

Peerassessment in universitair onderwijs. Een onderzoek naar bruikbare ontwerpen

Dr. B.A.M. van den Berg, dr. W. Admiraal en prof. dr. A. Pilot zijn werkzaam bij het Interfacultair Instituut voor de Lerarenopleiding, Onderwijsontwikkeling en Studievaardigheden (IVLOS) van de Universiteit Utrecht.

Peerassessment wordt in dit onderzoek opgevat als een vorm van samenwerken tussen studenten waarin ze elkaars werk tussentijds beoordelen en van commentaar voorzien. In dit ontwerponderzoek zijn zeven verschillende ontwerpen van peerassessment beproefd om tot een optimaal model te komen. Onderzocht wordt de uitvoering van het ontwerp, de samenstelling van de peerfeedback, de interactie tussen studenten, de leeropbrengst van peerassessment en de waardering van studenten en docenten. De ontwerpen zijn uitgevoerd binnen de opleiding Geschiedenis van de Universiteit Utrecht, daarbij waren 168 studenten en negen docenten betrokken. De studenten namen de beoordelingsstaak serieus en gebruikten de peerfeedback om hun werk te verbeteren. Studenten geven elkaar vollediger feedback als ze hun schriftelijke beoordeling mondeling aan elkaar toelichten.

INLEIDING

Leren schrijven op academisch niveau is een integraal onderdeel van universitair onderwijs. Leren schrijven op een aanvaardbaar academisch niveau gaat hand in hand met een diepere beheersing van de disciplinaire kennis. Het gaat hier om een vaardigheid die niet in het bestek van één cursus aan te leren is, maar die in alle fasen van het curriculum ondersteund moet worden. Veel opleidingen zoeken naar effectieve begeleidingsvormen bij dit leerproces, al was het maar omdat de toenemende aandacht voor vaardigheden in het curriculum veel docenttijd kost. De gedachte om medestudenten in te schakelen als medebeoordelaars en feedbackgevers wint aan populariteit. In literatuur over didactiek in het hoger onderwijs vinden we redenen om studenten bij de ondersteuning van het leerproces te betrekken, met name in de rol van feedbackgevers. Dit onderzoek richt zich op de inzet van medestudenten bij de begeleiding van de schrijfvaardigheid. Het is uitgevoerd binnen de opleiding Geschiedenis van de Universiteit Utrecht.

THEORETISCH KADER

Gallimore en Tharp, die de socioculturele theorie van Vygotsky uitwerkten voor het onderwijs (Tharp & Gallimore, 1988; Gallimore en Tharp, 1990), omschrijven onderwijzen als *assisted performance*. Zij onderscheiden zes manieren waarop die begeleiding gestalte kan krijgen, waaronder begeleiding in de vorm van feedback. Feedback staat bekend als een onmisbaar instrument voor het aanleren van complexe vaardigheden, zoals het kunnen schrijven op academisch niveau.

Zowel aan Vygotsky als aan Piaget zijn argumenten te ontleen om medestudenten in te schakelen als feedbackgevers. Volgens Neo-Piagetanen, weergegeven in Tudge (1990) biedt het contact met leeftijdgenoten aan het individu bij uitstek de gelegenheid om ervaringen van cognitieve dissonantie op te doen. Leren berust op *equilibratie*, aldus Piaget. Het individu, geconfronteerd met opvattingen die strijdig zijn met die van hemzelf, voelt zich gedwongen om zijn denken bij te stellen zodat er een nieuw evenwicht ontstaat. Neo-Vygotskianen beargumenteren de waarde van het contact met leeftijdgenoten vanuit de stimulans die het geeft om elkaar te begrijpen, meningsverschillen te overbruggen en zo naar een met anderen gedeeld begrip toe te groeien (Tharp & Gallimore, 1988). Tudge (1990) vergeleek bevindingen van Piagetaans met die uit Vygotskiaans onderzoek en komt tot de conclusie dat, of men nu redeneert vanuit Piaget of vanuit Vygotsky, de cognitieve groei het meest gestimuleerd wordt als de lerenden iets maar niet veel verschillen in capaciteiten.

In een reviewstudie van Topping (1998) worden positieve effecten gerapporteerd van beoordeling en feedback door medestudenten, oftewel *peerassessment*, bij schrijfonderwijs. Peerassessment wordt hierbij opgevat als 'an arrangement in which individuals consider the amount, level, value, worth, quality, or success of the products or outcomes of the learning of peers of similar status' (Topping, 1998).

Om studenten in te zetten als feedbackgevers bij het leren schrijven, moeten onderwijsontwerpers inzicht krijgen in het schrijfproces, zodat duidelijk is op welke momenten feedback het meest nodig is. Volgens Flower, Hayes, Carey, Schriver en Stratman (1986) zijn vooral de planfase en de revisiefase zinvolle momenten voor feedback door medestudenten. Saunders (1989) geeft een overzicht van de verschillende manieren waarop studenten elkaar kunnen begeleiden bij het leren schrijven en de daarbij te verwachten interactie, die hun weerslag hebben op de effectiviteit van de samenwerking. Eén van de vormen van samenwerkend schrijven die hij noemt is *co-responding*, waarbij studenten elkaar ondersteunen bij het reviseren. Deze vorm van begeleiding is wat wij beogen met peerassessment bij leren schrijven. De vraag is nu hoe de leersituatie op een manier kan worden ingericht die studenten ertoe aanzet om serieus commentaar te geven op een schrijfproduct van medestudenten en van elkaar te leren. Bestaand onderzoek met betrekking tot peerassessment geeft, aldus Topping, weinig inzicht in de factoren waarmee rekening gehouden moet worden bij de praktische vormgeving ervan. Hij vond in totaal 17 variabelen in de organisatie van peerassessment. Hierbij gaat het om keuzes die gemaakt moeten worden met betrekking tot vragen als: wie beoordeelt wie, is de beoordeling anoniem, gaat het alleen om een cijfer of ook/alleen om feedback?

Om meer zicht te krijgen op factoren die van belang zijn bij het organiseren van peer-assessment in een cursus startten we dit ontwerponderzoek, dat gericht is op het vinden van een optimaal model van peerassessment. In totaal zijn zeven ontwerpen van peerassessment gemaakt. Deze zijn geëvalueerd op de volgende punten:

- de uitvoering van de assessmentprocedure;
- de samenstelling van de feedback;
- de interactie tijdens de mondelinge toelichting;
- de leeropbrengst en waardering van het assessment.

ZEVEN ONTWERPEN VAN PEERASSESSMENT

Deze studie omvat de ontwikkeling, implementatie en evaluatie van zeven ontwerpen van peerassessment, met de bedoeling om een optimaal model te destilleren. De zeven ontwerpen zijn toegesneden op zeven cursussen. De selectie daarvan geschiedde op grond van de volgende argumenten. Het leek de onderzoekers wenselijk om in het hele curriculum ervaringen met peerassessment op te doen, zodat duidelijk kon worden of studenten van alle opleidingsjaren in staat zouden zijn om met peerassessment te werken. Daarbij zijn in elk geval de cursussen betrokken die een centrale rol hadden bij de schrijfontwikkeling. Om die reden zijn ontwerpen gemaakt voor de:

- eerstejaarscursus Scriptie Schrijven (SS);
- tweedejaarscursussen Historisch Ambacht (HA) en Doctoraalseminar (DS);
- de in het derde/vierde opleidingsjaar geprogrammeerde specialisatiecursus Rapportagepracticum (RP).

	Variabele	Variatiebreedte
1	Vakgebied/onderwerp	Range van opleidingen, thema's
2	Gebruiksdoel	Docenttijdsbesparing of leermiddel
3	Focus	Kwantitatief/summatief, kwalitatief/formatief
4	Product	Schriftelijk werk, vaardigheidsdemonstratie
5	Relatie docentbeoordeling	Vervangend of aanvullend
6	Gewicht van het oordeel	Bijdrage aan het eindcijfer
7	Oordeelsrichting	Eenzijdig of wederzijds
8	Privacy	Anoniem, vertrouwelijk of openbaar
9	Contact	In directe aanwezigheid of op afstand
10	Jaar	Zelfde of verschillend opleidingsjaar
11	Bekwaamheid	Vergelijkbare of verschillende bekwaamheden
12	Constellatie beoordelaars	Individueel, tweetal, groep
13	Constellatie beoordeelden	Individueel, tweetal, groep
14	Plaats	In of buiten contacttijd
15	Tijd	Lestijd/vrije tijd
16	Deelnameverplichting	Verplicht of facultatief
17	Beloning	Cijfer, studiepunt of andere extrinsieke beloning

Figuur 1 Typologie peerassessment (Topping, 1998)

Noot Variabelen waarop de ontwerpen verschillen, zijn met cursief aangegeven.

	SS	HA	DS
(4)product	Conceptwerkstuk (10 pp.)	Conceptopzet voor werkstuk (1-2 pp) en conceptversie kernhoofdstuk (3-5 pp)	Conceptwerkstuk (15 pp.)
(5)relatie docentbeoordeling	Aanvullend; docent geeft schriftelijke feedback op conceptversie	Aanvullend; docent geeft mondelinge feedback op concepten	Aanvullend; docent geeft schriftelijke feedback + cijfer voor conceptversie
(7)oordeelsrichting	Eén richting (2 beoordelingen)	Wederzijds (2 of 3 beoordelingen van beide producten)	Wederzijds (1 beoordeling)
(8)privacy	Openbaar (voor de docent en alle studenten)	Vertrouwelijk (binnen de groepen); de docent ontvangt een kopie	Vertrouwelijk (binnen de groepen); de docent ontvangt een kopie
(9)contact	Schriftelijke en mondelinge feedback; plenaire nabespreking met aanvullende feedback van docent	Schriftelijke en mondelinge feedback	Schriftelijke en mondelinge feedback; plenaire bespreking van discussiepunten
(11)bekwaamheid	Groepssamenstelling at random	Groepssamenstelling op basis van verwant onderwerp (door docent)	Groepssamenstelling op basis van verwant onderwerp (door studenten)
(12)constellatie beoordelaars	2 studenten + docent	Kleine groepen (3-4 studenten) met docent als participant	2 studenten + docent
(13)constellatie beoordeelden	2 andere studenten	Dezelfde kleine groepen	Dezelfde 2 studenten
(14)plaats	Schriftelijke feedback buiten /mondelijke feedback in contacttijd (plenaire bespreking en tutorial met docent)	Schriftelijke feedback buiten /mondelijke feedback in contacttijd (kleine groepen met docent)	Schriftelijke feedback buiten /mondelijke feedback in contacttijd (kleine groepen en plenaire bespreking)
(17)beloning	Geen beloning PA deelname	Geen beloning PA deelname	Geen beloning PA deelname

Figuur 2 Ontwerpkenmerken van zeven ontwerpen van peerassessment

HIST	RP	ICE	TGM
Conceptbiografie (10 pp.)	Conceptkrantenartikel (5 pp.)	Conceptanalyse van een tentoonstelling (3-5 pp)	Conceptartikel (5 pp)
Aanvullend; extra beoordeling	Aanvullend; docent geeft schriftelijke feedback + cijfer voor conceptversie	Aanvullend; extra beoordeling	Aanvullend; extra beoordeling
Wederzijds (2 gezamenlijke beoordelingen)	Wederzijds (2 beoordelingen)	Wederzijds (3 beoordelingen)	Eén richting (2)
Vertrouwelijk (binnen de groepen); de docent ontvangt een kopie	Vertrouwelijk (binnen de groepen); de docent ontvangt een kopie en beoordeelt	Vertrouwelijk (binnen de groepen); de docent ontvangt een kopie	Vertrouwelijk; de docent ontvangt kopie
Schriftelijke en mondelinge feedback; plenaire bespreking van discussiepunten	Schriftelijke en mondelinge feedback; plenaire bespreking van discussiepunten	Schriftelijke en mondelinge feedback	Schriftelijke feedback
Groepssamenstelling at random	Groepssamenstelling at random; zelfde onderwerp	Groepssamenstelling gebaseerd op tentoonstelling (door studenten)	Groepssamenstelling at random
Kleine groepen bestaande uit 2 duo's (studenten)	Kleine groepen (3 studenten) + docent	Kleine groepen (4 studenten)	2 studenten
Dezelfde kleine groepen	Dezelfde kleine groepen	Dezelfde kleine groepen	2 andere studenten
Schriftelijke feedback buiten /mondelijke feedback in contacttijd (kleine groepen en plenaire bespreking)	Schriftelijke feedback buiten / mondelinge feedback in contacttijd (kleine groepen en plenaire bespreking)	Schriftelijke feedback buiten / mondelinge feedback in contacttijd (kleine groepen)	Schriftelijke feedback buiten contacttijd
Geen beloning PA deelname	Uitgebracht schriftelijk oordeel beloond met max. $\frac{1}{4}$ punt extra	Geen beloning PA deelname	Geen beloning PA deelname

Ook het feit dat sommige cursussen door dezelfde docent in parallelgroepen werden gegeven was een reden voor selectie, daar dit de mogelijkheid gaf om met controle-groepen te werken. Dit was het geval bij de:

- derdejaarscursus Historiografie (HIST);
- de specialisatiecursussen Inleiding Cultuureducatie (ICE) en Cultuureducatie: Theorie, Geschiedenis en Media (TGM).

Ontwerpvariabelen

De verschillen tussen de ontwerpen hebben betrekking op tien van de 17 variabelen uit de typologie van Topping, die gepresenteerd wordt in figuur 1.

Op een aantal variabelen is niet gevarieerd, deels om onderwijskundige redenen (variabelen 2, 3, 10), deels omdat alle ontwerpen inpasbaar moesten zijn in het lopende programma (1, 15, 16) of omdat de docenten het een te groot risico vonden (6). De zeven ontwerpen worden samengevat in figuur 2.

Wat betreft de groep variabelen die te maken heeft met de functie van peerassessment als beoordelingsinstrument (1 t/m 6), verschillen de ontwerpen qua omvang en stadium van het te beoordelen product en de relatie met de docentbeoordeling. Bij vier ontwerpen (SS, HA, DS en RP) vormt peerassessment een aanvulling op de (voor dit doel globaal gehouden) docentbeoordeling van de conceptversie; in twee hiervan geeft de docent naast feedback ook een cijfer voor de conceptversie (DS en RP). In de andere drie ontwerpen (ICE, TGM en HIST) is peerassessment niet alleen aanvullend, maar vormt het ook een extra tussentijds beoordelingsmoment. Zonder welke de studenten geen feedback op hun conceptversie zouden hebben ontvangen. Met betrekking tot de variabelen interactie tussen de deelnemers (7 t/m 9) zitten de verschillen in de al of niet wederzijdse beoordeling, de mate van vertrouwelijkheid en de wijze van feedback geven: alleen schriftelijk of schriftelijk én mondeling. Voor wat betreft de groepsamenstelling (10 t/m 13) verschillen de ontwerpen qua omvang van de feedbackgroepjes; daarnaast wordt er in één ontwerp (HIST) door duo's beoordeeld die in overleg tot één gezamenlijk oordeel moeten zien te komen. Een ander verschil is de rol van de docent: in drie ontwerpen (HIST, ICE en TGM) is de docent geen medefeedbackgever, in de vier andere wel. Ten aanzien van de groep variabelen die te maken hebben met verplichting en beloning (14 t/m 17) is er in een ontwerp sprake van beloning van de inspanningen voor peerassessment omdat de kwaliteit van de uitgebrachte beoordeling meetelt in het eindcijfer (ontwerp RP). Behalve deze variabelen distilleert Topping nog een aantal aanwijzingen die betrekking hebben op de implementatie van peerassessment, zoals het vooraf expliciteren van doelen en verwachtingen van peerassessment, het verhelderen van de beoordelingscriteria en het oefenen in het geven en ontvangen van feedback. Ook onderstreept hij het belang van een goede procesbewaking tijdens de uitvoering. Deze aanwijzingen hebben we zo veel mogelijk opgevolgd.

Peerassessment-procedure

In Bean (1998) vonden we een transparante basisprocedure voor peerassessment, die goed in te passen was in de bestaande cursusopzet. Deze procedure is gebruikt in alle ontwerpen en werkt als volgt.

De studenten werken aan hun schrijfproduct, op een bepaald moment onderbreken ze dit om elkaars tussenproducten te beoordelen. Daarvoor gebruiken ze een beoordelingsformulier met dezelfde beoordelingscriteria als de docent gebruikt bij de beoordeling van de eindversie. De studenten vullen het formulier in, lichten per beoordelingsaspect hun oordeel toe en vatten hun bevindingen samen in de vorm van minimaal drie adviezen voor verbetering. Dit verslag geven ze aan degene wiens werk ze hebben beoordeeld. Daarna krijgt ieder de gelegenheid om het schrijfproduct te reviseren, alvorens het ter beoordeling voor te leggen aan de docent.

De docent heeft hierbij de rol van procesbewaker. Hij ontvangt tegelijk met de studenten een kopie van de producten die de studenten elkaar voorleggen en van de ingevulde beoordelingsformulieren, zodat hij kan zien of ieder de gemaakte afspraken nakomt. In de fase dat de studenten elkaar feedback geven, onthoudt de docent zich zo veel mogelijk van commentaar op de schrijfproducten.

Implementatie

Bij de implementatie is waar mogelijk rekening gehouden met Topping's aanbevelingen om de doelen en verwachtingen van peerassessment te expliciteren en de beoordelingscriteria te verhelderen. De studenten kregen op de eerste cursusbijeenkomst informatie over de doelen van peerassessment en de werkwijze, met de gelegenheid om er vragen over te stellen. De beoordelingsformulieren en de daarin opgenomen criteria werden door de docent toegelicht. Deze informatie werd vlak voor het moment waarop de studenten met de procedure startten, nog eens gedetailleerd op schrift gegeven. Studenten zijn voorafgaand aan de betreffende cursus niet getraind in het beoordelen van schriftelijke producten.

1	Hoeveel tijd heb je besteed aan het schriftelijk beoordelen van de 2 conceptbiografieën van je medestudenten?
1a	a) het <u>lezen</u> van de eerste versie van biografie 1 kostte mijmin/uren
1b	b) het <u>lezen</u> van de eerste versie van biografie 2 kostte mijmin/uren
1c	c) het opstellen van een gezamenlijk advies kostte ons ... min./uren
2	Wat vond je van deze tijdsinvestering? <i>Denk hierbij aan bv. het nut ervan, de hoeveelheid tijd die je eraan besteedde.</i>
3	Wat vond je van het commentaar dat je kreeg op jouw conceptversie van je medestudenten? Denk hierbij bv. aan: was het zinnig, 'to the point', stimulerend? Maak hierbij onderscheid tussen:
3a	a) het commentaar in de ingevulde beoordelingsformulieren
3b	b) de aantekeningen die ze maakten bij jouw tekst
3c	c) het mondeling gegeven commentaar daarna in de werkgroep
4	In hoeverre denk je dat je werkstuk beter zal worden als gevolg van het schriftelijke en mondelinge commentaar van je medestudenten?
5	Aan welk soort commentaar van je medestudenten had je het meest en aan welk het minst?
6	In hoeverre achtte je jezelf in staat om het werk van een medestudent te beoordelen?
7	In hoeverre had het feit dat je wist van wie het werk was dat je moest beoordelen invloed op je beoordeling?

8	Wat vond je van de opdracht om gezamenlijk met je duopartner tot 1 advies te komen?
9	Wat vond je van de beoordelingsformulieren?
10	Was de procedure m.b.t. peerassessment je voldoende duidelijk?
11	Ben je tevreden of ontevreden over je eigen inzet voor de werkgroep?
12	Ben je tevreden of ontevreden over je eigen inzet voor peerassessment?
13	Wat vond je van de inzet van je medestudenten tijdens de werkgroep tot nu toe?
14	Wat vond je van de inzet van je partners voor peerassessment?
15	Zijn er nog andere dingen die je kwijt wil over peerassessment in deze cursus?
16	Zou je vaker met peerassessment willen werken. Zo ja, op dezelfde manier of anders?
17	Wil je nog iets opmerken over deze cursus of heb je suggesties?

Figuur 3 Vragenlijst peerassessment

Noot. Voorbeeld, gebruikt bij het ontwerp Historiografie

6	Hoeveel uur heb je besteed aan het inlezen voor de biografie? A 0-10, B 11-20, C 21-30, D 31-40, E >41
7	Hoeveel uur heb je besteed aan het schrijven van de biografie? A 0-10, B 11-20, C 21-30, D 31-40, E >41
8	Vanaf welke cursusweek in blok 2 ging je intensief voor deze cursus aan de slag? A wk1-2 B wk3-4 C wk5-6 D wk7-8 E nooit
16	Over de inhoudelijke interactie met de docent tijdens de werkcolleges ben ik A zeer tevreden...E zeer ontevreden
17	Over de inhoudelijke interactie met mijn medestudenten tijdens de werkcolleges ben ik A zeer tevreden...E zeer ontevreden
21	Door dit deel van de cursus in blok 2 heb ik beter geleerd om een werkstuk te schrijven. A zeer mee eens...E zeer mee oneens
22	Door dit deel van de cursus in blok 2 heb ik meer zicht gekregen op de criteria waaraan een werkstuk moet voldoen. A zeer mee eens...E zeer mee oneens

Figuur 4 Cursusvragenlijst (aan cursusevaluatie toegevoegde vragen over peerassessment)

Noot. Voorbeeld, gebruikt in de cursus Historiografie.

METHODEN

Dit ontwerponderzoek kan worden opgevat als een multiple casestudy (zie Creswell, 1998). In dit geval worden de cases gevormd door de zeven cursussen, die bestudeerd worden op elementen die te maken hebben met peerassessment. De cases zijn zowel elk afzonderlijk als via crosscase analyse geëvalueerd om een optimaal ontwerp te kunnen destilleren.

De onderzochte groep omvatte 168 studenten die deelnamen aan onderwijs van de opleiding Geschiedenis van de Universiteit Utrecht. Op het moment dat zij zich voor een cursus inschreven, wisten de studenten niet in welke groepen met peerassessment

zou worden gewerkt. Op een enkeling na, had niemand ooit eerder met peerassessment gewerkt, althans niet in deze gestructureerde vorm. Van de totale onderzoeksgroep deden 131 studenten mee in groepen waarin gewerkt werd met peerassessment, de overige 37 maakten deel uit van een controlegroep. De controlegroepen werden gevormd door parallelwerkgroepen van eenzelfde cursus, waarin niet met peerassessment gewerkt werd en waarin het onderwijs en de toetsing door dezelfde docent werd verzorgd als in de peerassessment-groepen. Er waren negen docenten bij het onderzoek betrokken, te weten alle docenten die een of meer peerassessment-groepen begeleidden. Ook zij hadden nooit met peerassessment gewerkt.

	Verhelderen	Beoordelen	Toelichten	Reviseren	Inbreng Fn
Evaluatief	-	+	-	-	-
Ontdekkend	+	-	+	+	+

Figuur 5 Typologie van feedbackposities (naar Lockhart & Ng, 1995)

Noot. Bij de bepaling of een bepaalde feedbackfunctie veel of weinig wordt uitgeoefend, zijn de volgende criteria gehanteerd:

- bij Evaluatief + = 50% vd productbesprekingen of meer; - = 15% of minder;
- bij Ontdekkend + = 30% vd productbesprekingen of meer; - = 10% of minder;
- bij Fn (=feedbacknemer) + = evenredige inbreng, - = minder dan evenredige inbreng.

De gegevens over de uitvoering, leeropbrengst en waardering zijn deels verkregen door middel van observatie, twee schriftelijke vragenlijsten voor studenten en een interview met elke docent na afloop van de cursus. Daarnaast zijn de conceptversies met alle feedback en de definitieve versies van de schrijfproducten verzameld. Onder leeropbrengst wordt hier verstaan de leerwinst aan het einde van de cursus als gevolg van het werken met peerassessment. Deze wordt beschreven in termen van de ondernomen revisie, de behaalde cijferwaardering en de door de studenten ingeschatte leeropbrengst. Om zicht te krijgen op het type revisie en de mate waarin de peerfeedback door de studenten is verwerkt zijn de versies van de schrijfproducten voor en na peerassessment met elkaar vergeleken. Daarbij zijn de verschillen tussen beide versies naast de ontvangen peerfeedback en de eventuele docentfeedback gelegd.

Tabel 1 Uitvoering van de assessmentprocedure

SS	HA	DS	HIST	RP	ICE	TGM
88% op tijd met conceptversie, daarvan 80% volledige PA*	80% op tijd met opzet, daarvan 63% volledige PA; 75% op tijd met kernhoofdstuk, daarvan 47% volledige PA	67% op tijd met conceptversie, daarvan 100% volledige PA	65% op tijd met conceptversie, daarvan 31% volledige PA (in duo's); soms onvolledige conceptversie	100% op tijd met conceptversie, daarvan 69% volledige PA	91% op tijd met conceptversie, daarvan 75% volledige PA	100% niet op tijd met conceptversie (uiteindelijk wel ingeleverd), daarvan 50% volledige PA

*Noot. Volledige peerassessment, d.w.z. beoordeeld door alle leden van de feedbackgroep

Tabel 2 Verdeling van de feedbackfunctie in de schriftelijke peerfeedback (in % van het totale aantal feedbackuitspraken per ontwerp)

	SS	HA-1	HA-2	DS	Hist1	Hist2	RP	ICE	TGM	Gem
VERH*	20	25	12	19	22	17	24	24	2	19 (22**)
BEOORD	51	50	45	46	41	46	36	37	46	44 (47)
TOEL	16	14	32	20	16	19	17	9	27	18 (19)
REVIS	10	7	4	10	13	13	16	16	7	11 (12)
ORIENT	0	0	0	1	0	0	0	1	0	0
WERKW	1	2	1	0	0	0	0	2	0	1
FBnvt	1	2	6	4	8	4	7	11	17	6
Tot.	100	100	100	100	100	100	100	100	100	100 (100)

Noot. * de productgerichte feedbackfuncties zijn vetgedrukt;

** percentage van de productgerichte feedback;

FBnvt = uitspraken zonder feedbackfunctie.

Om de waardering van studenten voor peerassessment te achterhalen is een open vragenlijst afgenomen direct na de mondelinge toelichting en voordat de studenten het cijfer voor de gereviseerde versie van hun schrijfproduct hadden vernomen. Deze 'vragenlijst over peerassessment' bestond uit een basislijst van 16 open vragen die in alle peerassessment-groepen zijn gesteld. Deze lijst is per cursus aangevuld met een of twee vragen over specifieke elementen in het ontwerp. In deze lijst wordt gevraagd naar de hoeveelheid tijd die het lezen en beoordelen van andermans werk had gekost, de waardering van respectievelijk deze tijdsinvestering, de ontvangen feedback van medestudenten, de bruikbaarheid van het beoordelingsformulier, de eigen inzet en die van de medestudenten, de inschatting van de eigen beoordelingsbekwaamheid en het nut van de gekregen peerfeedback voor de revisie. Tot slot werd de studenten gevraagd of zij vaker met peerassessment wilden werken en hoe de methode eventueel verbeterd zou kunnen worden. Een voorbeeld is opgenomen als figuur 3, vraag 8 is ontwerp-specifiek en alleen in het ontwerp HIST gesteld.

De antwoorden op de vragen betreffende de waardering van de tijdsinvestering (vraag 2), van de schriftelijke, respectievelijk mondelinge feedback (vraag 3), van het beoordelingsformulier (vraag 9), van de inzet (vraag 12 en 13) en ten slotte van het vaker werken met peerassessment (vraag 16) zijn door de onderzoekers omgezet in een waardering op een driepuntsschaal, hiertoe is de interbeoordelaarsbetrouwbaarheid berekend tussen twee observatoren, die bleek voldoende. Daarnaast is bij alle studenten, zowel in de peerassessment-groepen als in de controlegroep een 'cursusvragenlijst' afgenomen, waarin onder meer gevraagd is naar de eigen inschatting van de schrijfvaardigheid. De antwoorden zijn gegeven in de vorm van scores op een vijfpuntsschaal. Deze lijst werd door de studenten ingevuld direct na afloop van het tentamen. Een voorbeeld is opgenomen als figuur 4.

Tabel 3 Verdeling in percentages van de feedbackfunctie in de mondelinge peerfeedback (in % van het totale aantal feedbackuitspraken per ontwerp)

	SS	HA-1	HA-2	DS	Hist1	Hist2	RP	ICE	Gem
VERH*	13	12	15	13	13	14	16	13	14 (23**)
BEOORD	34	12	19	19	21	22	24	26	20 (32)
TOEL	27	12	16	20	17	15	19	16	17 (26)
REVIS	3	13	9	12	5	14	11	10	11 (17)
ORIENT	14	15	10	10	16	12	9	7	11
WERKW	0	12	7	3	6	2	1	1	5
FBnvt	9	24	25	22	21	21	19	26	22
Tot.	100	100	100	100	100	100	100	100	100(100)

Noot * de productgerichte feedbackfuncties zijn vetgedrukt;

** percentage van de productgerichte feedback;

TGM is hier buiten beschouwing gelaten, omdat hierin alleen schriftelijke feedback is gegeven;

FBnvt = uitspraken zonder feedbackfunctie.

Tabel 4 Verdeling in percentages van het feedbackaspect in de schriftelijke peerfeedback (in % van het totale aantal feedbackuitspraken per ontwerp)

	SS	HA-1	HA-2	DS	Hist1	Hist2	RP	ICE	TGM	Gem
INH*	43	26	34	32	49	43	40	46	30	40 (43**)
STRUCT	13	21	18	21	7	13	14	0	16	12 (13)
STIJL	34	37	28	35	24	24	28	28	28	30 (33)
COMBI	7	11	12	6	12	17	7	12	10	9 (10)
NVT	4	4	8	6	8	3	10	13	15	8
Tot.	100	100	100	100	100	100	100	100	100	100 (100)

Noot * De productgerichte feedbackaspecten zijn vetgedrukt;

** percentage van de productgerichte feedback;

COMBI = feedback die op meerdere aspecten tegelijk betrekking heeft;

NVT = uitspraken die noch over de tekst, noch over de procedure gaan.

Aan het eind van de cursus is de docent bevraagd in de vorm van een halfgestructureerd interview. Dit vond plaats zo snel mogelijk nadat de eindversies van de schrijfproducten waren beoordeeld.

Tabel 5 Verdeling in percentages van het feedbackaspect in de mondelinge peerfeedback (in % van het totale aantal feedbackuitspraken per ontwerp)

	SS	HA-1	HA-2	DS	Hist1	Hist2	RP	ICE	Gem
INH*	38	25	32	23	25	29	23	23	27(45**)
STRUCT	5	6	6	2	3	3	2	0	4 (6)
STIJL	22	15	15	32	11	20	31	25	21(35)
COMBI	14	6	5	5	17	10	8	14	9 (14)
PROC	1	1	1	2	4	2	1	1	2
NVT	21	46	41	36	39	36	35	37	38
Tot.	100	100	100	100	100	100	100	100	100 (100)

Noot * De productgerichte (= tekstuele) feedbackaspecten zijn vetgedrukt;

** percentage van de productgerichte feedback;

TGM is hier buiten beschouwing gelaten, omdat hierin alleen schriftelijke feedback is gegeven;

COMBI = feedback die op meerdere aspecten tegelijk betrekking heeft;

PROC = uitspraken over de assessmentprocedure;

NVT = overige uitspraken die niet over de tekst gaan.

De vragen in dit interview bevatten ongeveer dezelfde items als de 'vragenlijst over peerassessment'. Evenals aan de studenten zijn ook aan de docent vragen gesteld in verband met zijn tijdsinvestering voor de beoordeling, eventuele waargenomen effecten van peerassessment op het schrijfproduct en de wenselijkheid van peerassessment als structureel onderdeel van het programma. Daarnaast zijn vragen gesteld ten aanzien van de interactie. Voor een overzicht van de interviewvragen zie Van den Berg, 2003.

Om de samenstelling van de feedback te kunnen bepalen en de interactie te typeren is alle schriftelijke en mondelinge feedback verzameld. Voor het analyseren van de mondelinge en schriftelijke peerfeedback is een codeersysteem ontwikkeld. Daarbij wordt de feedback gecodeerd in termen van de feedbackfunctie en feedbackaspect. Volgens Flower e.a. (1986) bestaat adequate feedback op schrijfproducten minimaal uit vier typen uitspraken. Om een schrijfproduct te kunnen beoordelen, moet het eerst geanalyseerd worden. De beoordeling hoort te worden toegelicht en de schrijver op weg te helpen bij de aanpak van de revisie. Aldus worden, mede op basis van Roossink (1990), vier feedbackfuncties onderscheiden: analyseren, evalueren, toelichten en reviseren. Dit type uitspraken heeft betrekking op het product, daarnaast onderscheiden we niet-productgerichte feedback, waarin bijvoorbeeld de werkwijze besproken wordt (Werkwijze) of de aanpak van de beoordeling (Oriënteren).

Naast 'feedbackfunctie' wordt de term 'feedbackaspect' gebruikt om aan te geven waarop de feedback is gericht: inhoud, structuur of stijl (zie Steehouder, Jansen, Maat, Staak & Woudstra, 1992). Het aspect 'inhoud' betreft de relevantie van de informatie, de helderheid van de probleemstelling, argumentatie en uitleg van begrippen. 'Structuur' staat voor de innerlijke samenhang, zoals de relatie tussen hoofd- en deelvragen en de relatie tussen probleemstelling, uitwerking en conclusie. Onder 'stijl' wordt al datgene verstaan wat te maken heeft met de uiterlijke vorm van de tekst, zoals taalgebruik, grammatica, spelling en lay-out. Ook met betrekking tot het feedbackaspect is er sprake

ke van productgerichte en niet-productgerichte feedback. Het laatste type betreft opmerkingen van gesprekstechnische aard, of onderwerpen die wel met de vakinhoud te maken hebben, maar niet met het schrijfproduct.

Om de interactie tussen studenten tijdens de mondelinge toelichting van hun schriftelijke beoordeling te kunnen typeren, is een typologie toegepast van Lockhart en Ng (1995). Zij onderscheiden vier feedbackposities: autoritair, interpretatief, onderzoekend en collaboratief. Feedback die gegeven wordt vanuit een autoritaire positie is er vooral op gericht om de ander duidelijk te maken wat er fout is of ontbreekt aan de tekst.

De interpretatieve feedbackgever wil ideeën bespreken die in hem opkomen naar aanleiding van de tekst. De onderzoekende en de collaboratieve feedbackgevers daarentegen houden rekening met de intenties van de schrijver. De onderzoekende feedbackgever wil de tekst verhelderen en beter aan laten sluiten bij de bedoelingen van de schrijver. De collaboratieve feedbackgever zoekt samen met de schrijver naar manieren om de tekst rijker en betekenisvoller te maken. De eerste twee feedbackposities worden door Lockhart en Ng (1995) 'evaluatief' genoemd, de derde en vierde 'ontdekkend'. Hoewel de evaluatieve manier van feedback geven volgens de auteurs niet per se slecht is en wellicht een noodzakelijke stap is in het leren feedback geven, leveren ze op den duur weinig leerwinst op voor beide partijen. De ontdekkende manier van feedback geven doet dat wel, als gevolg van het intensief samen zoeken en communiceren.

Om iets te kunnen zeggen over de feedbackpositie van de studenten in het onderzoek zijn de resultaten van de analyse van de mondelinge feedback met betrekking tot de feedbackfunctie opnieuw bestudeerd. Tevens is de gespreksinbreng van respectievelijk feedbackgever(s) en feedbacknemer geïnventariseerd. Omdat sommige categorieën in ons onderzoek niet traceerbaar waren, is alleen de grove typering in termen van evaluatieve en ontdekkende feedbackpositie toegepast (voor een toelichting zie Van den Berg, 2003). Onze operationalisatie van de typologie van Lockhart & Ng wordt samengevat in figuur 5.

De data zijn geanalyseerd met behulp van inhoudsanalyse en beschrijvende statistiek. Voor de analyse van onderlinge relaties is gebruikgemaakt van de χ^2 -toets; voor de analyse van eventuele verschillen tussen groepen is gebruikgemaakt van een t-toets voor verschil in gemiddelden en non-parametrische variantieanalyse. Om de resultaten in verband te kunnen brengen met ontwerpkenmerken is gebruikgemaakt van de matrixmethode, zoals omschreven in Miles en Huberman (1994). De kwaliteit van de instrumenten en methoden is gecontroleerd aan de hand van interbeoordelaarsbetrouwbaarheid, *member check* en *peer debriefing*.

RESULTATEN

Uitvoering procedure

De resultaten met betrekking tot de uitvoering van de peerassessment-procedure en de verschillen tussen de ontwerpen worden samengevat in tabel 1. Alle ontwerpen overziend, is de peerassessment-procedure volgens plan uitgevoerd. Ongeveer 80% van de studenten leverde op tijd een conceptversie in bij de leden van de feedbackgroep, die

kregen daarop feedback van minstens 1 medestudent en beoordeelden het schrijfproduct van minstens 1 medestudent. 67% van de studenten kreeg het aantal beoordelingen dat voorzien was in het ontwerp. Slechts enkele studenten kregen geen feedback. In twee ontwerpen (DS en TGM) besteedden de studenten te weinig tijd aan het lezen en beoordelen. Zie voor de hierbij gehanteerde tijdsnorm Van den Berg, 2003.

Samenstelling peerfeedback

De frequenties waarmee studenten de feedbackfuncties vullden in hun schriftelijke, respectievelijk mondelinge peerfeedback worden gepresenteerd in tabel 2 en 3.

De ontwerpen verschillen significant met betrekking tot de frequenties van de verschillende feedbackfuncties. Dit geldt voor de schriftelijke feedback ($\chi^2 = 143$; $df = 48$; $p \leq 0,001$, voor alle feedbackfuncties; $\chi^2 = 77$; $df = 24$; $p \leq 0,001$, voor uitsluitend de productgerichte feedbackfuncties) alsook voor de mondelinge peerfeedback ($\chi^2 = 351$; $df = 42$; $p \leq 0,001$, voor alle feedbackfuncties; $\chi^2 = 93$; $df = 21$; $p \leq 0,001$, voor uitsluitend de productgerichte feedbackfuncties).

In hun schriftelijke feedback waren de studenten vooral gericht op het beoordelen van de tekst. In de mondelinge toelichting besteedden ze veel aandacht aan het beargumenteren van hun oordeel en het geven van suggesties voor de revisie.

De samenstelling van de peerfeedback in termen van feedbackaspect wordt gepresenteerd in Tabel 4 en 5.

Ook met betrekking tot het feedbackaspect geldt dat de feedback per ontwerp verschilt. Dit geldt voor de schriftelijke feedback ($\chi^2 = 116$; $df = 40$; $p \leq 0,001$ voor alle feedback; $\chi^2 = 86$; $df = 24$; $p \leq 0,001$ voor uitsluitend productgerichte feedback) alsook voor de mondelinge feedback ($\chi^2 = 330$; $df = 35$; $p \leq 0,01$ voor alle feedback; $\chi^2 = 251$; $df = 21$; $p \leq 0,001$ voor uitsluitend productgerichte feedback).

De schriftelijke en mondelinge feedback blijkt vooral gericht te zijn op de aspecten 'inhoud' en 'stijl' van het te beoordelen product, maar niet op 'structuur.'

De tabellen 6 en 7 laten een significant verband zien tussen de manier van feedback geven (schriftelijk of mondeling) en de feedbackfuncties die worden vullend. Hetzelfde geldt voor het feedbackaspect. De kans dat alle feedbackfuncties worden vullend waarbij alle aspecten van de tekst worden bestreken, is het grootst als de schriftelijke feedback mondeling wordt toegelicht.

Tabel 6 Relatie feedbackfunctie en wijze van feedback geven

	C	χ^2	df	p
SS	0,42	105	6	$p \leq 0,001$
HA-1	0,30	86	6	$p \leq 0,001$
HA-2	0,29	91	6	$p \leq 0,001$
DS	0,37	92	6	$p \leq 0,001$
Hist1	0,30	44	6	$p \leq 0,001$
Hist2	0,23	42	6	$p \leq 0,001$
RP	0,21	46	6	$p \leq 0,001$
ICE	0,22	55	6	$p \leq 0,001$

Tabel 7 Relatie feedbackaspect en wijze van feedback geven

	C	χ^2	df	p
SS	0,33	59	5	$p \leq 0,001$
HA-1	0,24	48	5	$p \leq 0,001$
HA-2	0,25	62	5	$p \leq 0,001$
DS	0,42	108	5	$p \leq 0,001$
Hist1	0,32	49,1	5	$p \leq 0,001$
Hist2	0,28	58,8	5	$p \leq 0,001$
RP	0,33	109,1	5	$p \leq 0,001$
ICE	0,31	58,4	5	$p \leq 0,001$

Interactie

Om inzicht te krijgen in de interactie tussen studenten tijdens de peerassessment-procedure is eerst per ontwerp vastgesteld hoe groot de inbreng van de feedbacknemer was, gerekend in het aantal woorden. Onze conclusie is dat er in vijf ontwerpen sprake was van een evenredige inbreng, aangenomen dat ieder lid van de feedbackgroep ongeveer evenveel woorden mag inbrengen. In twee ontwerpen was de inbreng van de feedbacknemer te laag. De resultaten worden gepresenteerd in tabel 8.

Tabel 8 Inbreng feedbacknemer in de mondelinge peerfeedback (in % van het totale aantal woorden per ontwerp)

Inbreng	SS	HA-1	HA-2	DS	Hist1	Hist2	RP	ICE
Fg	78	24	26	60	66	72	69	67
d		49	53					
Fn	22	27	20	40	34	28	31	33
gr	3	3+d	3+d	2	3	4	3	3

Noot * = Fg=feedbackgever(s);

d = docent;

Fn = feedbacknemer;

gr = gemiddelde aantal leden van de feedbackgroep dat aanwezig was tijdens de mondelinge feedback.

Vervolgens is de feedback geanalyseerd op het niveau van de productbespreking, waarvan er afhankelijk van de groepsomvang in elk feedbackgroepje 2, 3 of 4 plaatsvonden. Daarbij is per bespreking de relatieve verhouding van feedbackfuncties bepaald en getypeerd in termen van onze typologie. De resultaten worden gepresenteerd in tabel 9.

Tabel 9 Feedbackpositie per ontwerp (in % van het aantal productbesprekingen per ontwerp)

	SS	HA-1	HA-2	DS	Hist1	Hist2	RP	ICE	Gem
Evaluatief	80	31	41	50	66	58	54	85	58
Ontdekkend	63	50	25		29	31		26	
Niet geclassificeerd	20	6	8	25	33	14	15	15	16
Tot.	100	100	100	100	100	100	100	100	100

Uit dit resultaat blijkt dat studenten vaker een evaluatieve dan een ontdekkende feedbackpositie innemen. In het licht van de taakinstructie die de studenten hadden gekregen is dit geen onverwacht resultaat. De studenten hadden immers de opdracht om het werk van medestudenten te beoordelen. Daar kwam bij dat de te beoordelen schrijfproducten meestal in een vergevorderd stadium waren, waardoor de bereidheid bij de feedbacknemer om zaken in de tekst opnieuw te overwegen wellicht niet erg groot was. Dit kan terugslaan op de feedbackgever, die daardoor zijn commentaar misschien liever zo snel mogelijk deponereert. In dit verband is het opvallend dat de ontdekkende feedbackpositie in het ontwerp HA relatief vaak wordt ingenomen. In dit ontwerp beoordeelde men het schrijfproduct in twee fasen, geruime tijd voordat er sprake is van een vrijwel afgerond werkstuk. De eerste beoordeling vond plaats in de planfase (de opzet van het werkstuk), de tweede in een vroeg stadium van uitwerking (de conceptversie van het kernhoofdstuk).

Leeropbrengst

De resultaten met betrekking tot de mate waarin de feedback is verwerkt, worden gepresenteerd in tabel 10. De percentages in de tabel geven aan hoeveel revisies van een bepaald type zijn voorgesteld en hoe vaak die zijn overgenomen. De omvang en diepgang van de revisie kan niet uit de tabel worden afgeleid.

Tabel 10 Verwerking van de peerfeedback (in % van het totale aantal revisieopmerkingen per ontwerp)

		% Inhoud	% Structuur	% Stijl	Aantal *
ICE	Verwerkt	50	0	50	22 (23%)
	Niet verwerkt	38	0	62	74
TGM	Verwerkt	43	6	50	16 (34%)
	Niet verwerkt	52	13	35	31
HIST-1	Verwerkt	50	5	45	20 (41%)
	Niet verwerkt	34	10	55	29
HIST-2	Verwerkt	41	12	47	17 (29%)
	Niet verwerkt	54	2	44	41
RP	Verwerkt	33	8	58	24 (34%)
	Niet verwerkt	24	7	70	46
Tot.	Verwerkt	44	5	51	98 (31%)
	Niet verwerkt	39	5	55	222

Noot. * het totale aantal revisieopmerkingen, uitgesplitst naar verwerkt en niet-verwerkt;

SS, HA en DS zijn niet opgenomen omdat het docentcommentaar interfereerde met de peerfeedback.

De studenten verwerkten gemiddeld 33% van alle revisievoorstellen van hun medestudenten. Van de revisies had 44% betrekking op de inhoud, en 51% op de stijl. De mate waarin de feedback is verwerkt verschilt per ontwerp.

De resultaten met betrekking tot de cijfers worden gepresenteerd in de tabellen 11 en 12. In drie ontwerpen (HIST, ICE en TGM) konden de cijfers voor de eindversie van het schrijfproduct van de peerassessment-groepen met die in de controlegroep worden vergeleken. Uit deze analyses kwamen geen significante verschillen in cijfers naar voren.

Tabel 11 T-test cijfergemiddelden peerassessment-groep (PA) en controlegroep (C)

	PA	C	t	df	p
Hist1	7.0	6.8	.36	21	.72
Hist2	7.2	6.8	1.07	25	.30
ICE	7.1	7.5	-1.29	30	.21
TGM	6.5	5.8	1.29	13	.22

Noot. De cijfergemiddelden in de kolom PA en C betreffen de cijfers voor de eindversie van het schrijfproduct.

In twee ontwerpen (DS en RP) gaf de docent een indicatief cijfer voor de aan de medestudenten voorgelegde conceptversie. Uit een t-test voor gemiddeld verschil komt naar voren dat in beide ontwerpen studenten een significant beter cijfer hebben gehaald voor de bijgestelde, tweede versie van het schrijfproduct. Door de scores op beide versie met elkaar in verband te brengen, kwam bovendien naar voren dat dit resultaat gelijk was voor de hele groep studenten ($r=0,88$ voor RP en $r=0,92$ voor DS). Bij RP, waar de feedback van de docent globaal was gehouden, was deze cijfervoortgang

aantoonbaar het gevolg van de verwerking van de peerfeedback. Uit vergelijking van de twee versies bleek dat alle studenten delen van de vaak gedetailleerde en daardoor in de verwerking gemakkelijk te traceren schriftelijke en mondelinge feedback van hun medestudenten hadden verwerkt. In DS verwerkten de studenten vaker het gedetailleerde en in de conceptversie bijgeschreven docentcommentaar, dat de studenten tegelijk ontvingen met de feedback van hun medestudenten.

Tabel 12 T-test cijfervoortgang eerste en tweede versie schrijfproduct

	v1	v2	t	df	p
DS	6.8	7.3	-3,38	9	.008
RP	5.7	6.5	-4,26	11	.001

Noot. v1= conceptversie; v2= gereviseerde versie.

Uit de antwoorden op de vragenlijst over peerassessment blijkt dat 80% van de totale groep studenten dacht dat hun schrijfproduct als gevolg van hun verwerking van de peerfeedback beter zou zijn dan de versie die ze aan hun medestudenten hadden voorgesteld. Uit een Kruskal-Wallis variantieanalyse blijkt dat studenten uit de verschillende ontwerpen hierover verschillend denken (KW=15.0; df=7; p=.04). Zo verwachtten de studenten in SS en in één van de twee peerassessment-werkgroepen van HIST veel verbetering, de studenten in TGM daarentegen, verwachtten weinig.

Waardering

Uit de antwoorden van de studenten naar aanleiding van de vragenlijst over peerassessment blijkt dat gemiddeld 75% van de studenten het lezen en beoordelen van het werk van hun medestudenten als een nuttige activiteit heeft ervaren, die niet te veel tijd vroeg. De ontvangen schriftelijke peerfeedback werd door 83% en het mondelinge commentaar door 78% van de studenten positief gewaardeerd. Als voordelen van peerassessment noemden de studenten vooral dat men ideeën kan opdoen door elkaars werk. Een nadeel vonden sommigen dat het lezen en beoordelen niet in de studielast gecompenseerd werd. Een paar studenten vonden de afhankelijkheid van medestudenten een groot nadeel. Gemiddeld 61% van de studenten zou vaker met peerassessment willen werken, daarbij zou 40% van deze groep de methode op enigerlei wijze willen aanpassen. Omdat de suggesties die daarvoor gegeven worden, sterk uiteenlopen zijn ze vooralsnog moeilijk op te nemen in een eventueel herontwerp.

De docenten zien voor- en nadelen aan het werken met peerassessment. Een docent noemt als voordeel van peerassessment dat de studenten op deze manier vertrouwd raken met een manier van samenwerken die ook in de geesteswetenschappen steeds meer gebruikelijk wordt. Een positief punt dat door zeven docenten genoemd wordt, is dat peerassessment een structuur biedt waardoor studenten gedwongen worden om zich in elkaars werk te verdiepen, zodat er een gezamenlijk gesprek over gevoerd kan worden. Vier docenten waren van oordeel dat de rol van de docent te onduidelijk is in het systeem van peerassessment. Zijzelf hadden meer begeleiding willen geven, maar ervoeren dat de tijd daarvoor ontbrak nu de studenten als medefeedbackgevers aan

bod moesten komen. Ondanks deze bedenkingen stonden alle docenten positief tegenover de gedachte om peerassessment als onderdeel in hun cursus op te nemen.

CONCLUSIE

In dit onderzoek zijn zeven ontwerpen geëvalueerd om meer zicht te krijgen op factoren die ertoe doen bij het organiseren van peerassessment in een cursus. Onderzocht is het verloop van de uitvoering, de samenstelling van de feedback, de interactie tijdens de mondelinge toelichting, de leeropbrengst en de waardering van de assessment. De uitkomsten kunnen aldus worden samengevat.

Over het geheel genomen is het assessment volgens plan uitgevoerd, zowel door de studenten als de docenten. In hun schriftelijke feedback waren de studenten sterk gericht op het beoordelen van de tekst. In de mondelinge toelichting werden ook andere feedbackfuncties vervuld, zoals het beargumenteren van hun oordeel en het geven van suggesties voor verbetering. De kans dat alle feedbackfuncties worden vervuld en dat daarbij alle tekstuele aspecten belicht worden, is groter bij een combinatie van schriftelijke en mondelinge feedback. Het blijkt dat studenten elkaar vaker op een evaluatieve dan op een ontdekkende manier feedback geven.

Vrijwel alle studenten gebruikten de peerfeedback bij de revisie; gemiddeld werd 33% van alle aanwijzingen van medestudenten verwerkt. Van de revisies was ongeveer de helft op inhoud gericht en de helft op stijl. Van de totale groep studenten dacht 80% dat hun tweede versie beter was dan de eerste als gevolg van hun verwerking van de peerfeedback. De mate waarin ze die mening zijn toegedaan, verschilt per ontwerp. De eindcijfers voor het schrijfproduct in de peerassessment-groepen verschilden niet significant van de eindcijfers in de controlegroep.

Het werken met peerassessment werd door de studenten positief gewaardeerd. Ze hadden het lezen en het beoordelen van het werk van hun medestudenten als een nuttige activiteit ervaren en vonden de feedback die ze van hen hadden gekregen over het algemeen bruikbaar. Ongeveer 67% van de studenten zou vaker met peerassessment willen werken.

De meerderheid van de docenten zag peerassessment als een effectieve manier om het eenrichtingsverkeer tussen docent en student te doorbreken. Enkelen verklaarden moeite te hebben gehad met het bepalen van hun rol. Ze hadden meer begeleiding willen geven dan mogelijk was, nu de studenten als medefeedbackgevers ook aan bod moesten komen.

Optimaal model

Op basis van de vergelijking van de resultaten per ontwerp trekken we de volgende conclusies over ontwerpkenmerken die de werking van peerassessment ondersteunen. Deze zijn samengevat in tabel 13.

Tabel 13 Kenmerken van een optimaal ontwerp van peerassessment

	Cursus X
(4) Product	Conceptversie artikel of werkstuk(5 –8pp)
(5) Relatie docentbeoordeling	Aanvullend; extra beoordeling
(7) Oordeelsrichting	Wederzijds (2 beoordelingen)
(8) Privacy	Vertrouwelijk (binnen de feedbackgroep)
(9) Contact	Schriftelijke en mondelinge feedback; plenaire bespreking o.l.v. docent
(11) Bekwaamheid	At random op basis van verwant of zelfde onderwerp
(12) Constellatie beoordelaars	Kleine groepen (3 studenten)
(13) Constellatie beoordeelden	Dezelfde kleine groepen
(14) Plaats	Schriftelijke feedback buiten resp. mondelinge feedback in contacttijd
(17) Beloning	Geen beloning van deelname aan peerassessment

De resultaten in de tabel kunnen worden uitgedrukt in de vorm van een aantal richtlijnen voor de organisatie van peerassessment.

- 1- Bij schrijfproducten met een omvang van vijf tot acht pagina's is de kans groter dat de studenten voldoende tijd besteden aan het lezen en beoordelen dan bij omvangrijkere producten.
- 2- Het moment van peerassessment moet duidelijk gescheiden zijn van de docentbeoordeling en enige tijd daarvoor plaatsvinden, zodat de kans groot is dat de studenten de peerfeedback daadwerkelijk gebruiken om hun werk te verbeteren.
- 3- Wederzijdse feedback is gemakkelijker te organiseren dan éénrichting-feedback. De uitwisseling van producten verloopt soepeler omdat het een kwestie is van 'gelijk oversteken'. De mondelinge toelichting in contacttijd bij wederzijdse beoordeling hoeft weinig tijd te kosten omdat de feedbackgroepen parallel met elkaar hun feedback kunnen bespreken.
- 4- Mondelinge toelichting, verheldering en zoeken naar verbetering zijn onmisbare element van het feedbackproces, dat vereist face-to-face contact.
- 5- Niet volledig bekend zijn met elkaars onderwerp hoeft niet te leiden tot een meer op stijl gerichte beoordeling. Een *at random* indeling voldoet doorgaans.
- 6- Wanneer de docent medefeedbackgever is tijdens de mondelinge feedback, zijn de studenten veel minder aan het woord dan in feedbackgroepen waarin de docent niet meedoet. Het is aan te bevelen dat de docent zich op het moment dat de studenten elkaar feedback geven, beperkt tot de rol van procesbewaker.
- 7- De feedbackgroep omvat bij voorkeur drie of vier studenten, zodat ze de kans hebben om verschillende commentaren te vergelijken en de relevantie ervan te bepalen. Wanneer de studenten in tweetallen werken is het risico groter dat bij niet goed functioneren van een of beide partners, de feedback niet gegeven wordt of te weinig oplevert.
- 8- De mondelinge toelichting kan het beste in contacttijd plaatsvinden, omdat de studenten er vaak zelf geen kans toe zien om een gezamenlijk moment en een geschikte plek buiten de les te vinden waarin ze dit geconcentreerd kunnen doen.

DISCUSSIE

Omdat het onderzoek is uitgevoerd in de onderwijspraktijk en niet in een laboratorium-situatie was het niet mogelijk om alle variabelen volledig onder controle te houden. Ook had het onderzoek vollediger kunnen zijn. Zo is er geen aandacht besteed aan het proces van verwerking van de feedback: hoe maken studenten een selectie uit het commentaar dat ze hebben gekregen en hoe leidt die tot de feitelijke revisie? Bovendien zijn alleen resultaten verkregen die op de korte termijn zichtbaar gemaakt konden worden. Niet duidelijk is hoe studenten op peerassessment reageren als ze er vaker mee te maken krijgen.

Interessant voor vervolgonderzoek is het ontwikkelen van vormen van peerassessment die minder beoordelend zijn en meer ontdekkend. Op grond van Saunders (1989) is de uitkomst dat de feedbackpositie van de studenten in ons onderzoek hoofdzakelijk evaluatief is, immers niet verrassend. De feedbackgever was niet medeverantwoordelijk voor het product, werd er pas in de eindfase bij betrokken en had bovendien als taak gekregen om het schrijfproduct te beoordelen. De vraag is hoe een minder evaluatieve en meer ontdekkingsgerichte taakstelling eruit zou kunnen zien. Leidt die ertoe dat de studenten vaker een ontdekkende feedbackpositie innemen? En hoe vertaalt zich dat in leerwinst? Hoe creëer je bij een meer op samenwerken gerichte taakstelling toch de distantie die nodig is om elkaar feedback te geven?

Peerassessment krijgt in het universitaire onderwijs meer en meer aandacht. Dat betekent echter ook dat er vele variaties in ontwerpen worden gebruikt, al dan niet gebaseerd op inhoudelijke argumenten of gefundeerd op onderzoek. De tijd lijkt aangebroken om dergelijk onderzoek op brede schaal uit te voeren om te bewerkstelligen dat in de toekomst peerassessment op een efficiënte en effectieve wijze kan worden ingezet als ondersteuning van het leerproces van studenten.

LITERATUUR

- Bean, J. C. (1998) *Schrijvend leren en andere didactische werkvormen voor actief leren*. Leiden: Spruyt, Van Mantgem & De Does bv.
- Berg, B.A.M. van den (2003) *Peerassessment in universitair onderwijs. Een onderzoek naar bruikbare ontwerpen*. Dissertatie, Universiteit Utrecht, Utrecht.
- Creswell, J. W. (1998) *Qualitative Inquiry and Research Design: choosing among five traditions*. Thousand Oaks: SAGE Publications.
- Flower, L., Hayes, J. R., Carey, L., Schriver, K., & Stratman, J. (1986) Detection, Diagnosis, and the strategies of revision. *College Composition and Communication*, 37, 16-55.
- Gallimore, R., & Sharp, R. (1990) Teaching mind in society: Teaching, schooling, and literate discourse. In: L. C. Moll (Ed.), *Vygotsky and education: instructional implications and applications of sociohistorical psychology* 175-205) Cambridge: Cambridge University Press.
- Lockhart, C., & Ng, P. (1995) Analyzing Talk in ESL Peer Response Groups: Stances, Functions and Content. *Language Learning*, 45, 605-655.

- Miles, M. B., & Huberman, A. M. (1994) *Qualitative data analysis: an expanded source-book* (2nd ed.) Thousand Oaks, California: SAGE.
- Roossink, H. J. (1990) *Terugkoppelen in het natuurwetenschappelijk onderwijs, een model voor de docent*. Dissertatie, Universiteit Twente, Enschede.
- Saunders, W. M. (1989) Collaborative writing tasks and peer interaction. *International Journal of Educational Research*, 13, 101-112.
- Steehouder, M., Jansen, C., Maat, K., Staak, J. v.d., & Woudstra, E. (1992) *Leren communiceren* (3e herziene druk ed.) Groningen: Wolters-Noordhoff.
- Tharp, R. G., & Gallimore, R. (1988) *Rousing minds to life: teaching, learning, and schooling in social context*. Cambridge etc.: Cambridge University Press.
- Topping, K. (1998) Peerassessment between students in colleges and universities. *Review of Educational Research*, 68, 249-276.
- Tudge, J. (1990) Vygotsky, the zone of proximal development, and peer collaboration: Implications for classroom practice. In: L. C. Moll (Ed.), *Vygotsky and education: instructional implications and applications of sociohistorical psychology*. Cambridge etc.: Cambridge University Press.