maken in zijn eigen tempo. Kortom, e-learning zou het de cursist mogelijk maken om ‘self-paced’, ‘just in time’ en ‘just enough’ de aangeboden leerstof te verwerken en tot zich te nemen (Rosenberg, 2001; Kirkpatrick, 1998).
2. De misverstanden worden voor een deel veroorzaakt door de onduidelijkheid over de term ‘accountant’. Wij bedoelen in dit artikel met deze term *Registeraccountants* die de afkorting RA achter hun naam mogen voeren (in de USA ‘Certified Public Accountant’, CPA) en die zich bij de grote maatschappen vooral bezighouden met de controle van de jaarrekening van grote en middelgrote organisaties. De beschermde RA-titel kan behaald worden via een postacademiale opleiding, bijvoorbeeld na een studie economie.
3. COS (= Nadere Voorschriften Controle en Overige Standaarden) verwijst naar regelgeving betreffende de accountantscontrole, zoals opgesteld door de NIVRA (= Koninklijk Nederlands Instituut voor Registeraccountants).
4. Voor de berekening van de betrouwbaarheid van de 20 onderscheiden episodes vormde de specifieke aard van de datamatrix een netelig probleem: voor elke episode zijn er aselect vier items uit een itempool van zeven getrokken, zonder teruglegging, wat tot gevolg heeft dat verschillende deelnemers per episode een ander scorepatroon vertonen c.q. kunnen vertonen: zo heeft de ene deelnemer bijvoorbeeld item 1, 2, 3 en 4 beantwoord, een tweede de items 2, 4, 6 en 7, een derde deelnemer de items 1, 4, 5 en 7, enzovoort. In de datamatrix komen dus structureel, zij het op willekeurige wijze, ‘missing values’ voor. Omdat betrouwbaarheidsberekening uitgaat van een correlatiematrix tussen de items die gezamenlijk een schaal constitueren, en omdat een dergelijke correlatiematrix alleen berekend kan worden voor zover de data van een deelnemer compleet zijn, moeten die structurele missing values geschat worden. De door ons toegepaste schattingstechniek behelst multiple regressie-analyses; daarbij worden de ontbrekende scores op een willekeurig gekozen item (een van de zeven items per episode) geschat met behulp van de scores op de resterende zes items (i.c. de predictoren). Na herhaalde toepassing van deze schattingstechniek beschikt elke proefpersoon uiteindelijk over precies hetzelfde aantal scores: 140 (20 episodes maal zeven). Van die 140 scores zijn er 80 realistisch (dat wil zeggen op de empirische werkelijkheid gebaseerd) en 60 geschat. Schattingen – het is bekend – kunnen systematische fouten bevatten die tot contaminatie van conclusies kunnen leiden; om dat uit te sluiten, worden de betrouwbaarheidsanalyses uitgevoerd op een variabel aantal items per episode: op 7, 6, 5, 4, 3 en 2 items. Indien er per episode systematische trends bij dit variërend aantal items aangetroffen kunnen worden en indien die trends generaliseerbaar zijn over de episodes, dan mag men er vanuit gaan dat de onvermijdelijke ruis die aanwezig is

in de schattingen niet van dien aard is dat daardoor consistente patronen in de data versluierd raken. Overigens, aan het schatten van scores hangt een 'prijskaartje': geschatte scores hebben de neiging te regresseren naar het gemiddelde, zodat de scorevariantie afneemt. Afname van de scorevariantie leidt doorgaans tot afname in de hoogte van de correlaties, wat een negatief effect heeft op de hoogte van de betrouwbaarheid.

5. De interne consistentie maat Cronbach's alpha levert waarschijnlijk een *overschatting* op van 'de' betrouwbaarheid, gezien de functie van de voortoets: vaststellen of een deelnemer een (sub)vaardigheid in voldoende mate beheerst. De interesse gaat dus uit naar de absolute prestatie van een deelnemer en niet naar diens relatieve positie ten opzichte van andere kandidaten. Cronbach's alpha levert echter de betrouwbaarheid (of de ondergrens daarvan) voor de relatieve prestatie van de kandidaat, niet de betrouwbaarheid voor de absolute prestatie; de betrouwbaarheidsschatting voor een absolute prestatie is doorgaans lager dan die voor een relatieve prestatie.

6. Het uitgangspunt van de validiteitsanalyses wordt gevormd door de correlatiematrix van de 16 items die gezamenlijk één module markeren (in het geval wanneer elk van de constituerende episodes vier items bevatten) en door de, daarmee corresponderende, correlatiematrix van de 12 items (in het geval wanneer elk van de episodes uit drie items bestaat). De structuur van de betreffende correlatiematrix zal telkens, voor elke module afzonderlijk, getoetst worden via covariantiestructuur-analyse (Jöreskog & Sörbom, 1989). Onder de aanname dat de items c.q. variabelen multivariaat-normaal verdeeld zijn, kan een toetsingsgrootte worden berekend die bij benadering chi-kwadraat verdeeld is: getoetst wordt het gespecificeerde model (als nulhypothese) tegen het alternatief dat de geanalyseerde matrix een willekeurige positief definitieve matrix is (Jöreskog, 1974).

Literatuur

- Dobbs, K. (2002). The State of Online Learning – What the Online World Needs Now: Quality. In Rosset, A. (Ed.) *The ASTD E-learning Handbook* (pp. 357-364). New-York: McGraw-Hill.
- Jöreskog, K.G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In D.H. Krantz, R.C. Atkinson, R.D. Luce & P. Suppes (eds.) *Contemporary Developments in Mathematical Psychology. Vol. II: Measurement, Psychophysics, and Neural Information Processing* (pp. 1-56). San Francisco: Freeman.
- Jöreskog, K.G. & Sörbom, D. (1989). *LISREL 7 User's Reference Guide*. Mooresville: Scientific Software Inc.
- Kirkpatrick, D.L. (1998). *Evaluating Training Programs, the Four Levels*. 2nd ed. San Francisco: Berrett-Koehler Publishers.
- Rosenberg, M. J. (2001). *E-learning. Strategies for Delivering Knowledge in the Digital Age*. New York: McGraw-Hill.
- Rubens, W. & Admiraal, W. (2003). Samenwerkend leren met behulp van ICT binnen het Europese onderwijs: ervaringen met Synergiea in vier landen. Paper gepresenteerd tijdens de OnderwijsResearchDagen van 7 t/m 9 mei 2003, te Kerkrade.

Maat & Meuffels