

Sneller en beter feedback geven met digitale tools



Onderzoeksrapportage SURF-project:

SCALA - Scaffolding Assessment for Learning

Februari 2014

De projectgroep:

Ineke van den Berg (UU)

Sumit Mehra (HVA)

Patris van Boxel (VU)

Janneke van der Hulst (VU)

Sanne Meeder (VU)

José Beijer (HU)

Annelies Riteco (HU)

Sanne Gratama van Andel (UU)

Colofon

Titel: *Sneller en beter feedback geven met digitale tools.*

Auteurs, projectgroep SCALA

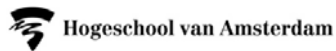
Datum
Februari 2014

Dit rapport is geschreven in het kader van het SURF-project SCALA – Scaffolding Assessment for Learning.

Penvoerende instelling
Universiteit van Utrecht, Centrum voor Onderwijs en Leren
Postbus 80127, 3508 TC Utrecht, (030) 253 3400, onderwijsleren@uu.nl.
Projectleider : Sanne Gratama van Andel, r.gratamavanandel@uu.nl

Overige participerende instellingen
Hogeschool van Amsterdam, Faculteit Economie en Management
Hogeschool Utrecht, Faculteit Educatie
Vrije Universiteit Amsterdam, Universiteitsbibliotheek

Dit project is tot stand gekomen met steun van SURFfoundation. De organisatie die ICT vernieuwingen in het hoger onderwijs en onderzoek initieert, regiseert en stimuleert door onder meer het financieren van projecten. Meer informatie over SURF is te vinden op de website (www.surf.nl).



Inhoud

1.	Inleiding	6
1.1	Aanleiding en probleemstelling	6
1.2	Onderzoeksvragen	7
2.	Onderzoek	7
2.1	Pilots	7
2.2	Opzet onderzoek	8
2.3	Instrumenten	9
2.4	Uitvoering dataverzameling	9
2.5	Wijze van analyse	10
3.	Resultaten	10
3.1	Respons	10
3.2	Tijdbesparing (vraag 1)	11
3.2.1	Tijdbesparing overall	11
3.2.2	Relatie systemen/tools	13
3.3	Kwaliteit van de feedback c.q. beoordeling (vraag 2)	15
3.3.1	Kwaliteit volgens docenten	15
3.3.2	Kwaliteit volgens studenten	17
3.4	Eindoordeel volgens docenten en studenten	21
4.	Conclusie en discussie	21
4.1	Conclusies	21
4.1.1	Tijdbesparing	22
4.1.2	Kwaliteit feedback-c.q. beoordeling	22
4.2	Discussie	22
5.	Literatuur	25
6.	Bijlagen	26
6.1	Docentvragenlijst	26
6.2	Studentvragenlijst	26
6.3	Interviewleidraad	26
6.4	Resultaten VU	26

6.5	Resultaten UU	26
6.6	Resultaten HU	26
6.7	Resultaten HvA	26

1. Inleiding

1.1 Aanleiding en probleemstelling

Verbetering van het onderwijsrendement door vermindering van studie-uitval staat in de meeste instellingen voor hoger onderwijs hoog op de beleidsagenda. Algemeen wordt ingezien dat studenten die voldoende voor de opleiding gekwalificeerd en gemotiveerd zijn in principe binnen de nominale tijd hun diploma behoren te behalen. Een cruciale rol bij de studievoortgang speelt de programmering en kwaliteit van de toetsing, zowel in summatieve als formatieve zin. Studeergedrag wordt in sterke mate gestuurd door de toetsmomenten (zie onder meer Ramsden, 2003; Gibbs, 1999; Sainsbury & Walker, 2008): studenten beginnen pas te studeren als ze beseffen dat de toets in zicht komt. Cursussen die maar één toetsmoment kennen, gewoonlijk aan het eind van de cursus, zetten niet alleen aan tot uitstelgedrag, studenten krijgen op deze manier ook te weinig feedback. Uit reviewstudies van Hattie en Timperley (2007), en Shute (2008), blijkt dat feedback een van de krachtigste strategieën is om het rendement van onderwijs te verhogen. Voorwaarde daarbij is dat de feedback voldoende hulp biedt om de ontvanger op het goede spoor te zetten, zodat hij weet hoe hij zichzelf aan moet sturen om aan de criteria te kunnen voldoen. Feedback in deze betekenis wordt aangeduid als feedforward (Hattie & Timperley, 2007). Het is een bekend gegeven dat studenten vaak weinig opsteken van de feedback die ze krijgen (Gibbs, 2010). Verschillende redenen kunnen daaraan ten grondslag liggen, zoals het late tijdstip van de feedback waardoor studenten er niets meer mee kunnen doen om hun prestatie te verbeteren. Daarnaast gebeurt het ook vaak dat de feedback niet wordt begrepen (Chanock, 2000) of te weinig houvast biedt (Weaver, 2006).

Terecht krijgt de toetsing dan ook veel aandacht in het onderwijsbeleid van universiteiten en hogescholen. Als bijpassende maatregelen worden vaak meerdere toetsmomenten in een cursus voorgeschreven en procedures ontwikkeld voor het verbeteren van de afstemming en transparantie in de beoordeling en feedback. De keerzijde hiervan is dat de werkdruk van docenten toeneemt. Verantwoord toetsen en feedback geven vereist nu eenmaal een stevige tijdsinvestering, vooral als het gaat om schrijfproducten. Aangezien bij de meeste opleidingen in het hoger onderwijs bij uitstek getoetst wordt in de vorm van geschreven producten, en dat meermalen per cursus, legt dit een groot beslag op de docenttijd. Het werken met digitale beoordelings- en feedbacktools kan helpen om die tijdsinvestering beheersbaar te houden.

In het SCALA-project zijn 28 pilots uitgevoerd waarbij voor de beoordeling en feedback gebruik is gemaakt van de softwareapplicaties Turnitin (op de Vrije Universiteit, Hogeschool van Amsterdam en Universiteit Utrecht) en WinVision (Hogeschool Utrecht).

Turnitin is een Amerikaans online computersysteem waarmee digitaal schriftelijke producten van feedback kunnen worden voorzien en worden beoordeeld. Het systeem bestaat uit drie delen: Originality check, GradeMark en PeerMark. De Originality check is een systeem voor plagiaatdetectie. GradeMark is het deel van Turnitin voor docenten waarin ze digitale feedback geven kunnen geven op schriftelijke producten, PeerMark is het deel van Turnitin dat peerfeedback faciliteert. In het SCALA project is vooral gebruik gemaakt van GradeMark. In GradeMark kan een docent op drie verschillende manieren feedback geven: door middel van een beoordelingsformulier (Rubric), door feedback in de tekst en door een geschreven of gesproken algemeen commentaar (GeneralComments). De feedback in de tekst kan worden opgeslagen, zo ontstaan de zogenaamde QuickMarks. Deze kunnen eenvoudig worden hergebruikt door ze met een muis naar de gewenste plek in de tekst te slepen.

Winvision Digitaal Portfolio richt zich op de ondersteuning van de onderwijsprocessen rond competentiegericht opleiden. Met deze tool leggen studenten een digitaal portfolio aan om hun competentieontwikkeling aan te tonen, te monitoren en te laten beoordelen.

Tijdens de opleiding fungeert het digitaal portfolio als instrument voor beoordeling en feedback. Het portfolio is dus zowel een communicatiemiddel voor feedback als een digitaal archiveringssysteem voor prestaties en beoordelingen. Ook medestudenten kunnen via het systeem feedback/feedforward aan hun studiegenoten geven. De docent kan zowel met gestandaardiseerde als open feedback en beoordeling werken. Voor gestandaardiseerde feedback/ beoordeling wordt gebruik gemaakt van beoordelingscriteria die docenten voor specifieke opdrachten ontwikkelen.

In dit project wordt het gebruik van bovenbeschreven digitale tools voor het beoordelen en feedback geven geïntroduceerd en geëvalueerd. Het gebruik van dergelijke tools biedt potentieel een aantal voordelen:

1. Doordat ze de 'turn-around-time' verminderen en de administratie vereenvoudigen, wordt de logistieke rompslomp voor docenten beperkt.
2. De in de digitale tools ingebouwde structuur van criteria en bijbehorende feedback maakt het inzichtelijker voor de studenten in welke opzichten hun prestatie aan de norm beantwoordt en op welke punten ze zich moeten verbeteren.
3. Door de ingebouwde feedbackmogelijkheden hoeft de feedback niet steeds te worden bedacht en uitgeschreven, wat tijd spaart.
4. Nevendoel van het project is het verbeteren van de objectiviteit en transparantie van de toetsing binnen de opleiding. Het formuleren van criteria voor schrijfproducten, rekening houdend met de opbouw van niveaus, en het bedenken van bijbehorende feedbackmogelijkheden, is een taak die vraagt om overleg tussen collega's die onderwijs geven binnen eenzelfde cursus alsook met collega's van aangrenzende cursussen, horizontaal en verticaal. Aldus worden een gemeenschappelijk stramien en taal ontwikkeld voor het beoordelingsproces, wat de transparantie en samenhang van de toetsing ten goede komt.

1.2 Onderzoeksvragen

In deze rapportage wordt beschreven in hoeverre bovenstaande claims waargemaakt zijn in de pilots die in het kader van SCALA zijn ondernomen. Daartoe zijn de volgende onderzoeksvragen geformuleerd:

1. In hoeverre heeft het gebruik van een digitale feedback- en beoordelingstool bijgedragen aan tijdswinst voor de docent bij het feedback geven en beoordelen?
2. In hoeverre heeft het gebruik van een digitale feedback- en beoordelingstool bijgedragen aan een verbetering van de feedback- en beoordelingskwaliteit?

Bij de weergave van de resultaten zullen eventuele verschillende uitkomsten van de pilots worden getoetst op mogelijke verbanden met de variabelen: 1. systeem (Turnitin vs WinVision), en 2. herhaling (eerste pilots vs herhaalpilots).

2. Onderzoek

2.1 Pilots

Om geïnteresseerde docenten te werven voor het uitvoeren van een pilot is een openingsbijeenkomst georganiseerd op de Universiteit Utrecht waar de mogelijkheden van Turnitin en WinVision zijn gedemonstreerd. Daarnaast is er binnen de instellingen informatie over het SCALA-project verspreid via facultaire nieuwsbrieven, gesprekken met onderwijscoördinatoren, onderwijsdirecteuren en contacten van de projectteamleden (allen onderwijskundig adviseur) op de werkvloer. Het was geen probleem om voldoende pilots te vinden. Vanwege de beperkte begeleidingscapaciteit moesten er zelfs geïnteresseerden worden afgewezen.

Voordat deze pilots startten waren binnen de HvA twee zogeheten pre-pilots met Turnitin uitgevoerd. Omdat deze destijds niet met het hierboven beschreven instrumentarium geëvalueerd konden worden zijn deze pre-pilots niet meegenomen in deze rapportage. Niettemin hebben de pre-pilots wel waardevolle ervaringen en materialen opgeleverd, met name voor de training van docenten.

De pilots zijn uitgevoerd gedurende het studiejaar 2012-2013 en in 2013-2014. Het betreft reguliere cursussen/ onderwijsmodules van verschillende opleidingen aan de vier instellingen. Omdat de pilots plaatsvonden in het kader van 'lopend onderwijs', waren de docenten verantwoordelijk voor de manier waarop zij de digitale feedbacktools gebruikten. Er is tussen de pilots onderling een grote variatie, bijvoorbeeld in het aantal betrokken docenten per pilot; studentenaantal, opleidingsjaar, vakgebied, functie van de feedback (formatief of summatief), type werkstuk dat beoordeeld werd, beoordeling door docent, door medestudenten (peer review) of beiden.

De docenten die deelnamen aan de pilots hebben wat voorlichting en instructie gekregen over het werken met de digitale nakijktools en enige ondersteuning tijdens het gebruik ervan. De tijd die de hieraan besteed werd verschilde niet of nauwelijks van de tijd die scholing en ondersteuning normaal gesproken kosten bij het in gebruik nemen van nieuwe digitale middelen. De tijd die van docenten gevraagd werd voor de evaluatie was wel 'extra', hier stond tegenover dat deze tijd vanuit het project vergoed werd. De variëteit tussen de pilots heeft als nadeel dat de onderlinge vergelijkbaarheid van de pilots laag is, wat de interpretatie van de uitkomsten van de evaluatie bemoeilijkt, het voordeel is echter dat de pilots een aardige dwarsdoorsnede geven van de wijze waarop docenten aan onze instellingen onder normale omstandigheden gebruik maken van de digitale nakijktools.

2.2 Opzet onderzoek

Om de onderzoeksvragen te beantwoorden is gekozen voor een combinatie van kwantitatieve en kwalitatieve methoden. Om de benodigde kwantitatieve data te verzamelen zijn twee online vragenlijsten ontwikkeld, een voor de docenten en een voor de studenten. De link naar deze vragenlijsten is in alle pilots ongeveer twee weken na de laatste bijeenkomst van de cursus verstuurd. De kwalitatieve data zijn verzameld met behulp van open antwoordmogelijkheden in genoemde vragenlijsten en via interviews. Voor de start van elke pilot is een intakegesprek met de docent gehouden, of in geval er meerdere docenten bij de uitvoering van dezelfde pilot betrokken waren, met de cursuscoördinator. Na afloop van de cursus vond een interview plaats met de docent dan wel het docententeam waarin de ervaringen tijdens de pilot werden besproken. Dit interview vond plaats nadat de docentvragenlijst en de studentvragenlijst waren ingevuld, de uitkomsten hiervan werden betrokken in het gesprek.

Om het effect van de inzet van een digitaal beoordelings- c.q. feedbacksysteem op de benodigde feedback- en beoordelingstijd van docenten te kunnen meten, was het aanvankelijke plan om waar dit kon een quasi-experimentele opzet te gebruiken met een controlegroep. Hierbij zou de tijdsbesteding van docenten binnen eenzelfde cursus met en zonder de digitale tool met elkaar worden vergeleken. Deze opzet kon echter in geen enkele pilot worden uitgevoerd omdat over het algemeen –een enkeling daargelaten– alle docenten binnen dezelfde cursus gebruik wilden maken van de softwareapplicatie. Aangezien de focus van dit project minder op een verandering van de objectieve dan op een verandering van de beleefde werkelijkheid ligt, het doel is immers een betere tijdsbeheersing voor de docent, is ervoor gekozen om de effecten op de tijdsbesteding alleen op perceptieniveau te meten. Dit gebeurde als volgt: bij de 'intake' formuleerde de docent/cursuscoördinator zijn of haar ambitie met betrekking tot de te behalen tijdsbesparing. In de vragenlijst kregen alle docenten, dus ook als ze niet bij de intake aanwezig waren, de vraag voorgelegd om de vergelijking te maken tussen de tijd die ze gewoonlijk kwijt waren aan het nakijken van dezelfde of een vergelijkbare opdracht. In

het interview na afloop werd om toelichting gevraagd op de redenen voor het al dan niet behaald hebben van een eventuele tijdswinst.

2.3 Instrumenten

De vragenlijsten voor de docenten en studenten zijn door het projectteam ontwikkeld. Dit gebeurde op basis van literatuurstudie, aangevuld met praktijkervaringen van de projectteamleden van de VU en de HvA, waar Turnitin al eerder in gebruik is genomen.

Beide vragenlijsten zijn zoveel mogelijk analoog opgebouwd. Ze openen met enkele vragen die bedoeld zijn om respondentgegevens te verzamelen, te weten: over de instelling waar men doceert/studeert, de naam van de opleiding en die van de cursus. Aan de studenten is daarnaast gevraagd naar de naam van de docent.

Het inhoudelijke deel bestaat uit vragen over de thema's Tijdbesparing (alleen in de docentvragenlijst), Feedbackkwaliteit en Gebruiksgemak. Daarnaast is er zowel om een algemeen alsook om een toolspecifiek eindoordeel gevraagd (in de zin van 'ik zou vaker met dit systeem/tool willen werken'). Per thema zijn 5-10 items opgenomen, meestal in de vorm van stellingen waarbij de respondent gevraagd wordt om op een Likertschaal van 1 (negatief) tot 5 (positief) aan te geven in hoeverre hij/zij het met de stelling eens is. In beide vragenlijsten werd bij een aantal vragen de mogelijkheid tot toelichting gegeven. Omdat er twee verschillende softwareapplicaties zijn gebruikt (Turnitin en Winvision) en er binnen die systemen gekozen kon worden voor het gebruik van een of meerdere tools (Rubrics, QuickMarks, GeneralComments en PeerMark in Turnitin, en Beoordelingscriteria in WinVision), is een aantal vragen ook toolspecifiek gesteld. De totale docentvragenlijst omvatte 67 vragen, de studentvragenlijst omvatte 45 vragen. De respondenten zijn via een routing, afgestemd op het systeem en de tools waarmee ze werkten, door de vragenlijst geleid. Alle vragen waren verplicht, met uitzondering van de vraag in de docentvragenlijst naar de naam van de docent.

De invulling van de studentvragenlijst was anoniem, in de docentvragenlijst werd de docenten gevraagd om hun naam in te vullen, maar dit was niet verplicht.

De docent- en studentvragenlijsten zijn opgenomen als bijlage 1 en 2.

Voor het interview na afloop met de docent of cursuscoördinator is een interviewleidraad ontwikkeld. Tijdens dit interview komt de opzet en het verloop van de pilot in het licht van de tijdens het intakegesprek geformuleerde ambities. Hierbij worden ook de resultaten van de docent- en studentevaluatie samen besproken en geïnterpreteerd. Dit kan resulteren in eventuele verbeterwensen, aanpassingen in het eigen vak, 'lessons learnt', en tips voor andere docenten.

De interviewleidraad is opgenomen als bijlage 3.

2.4 Uitvoering dataverzameling

De vragenlijsten zijn omgezet in een online format met behulp van SurveyMonkey. Direct na afloop van elke pilot kreeg de docent c.q. cursuscoördinator twee e-mails van de onderzoekers, een email met daarin de link naar de docentvragenlijst en een met daarin de link naar de studentvragenlijst. De eerste e-mail bevatte het verzoek om de docentvragenlijst in te vullen en –in geval van een docententeam- de email met link door te sturen naar de andere docenten (voor zover betrokken bij het feedback geven c.q. beoordelen). De tweede e-mail was bestemd voor de studenten, de docent of cursuscoördinator werd gevraagd om deze door te sturen naar alle studenten die aan de cursus hadden deelgenomen.

In totaal heeft de uitnodiging voor de docentvragenlijst 79¹ docenten en 2406 studenten bereikt. De dataverzameling van beide enquêtes is afgesloten op 10 december 2013.

De interviews met de docenten zijn gemiddeld een maand na afloop van de pilot gehouden door de leden van het projectteam. De leden interviewden de docenten van hun eigen instelling, de gesprekken werden een op een gevoerd, de gespreksduur varieerde tussen 45 en 60 min.

2.5 Wijze van analyse

De antwoorden op de interviewvragen zijn door de interviewers samengevat in het format van de interviewleidraad.

De kwantitatieve data worden gepresenteerd met behulp van beschrijvende en inductieve statistiek (dat laatste alleen bij de uitkomsten van de docentvragenlijst), waarbij correlaties en eventuele verschillen in gemiddelden waar dit relevant leek zijn onderzocht met behulp van parametrische en non-parametrische toetsen. De toelichtingen bij antwoorden uit de vragenlijsten en uitspraken in interviews worden in deze rapportage selectief gebruikt om de kwantitatieve uitkomsten te illustreren. De volledige verzameling toelichtingen en interviewverslagen is opvraagbaar bij de auteurs van dit rapport.

De interviews zijn samengevat in het format van de interviewleidraad en zijn gebruikt bij de interpretatie van de andere data.

3. Resultaten

3.1 Respons

In totaal zijn er data aangeleverd over 26 pilots. Eén pilot (pilot 44) werd op het laatste moment door de betreffende docent afgelast toen het niet mogelijk bleek om de zelf geformuleerde beoordelingscriteria –die waren opgemaakt met een van het systeem afwijkende systematiek- in te voeren. Bij een andere pilot (pilot 22) is de evaluatie om roostertechnische redenen zowel bij de docenten als studenten achterwege gebleven.

Bij alle 26 pilots is de studentvragenlijst afgenomen. De docentvragenlijst is bij 24 pilots ingevuld. Bij twee pilots hebben de docenten de vragenlijst niet ingevuld, ook niet na een herhaald verzoek. Daarnaast is bij alle 26 pilots de docent, of in geval van een docententeam, de cursuscoördinator geïnterviewd.

Van de 26 pilots zijn er vijf als 'herhaalpilot' uitgevoerd, te weten de pilots 23,24,25,45 en 46. Een herhaalpilot is een cursus die in hetzelfde of volgende studiejaar opnieuw is uitgevoerd in dezelfde opzet en met hetzelfde systeem, al dan niet met dezelfde docent(-en).

Van de 94 bij de pilots betrokken docenten hebben er 67 de vragenlijst ingevuld, dat betekent een responspercentage van 71,3 %. Van de 2406 studenten die aan een van de pilotcursussen hebben meegedaan hebben er 575 de vragenlijst geopend. De 23 studenten die na de eerste paar vragen over instelling en opleiding geen enkele vraag hebben ingevuld van het inhoudelijk deel zijn uit de dataset verwijderd, het aantal studenten dat de vragenlijst heeft ingevuld komt daarmee op 552, dus de overall respons van de studenten is 23,0% (zie de tabellen 1 en 2 voor de respons per instelling). Bij sommige pilots (Pilot 23 en 26) is de respons van studenten minder dan 10%, dus verre van representatief. Een verklaring voor de lage respons bij de studenten kan zijn dat ze vlak voor onze online vragenlijst ook de standaardevaluatie die door de opleiding aan het eind van elke cursus wordt afgenomen hadden moeten invullen. Het kan dus zo zijn dat vooral de meer gemotiveerde studenten (positief dan wel negatief) de moeite hebben willen nemen om beide vragenlijsten in te vullen. Een andere reden

¹ Hierbij (zowel docenten als studenten) zijn de pre-pilots van de HvA buiten beschouwing gelaten.

voor de lage respons kan zijn dat niet alle docenten de link naar de vragenlijst snel genoeg doorstuurden naar hun studenten, waardoor er teveel tijd overheen ging en het moment voor de studenten gepasseerd was waarop ze nog terug wilden blikken.

Tabel 1. Respons docenten per instelling.

Instelling	Aantal respondenten	Tot.aantal docenten	Responspercentage
VU	20	28	71
HvA	26	39	67
UU	14	20	70
HU	7	7	100
<i>Totaal</i>	<i>67</i>	<i>94</i>	

Tabel 2. Respons studenten per instelling.

Instelling	Aantal respondenten	Tot.aantal studenten	Responspercentage
VU	184	709	26
HvA	111	1085	10
UU	168	431	39
HU	89	181	49
<i>Totaal</i>	<i>552</i>	<i>2406</i>	

Al met al vinden we de respons van de bij de pilots betrokken docenten acceptabel: 71% (67 van 94) heeft de vragenlijst ingevuld. Dit is belangrijk omdat het SCALA-project primair op de docenten is gericht, het gaat immers uiteindelijk om het verlichten van hun werkdruk, met behoud van de feedback- en beoordelingskwaliteit. Om iets te kunnen zeggen over dat laatste hebben wij ook vragen gesteld aan de 'afnemers', de studenten. Hun respons mag met 23% aan de lage kant zijn, het is wel een gangbaar gemiddelde voor dit soort evaluaties. Doordat het toch alles bij elkaar veel studenten zijn die hun ervaringen met Turnitin en WinVision aan ons meedelen beschouwen we hun oordelen wel degelijk als waardevol om meer inzicht te krijgen in de uitwerking van de gebruikte systemen en tools op de ontvangers. Hun antwoorden mogen dan niet representatief genoemd worden, ze zijn wel informatief.

3.2 Tijdbesparing (vraag 1)

De eventuele tijdbesparing als gevolg van het werken met een digitaal feedback- / beoordelingssysteem is alleen onderzocht bij de docenten.

3.2.1 Tijdbesparing overall

Van de 65 docenten die de betreffende vraag hebben ingevuld heeft 38,5% minder dan de gebruikelijke² tijd besteed bij het feedback geven c.q. beoordelen, 46,2% heeft evenveel tijd als gewoonlijk nodig gehad en 15,4% heeft meer tijd moeten investeren. Van degenen (n=25) die tijd bespaard hebben heeft een derde 10%, een derde 20-30% en een derde 40-50% minder tijd dan gebruikelijk nodig gehad. De groep docenten die tijd hadden bespaard gaven aan gemiddeld ongeveer 25% tijd te hebben bespaard in vergelijking met voorgaande jaren.

Minder tijd

Docenten die tijd hadden bespaard lichten dit toe met uitspraken als bijvoorbeeld:

² Gevraagd is in dit verband ook naar de voor de docent gebruikelijke **manier** van nakijken: op papier, via de reviewoptie van Word, via e-mail e.d. Aangezien vrijwel iedereen hier aangaf de aangegeven opties in combinatie te gebruiken en het ondoenlijk bleek om daar substantiële groepen in te onderscheiden, is het niet mogelijk om de manier waar op men voorheen gewend was na te kijken als een variabele te beschouwen.

Handig om voorgeprogrammeerde feedback te slepen. (Turnitin)

Het was aan het begin wennen om van het scherm te lezen en om de QuickMarks en feedbackknoppen goed door te nemen. Maar na 2 adviesrapporten nagekeken te hebben ging het erg snel. Dus tijdswinst heb ik zeker behaald. (Turnitin)

Er zit veel winst in de overzichtelijkheid die je met dit systeem bereikt, zowel voor de docenten zelf als voor de studenten. De infrastructuur en organisatie van de cursus en opdrachten is nu veel overzichtelijker, dat is ook een vorm van tijdswinst. (Turnitin)

Ik maakte eerder gebruik van de 'review' optie in Word, waar je comments in de tekst kunt plaatsen, het lijkt een beetje op Turnitin maar daar moest ik alles steeds zelf in de commentswolkjes typen'. (Turnitin)

Turnitin biedt een handig overzicht van ingeleverde opdrachten. Ook je gegeven feedback wordt bewaard. Dit bespaart me veel e-mailverkeer. (Turnitin)

Het is handig dat je de originality score in Turnitin direct kunt zien als je gaat nakijken. Bij andere systemen moet je een document twee keer openen: om de plagiaat te checken en om na te kijken. In Turnitin kun je tijdens het nakijken de originality score zien en kijken waar de overeenkomsten vandaan komen. (Turnitin)

Tijdens de cursus gaven de studenten elkaar onderling feedback en heb ik geen feedback gegeven. Door de vooraf opgestelde criteria wisten de studenten waaraan de producten moesten voldoen. (WinVision)

Met de vaste set criteria kan ik een grotere hoeveelheid producten nakijken. (Winvision)

De gemelde tijdswinst wordt overigens niet altijd zonder meer als positief gewaardeerd:

Alleen maar minder tijd, omdat ik veel feedback die ik eerst wel zou geven nu heb laten zitten, omdat het niet makkelijk was uit te voeren in Turn it In. (Turnitin)

Meer tijd

Docenten die meer tijd hadden besteed verklaren dat soms uit technische problemen, omdat het systeem niet altijd goed functioneerde:

Het werkt niet altijd, het is niet gebruikersvriendelijk. De studenten snappen het niet, het kost meer uitleg dan het aan tijd oplevert. (WinVision)

Vaker werd als reden aangegeven dat men tijd nodig had om het systeem te leren kennen, ontwikkeltijd nodig had voor het 'vullen' van de tools (aanmaken rubrics, QuickMarks, beoordelingscriteria).. Illustratief is bijvoorbeeld de uitspraak:

Omdat het nog nieuw voor me is, was ik toch wel wat tijd kwijt aan bekend raken met het systeem. Bovendien waren de standaardcommentaren voor het gebied 'taal' niet erg nauwkeurig omschreven, waardoor ik toch wel vaak zelf een opmerking moest aanvullen. Dat kost ook tijd. (Turnitin)

Evenveel tijd

Docenten die evenveel tijd als gewoonlijk nodig hadden waarden over het algemeen wel de voordelen van het systeem op het logistieke vlak, dat geldt voor beide systemen:

Voordeel van het systeem is dat het de mogelijkheid biedt om bestanden op een gestructureerde overzichtelijke manier te bewaren/te plaatsen. Deadlines kunnen makkelijker gehanteerd worden omdat studenten documenten niet meer kunnen plaatsen. Ook is het duidelijk voor hen wanneer er feedback verwacht kan worden. (Turnitin)

De takenlijst die zichzelf opschoont was erg handig. (WinVision)

3.2.2 Relatie systemen/tools

Van de 59 docenten die Turnitin gebruikten heeft 40,7% minder tijd besteed, 45,7% evenveel tijd als daarvoor en 13,6% meer tijd besteed. Van de zes docenten die met WinVision werkten heeft één docent (16,7%) tijd bespaard, twee andere docenten (33,3%) hadden meer tijd nodig.

Tabel 3. Ervaren tijdbesparing docenten bij gebruik Turnitin resp. WinVision (item 12)

	Minder tijd	Evenveel tijd	Meer tijd	Tot.
Turnitin	40,7% (n=24)	45,7% (n=27)	13,6% (n=8)	100% (n=59)
Winvision	16,7% ³ (n=1)	50,0% (n=3)	33,3% (n=2)	100% (n=6)
Overall	38,5% (n=25)	46,2% (n=30)	15,4% (n=10)	100% (n=65)

Turnitin kent vier verschillende tools waarmee feedback gegeven c.q. beoordeeld kan worden, te weten: Rubrics, QuickMarks, GeneralComments en PeerMark. Docenten kunnen één of meerdere van deze tools gebruiken. In onze pilots bepaalden docenten zelf welke tools zij gebruikten, en in welke combinaties. De manier waarop de verschillende tools door de docenten werden gecombineerd verschilt, het meest gebruikt werd de combinatie van drie tools: Rubric/QuickMarks/GeneralComments (zie Tabel 4).

WinVision bevat één tool, die zowel door docenten als studenten (in het kader van peer feedback) kan worden gebruikt: Beoordelingscriteria. In vier van de zes WinVision-pilots werd deze tool door beide groepen gebruikt.

Tabel 4 Meest gebruikte combinaties van tools door docenten.

Combinatie tools	Aantal docenten dat deze combinatie heeft gebruikt
Rubric/QM/GenCom	41
QM/GenCom/PeerMark	11
Rubric/PeerMark	8
Beoordelingscriteria (docent)	2
Beoordelingscriteria (docent en student)	4
	N=66

In Tabel 5 worden de gemiddelde oordelen over de tijdbesparing per tool gepresenteerd. Binnen de combinatie van tools wordt de Rubric relatief als de meest tijdbesparende tool ervaren, GeneralComments en PeerMark worden niet als tijdbesparend gezien.

Tabel 5 Gemiddeld oordeel tijdbesparing afzonderlijke tools (items 22,32,44,56,65)

Tool	n	M*	SD
Rubric	45	3,56	0,99
QM	53	3,32	1,00
GenComm	43	2,80	0,89
PeerMark	16	2,88	1,26

³ Hoewel het hier om kleine aantallen gaat is er toch voor gekozen om ook de percentages te vermelden, aangezien het om relatief grote aantallen gaat binnen de groep.

Beoordcrit	4	3,25	0,96
------------	---	------	------

* 1 = zeer mee oneens, 5 zeer mee eens

Tabel 6 rapporteert de gemiddelde oordelen m.b.t. de gebruiksvriendelijkheid van het systeem.

Tabel 6 Gemiddelde scores items over gebruiksvriendelijkheid van de systemen (items 7,8,9 en 10)

	Turnitin (n=59)		Winvision (n=6)	
	M*	SD	M*	SD
Het systeem..				
wende snel.	3,58	0,93	3,17	1,60
functioneerde naar behoren.	3,34	0,84	3,17	1,72
hielp overzicht houden.	3,66	1,08	3,67	1,03
hielp bij feedback geven/ beoordelen.	3,75	0,98	3,17	1,17

* 1 = zeer mee oneens, 5 zeer mee eens

Verschillen tussen gebruikers Turnitin en WinVision

Vergelijking van de gemiddelde scores op de vier items over de gebruiksvriendelijkheid van de systemen laten iets lagere gemiddelden zien voor de gebruikers van WinVision, maar een t-toets voor onafhankelijke gemiddelden wijst uit dat deze verschillen niet significant zijn. De docenten die Turnitin gebruikten zijn dus niet meer of minder tevreden over de gebruiksvriendelijkheid van het systeem dan de docenten die WinVision hebben gebruikt.

Samenhang tijdbesparing met herhaling

Van de 65 docenten waren er 12 die de cursus uitvoerden als herhaalpilot⁴. Tien van hen gaven onderwijs in een van de drie herhaalpilots met Turnitin op de HvA. Bij de HU waren twee docenten die een cursus verzorgden waarin eerder al feedback werd gegeven c.q. beoordeeld via WinVision. In Tabel 7 wordt de tijdbesparing van de docenten die een pilot voor het eerst uitvoerden en die van de docenten in de herhaalpilots gepresenteerd.

Tabel 7 Tijdbesparing docenten herhaalpilots vs docenten eerste pilot

	N	M*	SD
Eerste pilot	53	2,13	0,68
Herhaalpilot	12	2,67	0,65
Overall	65	2,23	0,70

* Gemiddelde tijdbesparing op een schaal van 1 (meer tijd besteed)-2 (evenveel tijd besteed)-3 minder tijd besteed

De Mann-Whitney U test wijst uit dat de docenten van de herhaalpilots meer tijd hebben bespaard bij het feedback geven c.q. beoordelen dan de docenten voor wie dit de eerste keer was. De docenten in de herhaalpilots hadden minder tijd nodig (gemiddelde rangordescor 44,46) dan de docenten in de eerste pilots (gem. rangordescor 30,41).

⁴ Een herhaalpilot is een cursus die al eerder (in hetzelfde of voorgaande studiejaar) is uitgevoerd in dezelfde opzet en met hetzelfde systeem, al dan niet door dezelfde docenten. In deze herhaalpilots kon dus worden voortgebouwd op ervaring met de tool die al eerder, in dezelfde onderwijscontext, was opgedaan.

Dit verschil is significant ($Z=-2,535; p=0,011$). Van de 12 docenten in de herhaalpilots hebben er 9 tijd bespaard (75%).

3.3 Kwaliteit van de feedback c.q. beoordeling (vraag 2)

De eventuele kwaliteitsverandering als gevolg van het werken met een digitaal systeem voor feedback c.q. beoordeling is onderzocht bij de docenten en studenten.

3.3.1 Kwaliteit volgens docenten

Met betrekking tot de stelling 'De kwaliteit van mijn feedback cq beoordeling werd beter toen ik dit systeem ging gebruiken' zijn de meningen sterk verdeeld: 40,0% is het eens met de stelling, 40,0% antwoordt neutraal en 21,5% is het oneens. (zie Tabel 8).

Tabel 8. Door docenten ervaren kwaliteitsverbetering feedback c.q.beoordeling bij gebruik Turnitin resp. Winvision (item 11)

	(Zeer) eens	Eens noch oneens	(Zeer)oneens	Tot.
Turnitin	38,9% (n=21)	38,9% (n=21)	22,2% (n=12)	100% (n=54)
Winvision	50,0 (n=3)	16,7% (n=1)	33,4% (n=2)	100% (n=6)
Overall	40,0% (n=24)	36,7% (n=22)	23,3 (n=14)	100% (n=60)

Kwaliteitsverbetering

Docenten die vinden dat de kwaliteit verbeterd is lichten dit toe met uitspraken als bijvoorbeeld:

Kwaliteit van feedback verbetert aanzienlijk aangezien je veel meer tekst in de opdrachten als feedback snel kunt plaatsen. (Turnitin)

Het lukt beter om de grote lijn vast te houden en helder te hebben waar je op moet letten!(Turnitin)

Gedeelde feedbacksets met collegadocenten zorgt voor intercollegiale check = verbetering; feedback werd minder individueel toegesneden = geen verbetering; efficiënter en sneller = verbetering. (Turnitin)

Vooraf omdat consequent bij elk verslag dezelfde feedback (Quickmarks) werd gebruikt. Hierdoor gebruik je consequent dezelfde criteria en eenzelfde formulering. (Turnitin)

Dit komt vooral door de criteria die ik voor het systeem ontwikkeld heb. Daarnaast was mijn feedback vooral gericht op de inhoud en de vorm van het product, i.p.v. op het niveau van taalfouten. (WinVision)

Het werken met vooraf opgestelde criteria was erg prettig. De studenten wisten goed waar ze op moesten letten bij het maken van hun producten en het feedback geven op andermans producten. (WinVision)

Vaste beoordelingscriteria beschermen je tegen jezelf om alle fouten aan te strepen. Je beoordeelt nu alle producten op dezelfde wijze. Dat is kwalitatieve winst en tijdswinst. (Winvision)

Geen kwaliteitsverbetering

Sommige docenten die het noch eens noch oneens zijn met de stelling 'De kwaliteit van mijn feedback c.q. beoordeling werd beter toen ik dit systeem ging gebruiken' geven hierbij toelichtingen die duidelijk maken dat ze problemen hadden gehad door uitval van het systeem:

Tegen het einde van de cursus bleek dat er dit jaar problemen waren met Winvision en Internet Explorer. Veel studenten konden geen feedback geven/vragen en liepen telkens vast. Dit heeft voor veel frustratie geleid onder studenten en heeft het proces van peerfeedback uitwisselen belemmerd. Sommigen zijn via de mail feedback gaan geven om toch te kunnen voldoen aan de cursusdoelen.

Herhaaldelijk crashte het programma. Na installatie van Google Chrome ging het beter. (Turnitin)

Anderen gaven aan dat ze de feedbackmogelijkheden van het systeem ontoereikend vonden:

Het is niet mogelijk aanpassingen in de tekst zelf te maken. Dit ervaar ik als een gemis bij het feedback geven. (Turnitin)

Soms ben je door gestandaardiseerde feedback te geven niet genuanceerd genoeg waardoor studenten niet goed begrijpen wat je bedoelt. Dat vind ik een valkuil. (Turnitin)

Voor een van de opdrachten hebben we voor de beoordeling een rubric ontwikkeld, maar het was om technische redenen niet mogelijk om die in Winvision op te nemen.

Overigens was er veel variatie tussen de pilots in de manier waarop de tools werden ingezet, waarbij niet altijd goed te bepalen is in hoeverre de ervaringen het gevolg van het systeem an sich, en in hoeverre ze bepaald zijn door de wijze waarop de tool is ingezet. Een voorbeeld is een docent die geen verbetering ervoer van de feedbackkwaliteit in een cursus waarin QuickMarks waren geformuleerd die zich sterk richtten op tekstuele details:

Ik heb door Turnitin meer feedback op detailniveau gegeven. Doordat het plakken van QuickMarks op de papers zo gemakkelijk gaat gaf ik per paper veel méér commentaar dan voorheen, maar dat was feedback op detailniveau. Door het systeem werd ik de kant van feedback op detailniveau op gestuurd, want als je feedback op structuurproblemen wilt geven, dan moet je de feedback veel meer voor iedere student apart formuleren. Dat soort feedback heb ik nu dus veel minder gegeven dan voorheen.

Ondanks de genoemde problemen en/of minpunten geeft 86,2% van de docenten (56) aan ook in de toekomst met het gebruikte systeem te willen werken (item 13). Van deze 56 docenten werkten er 51 met Turnitin en 5 met Winvision.

Daarbij is de gemiddelde score op de vraag of men vaker met de binnen het systeem gebruikte tools feedback zou willen geven c.q. beoordelen rond de 4 (schaal 1-negatief tot 5-positief), dus er zijn geen tools die hierbij opvallend positief of negatief scoorden (zie Tabel 9).

Tabel 9 Gemiddelde score mbt: ik wil vaker met deze tool feedback geven cq beoordelen (items 25,36,47,57,67).

Ik wil vaker werken met...	N	M*	SD
Rubrics(Tii)	43	3,98	0,94

QuickMarks(Tii)	53	3,85	1,05
GeneralComments(Tii)	43	3,67	0,94
PeerMark	14	3,92	0,73
Beoordelingscriteria (WV)	7	4,57	0,53

* 1 = zeer mee oneens, 5 zeer mee eens

Verschillen tussen gebruikers Turnitin en WinVision

Een t-toets voor onafhankelijke gemiddelden levert geen significante verschillen op tussen de groep Turnitin- en WinVisiongebruikers ten aanzien van de scores op item 11 ('De kwaliteit van mijn feedback cq beoordeling werd beter toen ik dit systeem ging gebruiken'). De docenten die Turnitin gebruikten zijn dus niet meer of minder tevreden over de kwaliteit van hun feedback c.q. beoordeling dan de docenten die WinVision hebben gebruikt.

Samenhang feedback-c.q.beoordelingskwaliteit in relatie tot tijdbesparing (item 11, 12)

Uit een correlatieberekening (Spearman's rho) blijkt dat er geen samenhang is tussen het oordeel over tijdbesparing en het oordeel over de feedback- c.q. beoordelingskwaliteit. Dat betekent dat een eventuele tijdbesparing niet ten koste is gegaan van de feedback- c.q. beoordelingskwaliteit.

3.3.2 Kwaliteit volgens studenten

De studenten geven gemiddeld het rapportcijfer 7,1 (SD 1,38) voor de feedback die ze op hun werk hebben gekregen. Bijna driekwart (71,1%) vindt de hoeveelheid feedback voldoende. Degenen die met Turnitin hadden gewerkt hebben geen problemen ondervonden met het inleveren van hun werk en het opzoeken van de feedback. Bij drie van de zes pilots met WinVision gaf dit wel problemen omdat het systeem op cruciale momenten niet werkte. De studenten is ook gevraagd naar hun oordeel over de afzonderlijke tools, voor de resultaten per pilot zie de overzichtstabellen in de bijlagen 4, 5, 6 en 7.

Rubric⁵

Van de 291 studenten die met een rubric zijn beoordeeld wil 81,8% vaker op deze manier worden beoordeeld. Uitspraken die dit illustreren zijn bijvoorbeeld:

Er staat duidelijk in wat er van je wordt verwacht en wat er in je product moet komen te staan dus dat is goed.

Door de rubrics is het duidelijk waar de focus ligt en hoe je dat aan kan passen en verbeteren. Erg handig.

Je wist precies wat je moest doen om een hoger cijfer te krijgen en hoe jouw werk op dat moment was door te lezen in welke rubric je valt.

Anderen maken een voorbehoud in de trant van 'de rubric moet dan wel duidelijker zijn dan nu het geval was':

De rubric werkt wel als de richtlijnen en criteria in de rubric concreet zijn en niet abstract blijven.

Ik vind dat de rubric te veel verschillende onderdelen onder elk kopje had, waardoor je soms voor het ene onderdeel goed scoort en voor het andere minder goed, zonder dat hier onderscheid in wordt gemaakt.

⁵ Deze tool zit alleen in Turnitin.

Ik vind rubrics over het algemeen goed werken maar deze was inhoudelijk misschien wat minder: meerdere factoren in een beoordeling en niet makkelijk objectief te beoordelen. (Turnitin)

Een rubric alléén geeft volgens sommigen te weinig houvast:

Voor mij was de Rubric een handig middel om erbij te hebben bij het schrijven van de opdracht, omdat het duidelijk aangeeft wat er in ieder gedeelte van mij verwacht wordt. Maar als beoordeling heb ik meer aan constructieve korte opmerkingen dan een globaal beeld (de categorieën) van op welk niveau ik zit. De handigste feedback waren de opmerkingen (comments) in de tekst.

Hiermee kun je zien wat je sterke en zwakke onderdelen van een betoog zijn. Het is wel noodzakelijk dat er bij wordt gezet wat er dan precies goed of fout aan is anders kom je nog niet veel verder.

De rubric is een prettige manier om feedback te krijgen, ik vond het alleen niet altijd duidelijk welke aantekening bij welk stukje tekst hoorde. Voor mij mochten de aantekeningen dus wat beter aangegeven worden.

QuickMarks⁶

Van de 391 studenten die via QuickMarks (meestal in combinatie met een rubric) feedback in de tekst hebben gekregen zou 81,3% vaker op deze manier feedback willen krijgen. Enkele toelichtingen:

Feedback in de tekst heb ik als duidelijker en handiger ervaren dan de RUBRIC feedback, die feedback blijft erg algemeen en als je pech hebt krijg je ook nog eens twee 'slechte' specifieke suggesties waardoor je er eigenlijk niets mee kunt. Door cijfers te geven zoals in RUBRIC weet je wel dat er iets mis is maar niet precies wat.

Een rubric is wat oppervlakkiger, feedback in de tekst geeft echt specifieke aanwijzingen. Ik vind het fijn als iemand echt punten benoemd van wat beter moet of veranderd moet worden. Vaak zijn dat ook de dingen namelijk waar je zelf over heen kijkt.

Dit vond ik een prettige manier van feedback krijgen. Het is erg duidelijk met de ballonnen en dergelijke. Mijn tekst is sterk verbeterd na het verkrijgen en verwerken van de feedback.

Het werkt prettig, het is overzichtelijk en duidelijk. Ik denk dat het voor de docenten ook efficiënt is omdat deze op dezelfde manier op werk feedback moeten geven en naar dezelfde criteria moeten kijken.

Maar niet alle studenten zijn enthousiast over de QuickMarks:

Lastig dat de feedback pas aan het einde uitgeschreven werd. Hierdoor moest je telkens heen en weer bladeren. Medestudenten hadden geschreven feedback gekregen direct in de tekst en dit leek mij toch nog handiger.

Wanneer de tekst wordt aangepast in een Word document kan je dat direct veranderen. Nu moet je telkens switchen tussen je internetscherm en je tekstverwerker. Dit is erg irritant en onduidelijk. Daarnaast stond er vaak ook een comment waarvan niet duidelijk was op welk gedeelte van de tekst dit sloeg.

Het is vaak niet duidelijk op welke woorden de feedback precies slaat. Je kan dat niet precies genoeg aangeven.

GeneralComments⁷

Van de 426 studenten die commentaar op hun werk via GeneralComments kregen (meestal in aanvulling op de rubric en QuickMarks) kreeg tweederde (62,7%) daardoor een goed totaalbeeld van het werk. Sommige studenten verklaren dat ze dergelijk

⁶ Deze tool zit alleen in Turnitin.

⁷ Deze tool zit alleen in Turnitin.

afsluitend commentaar misten en maar al te graag hadden gehad, vooral omdat de rubric en QuickMarks volgens hen uitnodigen tot vooral negatieve feedback:

Ik heb niet het idee dat ik een goed samenvattend afsluitend commentaar heb gekregen. Dit had ik graag wel gehad, want dat lijkt me heel verduidelijkend en misschien ook motiverend voor hoe je het beste verder kan gaan.

Ik heb veel te lezen gekregen wat ik niet goed had gedaan, maar niet hoe ik het wel goed moest doen. Hier liep ik nog wel tegen aan.

Vertel wat er goed is, indicatie of het al op een voldoende zit of dat er nog heel veel aan moet gebeuren.

Beoordelingscriteria⁸

Van de 92 studenten die hun beoordeling aan de hand van beoordelingscriteria in WinVision hadden gekregen wil ongeveer de helft (48,9%) vaker op deze manier worden beoordeeld. Een paar uitspraken van studenten illustreren waarom zij deze manier van beoordelen positief waarderen:

Het werd duidelijk beoordeeld en je had het meteen op je laptop staan waardoor het makkelijk te verwerken was. Dit werkte heel goed voor mij.

Ik vond het een fijne manier van beoordelen. Er werd duidelijk bijgezet wat ik fout deed, waarom het fout was en wat ik zou kunnen doen om het te verbeteren. Ik had 1 punt wat ik onduidelijk vond maar in de les was er ruimte om hier vragen over te stellen.

Deze manier van beoordelen bevalt mij erg, omdat ik hierdoor een duidelijk beeld krijg van mijn sterke punten en punten waar ik nog aan moet werken. Hierdoor word ik gemotiveerd om mijzelf te ontwikkelen.

Het is erg makkelijk voor mij om mijn werk via winvision in te leveren. Ik hoef namelijk niks meer uit te printen en/of met losse blaadjes rond te lopen, wat er voor zorgt dat ik mijn feedback formulier of juist mijn werk sneller kwijt raak.

Negatieve waarderingen verwijzen behalve naar uitval van het systeem soms ook naar gebrek aan instructie in het systeem:

Ik heb via WinVision geen feedback kunnen geven en vragen. Dit omdat het programma bij mij via Chrome niet werkte, terwijl dit het probleem had moeten zijn. Ik ben totaal niet te spreken over het programma en heb er dus ook weinig van gezien.

Het digitaal inleveren en feedback krijgen is ideaal, maar Winvision zelf vind ik persoonlijk te gebruiksonvriendelijk om goed te kunnen gebruiken.

Een digitale beoordeling is prima, maar het moet van tevoren duidelijk zijn hoe je de vraag naar een beoordeling in het digitale portfolio zet en hoe je de beoordeling weer terug krijgt. Op winvision is het soms erg lastig terug te vinden. Een melding over goed ontvangst/verstuur bericht etc. zal helpen om dit te checken.

Peer feedback⁹

Van de 204 studenten die met peer feedback hebben gewerkt (waarvan 130 via PeerMark en 74 in WinVision (via het toepassen van de beoordelingscriteria op het werk van medestudenten) vond 46,8 % de feedback van hun medestudenten nuttig om hun eigen werk te verbeteren (item 38). Voor het zelf geven van feedback op andermans

⁸ Het betreft hier de tool Beoordelingscriteria in WinVision.

⁹ De tool PeerFeedback zit alleen in Turnitin, maar WinVision kent wel de mogelijkheid om de tool Beoordelingscriteria te gebruiken voor peer feedback.

werk ligt dit percentage op 51,2% (item 43). Ruim de helft (54,4%) had daarnaast graag meer feedback gehad van de docent (item 46). Eveneens de helft (50,8%) zou vaker met (online) peer feedback willen werken (item 48). De waarderingen, uitgesplitst voor Turnitin en WinVision, worden gepresenteerd in Tabel 10.

Tabel 10 Gemiddelde scores items over het door studenten ervaren nut van peer feedback (items 38,43,46 en 48)

	Turnitin (n=130)		Winvision (n=74)	
	M*	SD	M*	SD
Item 38	3,42	0,95	3,64	0,86
Item 43	3,51	0,95	3,76	0,73
Item 46	3,84	1,00	3,76	0,82
Item 48	3,62	0,89	3,36	0,99

* 1= zeer mee oneens, 5 zeer mee eens

Positieve waarderingen voor het werken met peer feedback gaan wel vaak gepaard met aanwijzingen voor verbetering:

Geef aan waar de studenten precies op moeten letten, dit werd goed gedaan in GAWWAG). (Turnitin)

Ik vind peer feedback bij het schrijven van een paper een nuttige bezigheid. Maar bij de procedure moet mijns inziens iets veranderd worden. Als de student voor de eerste keer feedback geeft krijgt hij soms papers die niet zijn afgewerkt. Omdat alleen de studenten elkaar beoordelen, en niet de docenten, schrijven sommige studenten deze eerste paper niet serieus. Conclusies ontbreken of stukken zijn afgeraffeld. Op die manier is het moeilijk om goede feedback te geven, een eerste poging die totaal niet serieus genomen is, valt moeilijk te beoordelen. Ik zou als verbetering/tip willen aanbevelen dat er een systeem komt waarbij ook de eerste poging (met enkel feedback van studenten) serieuzer kan worden genomen. (Turnitin)

Door het zelf beoordelen van andermans werk aan de hand van een aantal criteria is het makkelijker om bij volgende opdrachten ook tijdens het schrijven eigen werk te beoordelen. Eigen fouten worden ook duidelijker. Peer feedback geven moet echter niet teveel tijd in beslag nemen: op die manier wordt er toch nog genoeg aandacht aan besteed, en de beoordeling van de docent is uiteindelijk de enige die telt. (Turnitin)

Maak duidelijk dat niet alleen de beoordelingstabel ingevuld moet worden, maar dat er ook een toelichting gegeven moet worden. (WinVision)

Negatief:

Persoonlijk heb ik liever dat de docent feedback geeft, omdat dat gegarandeerd kwalitatieve feedback is. Als een student feedback vraagt aan medestudenten die minder goed zijn in het onderwerp, kan deze medestudent het werk verkeerd beoordelen, bijv. door goede zinnen aan te geven als slecht geformuleerd. (WinVision)

Ik vond het jammer dat de peer feedback van mijn medestudenten vaak helemaal niet aansloot met de peer feedback van de docent, dit maakte het erg verwarrend. (Turnitin)

Ik vond de kwaliteit van de peer feedback sterk wisselen. Ik had er in eerste instantie heel veel tijd aan besteed en toen kreeg ik echt minimaal commentaar terug, dat vond ik jammer en de peer

feedback werd ook niet door de docenten bekeken/beoordeeld dus in feite kon je gewoon maar wat doen. (Turnitin)

Maak duidelijk dat niet alleen de beoordelingstabel ingevuld moet worden, maar dat er ook een toelichting gegeven moet worden. (WinVision)

3.4 Eindoordeel volgens docenten en studenten

Het eindoordeel van de docenten over het digitaal feedback geven c.q. beoordelen via de door hen gebruikte systemen is positief: 86,2% (56) geeft aan ook in de toekomst met het gebruikte systeem te willen werken. Van deze 56 docenten werkten er 51 met Turnitin en 5 met Winvision. Voor beide systemen geldt dat niet alleen de specifieke tools positief gewaardeerd worden, maar ook de logistieke ondersteuning die het systeem biedt bij het inleveren en archiveren.

De handleiding die voor het gebruik van Winvision is ontwikkeld, is door de docenten als zeer bruikbaar ervaren. Er blijkt alleen wel meer tijd nodig te zijn in de cursus om de studenten met de handleiding te laten werken.

Het eindoordeel van de studenten die hun feedback c.q. beoordeling kregen via rubric en QuickMarks (beide tools van Turnitin, meestal in combinatie gebruikt) is eveneens positief: een ruime meerderheid zou vaker op diezelfde manier hun feedback c.q. beoordeling willen krijgen. De studenten die via WinVision hun feedback c.q. beoordeling hadden ontvangen zijn minder positief: de helft wil dit vaker op deze manier. Dit heeft vooral met de in hun ogen tekortkomende gebruiksvriendelijkheid van het systeem te maken en/of het gebrek aan instructie.

4. Conclusie en discussie

4.1 Conclusies

In dit onderzoek zijn 26 pilots met de beschreven onderzoeksinstrumenten geëvalueerd. Het ging hierbij om reguliere onderwijsmodules waarin op uiteenlopende wijzen digitaal feedback is gegeven op studentwerkstukken, c.q. werkstukken digitaal zijn beoordeeld.

Het eindoordeel van docenten over het digitaal feedback geven c.q. beoordelen is positief; een grote meerderheid (86%) geeft aan ook in de toekomst met het gebruikte systeem te willen werken. Ook van de studenten die zijn beoordeeld met Turnitin is de ruime meerderheid positief, zij willen in de toekomst vaker met het systeem werken. De studenten die met WinVision zijn beoordeeld, zijn iets kritischer: ongeveer de helft wil vaker via dit systeem beoordeeld worden of feedback ontvangen. Dit is te wijten aan de tekortschietende gebruiksvriendelijkheid van het systeem en/of het gebrek aan instructie.

In dit rapport willen we antwoord geven op de volgende twee onderzoeksvragen:

1. In hoeverre heeft het gebruik van een digitale feedback- en beoordelingstool bijgedragen aan tijdwinst voor de docent bij het feedback geven en beoordelen?
2. In hoeverre heeft het gebruik van een digitale feedback- en beoordelingstool bijgedragen aan een verbetering van de feedback-en beoordelingskwaliteit?

De resultaten ten aanzien van de twee onderzoeksvragen worden hieronder per onderzoeksvraag samengevat:

4.1.1 Tijdbesparing

1-Van de docenten heeft 38,5% minder dan de voor hem/haar gebruikelijke tijd besteed bij het feedback geven c.q. beoordelen. De opgegeven tijdbesparing loopt daarbij uiteen van 10 tot 50%. 46,2% heeft evenveel tijd als gewoonlijk nodig gehad en 15,4% heeft meer tijd moeten investeren. Dat er meer tijd besteed is komt vooral door de investering die nodig was om inhoudelijk vulling te geven aan de tools, dit geldt met name voor de rubrics, beoordelingscriteria en QuickMarks.

2- Ervaring in het werken met het systeem lijkt te lonen: docenten die een cursus uitvoerden als een 'herhaalpilot' gaven vaker aan dat ze tijdswinst hadden geboekt dan docenten in cursussen waarin de tool voor het eerst werd ingezet. Dit zal er deels aan liggen dat docenten van herhaalpilots over het algemeen al vertrouwd waren met de nakijktools en dus minder scholings- of inwerktijd nodig hadden om met het systeem te leren werken. De belangrijkste reden zal zijn dat ze konden profiteren van het ontwikkelwerk dat zijzelf of hun collega's hebben moeten uitvoeren om de tools inhoudelijk te vullen.

We concluderen dan ook dat digitaal feedback geven en beoordelen inderdaad kan leiden tot tijdswinst voor docenten; bijna veertig procent van de docenten ervoer al bij de eerste keer dat deze werkwijze werd toegepast tijdswinst en dit neemt toe als deze werkwijze in latere jaren herhaald wordt.

4.1.2 Kwaliteit feedback-c.q. beoordeling

3-Volgens 40,0 % van de docenten is de kwaliteit van hun feedback c.q. beoordeling verbeterd. De docenten beschouwen het als een verbetering dat er uniformer wordt beoordeeld omdat het hele docententeam gebruik maakt van dezelfde rubric, QM of beoordelingscriteria. Ze vinden het een verbetering dat voor de beoordeling van schrijfproducten consequent dezelfde criteria worden gebruikt en dat de beoordeling transparanter is geworden voor de studenten. De docenten die geen kwaliteitsverbetering zagen waren vaak niet tevreden over de mogelijkheden of de inhoudelijke invulling van de tools, of hadden technische problemen gehad met het systeem.

4- Docenten die tijd bespaard hebben bij het beoordelen c.q. feedback geven vinden niet dat de kwaliteit van hun feedback c.q. beoordeling erop achteruit is gegaan. Met andere woorden: ze hebben hun tijdswinst niet moeten 'bekopen' met een achteruitgang in feedbackkwaliteit.

5- De studenten zijn over het algemeen tevreden over de feedback die ze hebben gekregen. Ze waarderen de kwaliteit daarvan met gemiddeld het rapportcijfer 7,1 (SD 1,38). Bijna driekwart (71 %) vindt de hoeveelheid feedback voldoende.

We concluderen dan ook dat het efficiënter nakijken met digitale hulpmiddelen in de ogen van de docenten niet leidt tot kwaliteitsverlies en dat deze wijze van feedback krijgen en beoordeeld worden ook voor studenten acceptabel is.

4.2 Discussie

De in dit onderzoeksrapport gepresenteerde resultaten maken duidelijk dat digitaal feedback geven en beoordelen –wanneer de benodigde investeringen eenmaal gedaan zijn- helpt om de werkdruk van docenten beheersbaar te houden, zonder dat dit ten koste gaat van de feedbackkwaliteit zoals gepercipieerd door docenten. Over de vraag of dat het beste via Turnitin of WinVision kan verlopen doen wij geen uitspraak, binnen het SCALA-project wilden we onderzoeken of de digitale systemen leiden tot tijdswinst en

voldoende kwaliteit, het was niet ons doel om beide systemen te vergelijken. Bovendien is het aantal pilots dat met WinVision is uitgevoerd te klein in verhouding tot het aantal pilots waarin met Turnitin is gewerkt.

Ons onderzoek, dat is uitgevoerd als onderdeel van een opschalingsproject, kent een aantal haken en ogen, die van belang zijn bij de interpretatie van de gegevens.

De wijze van gebruik van de systemen was niet altijd ideaal

Zoals gezegd gaat het om pilots die zijn uitgevoerd door gewone docenten, vaak als hun eerste kennismaking met Turnitin dan wel WinVision, in alledaagse onderwijscontexten met de onvermijdelijke malheur die daarbij komt kijken (tijdgebrek, last minute wijzigingen, verschil van interpretatie door collega's, technische problemen, etc.). Dit maakt dat de feedback- en nakijktools niet altijd op de meest doordachte of best mogelijke manieren zijn ingezet. Uit de per instelling gegeven overzichten van de uitkomsten per pilot is af te lezen dat niet in alle pilots alles naar wens is verlopen. We geven een aantal voorbeelden van dit soort onvolkomenheden:

- Het kwam voor dat studenten werden beoordeeld met een rubric die nauwelijks was ingevuld. Er werd soms volstaan met het invullen van de categorie 'voldoende' met de punten die er op elk criterium te verdienen zijn. Op zich is dat begrijpelijk omdat in veel pilots de rubric alleen voor summatieve doeleinden werd gebruikt en het ontwikkelen van een volledige rubric (met beschrijvingen van prestatieniveaus per criterium) veel tijd en afstemming vraagt. Maar op die manier is de rubric in feite hetzelfde als een beoordelingsformulier, waarbij niet het maximale uit het instrument gehaald wordt.
- Docenten hadden soms bepaalde functies nog niet ontdekt. Hoewel het een goed verborgen functie is heeft Turnitin de mogelijkheid om met één druk op de knop te archiveren, waardoor alle ingeleverde opdrachten en de feedback in een zipbestand opgeslagen worden.
- Er waren ook docenten die –achteraf gezien- te creatief omgingen met het systeem: zo transformeerde een docent de rubric tot louter een scoringsinstrument dat via PeerMark gebruikt kon worden in het kader van peer feedback. Hiermee wordt het potentiële leereffect van peer feedback de das omgedaan: in de versie van PeerMark zoals die in Turnitin zit wordt juist de toelichting van de beoordeling uitgelokt en worden de studenten gestimuleerd om specifiek commentaar te geven in de tekst.
- Ook bleken er misvattingen te bestaan over wat het systeem wel en niet kan. Zo wijten sommige studenten het aan het systeem dat zij alleen negatieve feedback kregen: de rubric en QuickMarks zouden daartoe uitnodigen. Dat is niet het geval: in een goede rubric ziet een student juist ook wat er al goed gaat en QuickMarks kunnen zowel negatief als positief geformuleerd worden.

Het ging, met andere woorden, in deze pilots niet om 'strak' ontworpen interventies, maar om 'gewoon gebruik' in normale onderwijssettings. Dit heeft een conceptueel nadeel: de uitkomsten tonen niet de volledige potentie van het nakijken en beoordelen door middel van deze systemen. Evenmin wordt exact duidelijk wat precies de merites zijn van de verschillende nakijktools, afzonderlijk of in bepaalde combinaties gebruikt. Om dat soort vragen te kunnen beantwoorden, hadden de implementaties meer systematisch ontworpen moeten worden.

Bezien door een pragmatische bril, heeft onze aanpak ook een voordeel. Omdat de resultaten verkregen zijn in settings die heel 'normaal' zijn, is te verwachten dat de opbrengsten, in termen van tijdswinst en het op peil houden van de kwaliteit van feedback en beoordeling, vergelijkbaar zal zijn wanneer het gebruik van deze systemen verder verbreed zal worden.

Een vraag die zich hierbij opdringt, is of het verschil uitmaakt dat we hier van doen hadden met gebruik door 'early adopters'. De docenten – of tenminste de cursuscoördinatoren- hebben zich vrijwillig aangemeld en waren bereid om over de onvermijdelijke kinderziektes van de systemen heen te kijken. De vraag is of het gebruik van de systemen anders, of moeizamer, zal zijn als een systeem instellingsbreed wordt geïmplementeerd, zoals men nu bij de Universiteit Utrecht van plan is. Ervaringen tot nu toe bij de VU, waar Turnitin 'low profile' als een standaardvoorziening wordt aangeboden– met bijbehorende ondersteuningsmogelijkheden- stemmen optimistisch: er zijn bij de VU inmiddels 600 actieve gebruikers (docenten) geregistreerd die er vrijwillig aan begonnen zijn. De brede acceptatie van het systeem verloopt bij de VU vrij geruisloos.

Methodologische beperkingen

Er was veel verschil tussen de wijze waarop in de verschillende pilots gebruik is gemaakt van de feedback en nakijksystemen en soms bleken ook docenten binnen één pilot het systeem op uiteenlopende wijze te gebruiken. Dit maakt het moeilijker om algemene conclusies te trekken uit de (kwantitatieve) evaluatiegegevens.

Zo lopen de gemiddelde rapportcijfers die de studenten geven voor de kwaliteit van de feedback c.q. beoordeling nogal uiteen, in sommige pilots ligt het gemiddelde bij een zes, bij andere rond de acht. Bij een pilot kon een laag rapportcijfer worden verklaard door het feit dat de betreffende docent geen instructie over het werken met WinVision had gegeven aan de studenten, maar volstond met het uitreiken van de handleiding. Het bleek dat de studenten er niet zelfstandig mee uit de voeten konden en niet wisten waar ze hun feedback c.q. beoordeling konden vinden. Er is ook een pilot met het rapportcijfer negen, maar dat cijfer heeft niets te maken met digitaal nakijken. De docent had hier namelijk al vrij snel besloten om haar (enige) student niet via Turnitin feedback te geven maar via de reviewoptie van Word. Zij koos daarvoor omdat ze zo met vorige versies kon vergelijken. Kortom: het rapportcijfer zegt niet per se iets over de kwaliteit van de via Turnitin of WinVision gegeven beoordeling c.q. feedback maar waarschijnlijk meer over de algehele tevredenheid van de studenten met de cursus.

Onze evaluatie richtte zich op de gepercipieerde tijdsinstroom door de docent en de beleving van de kwaliteit van feedback en beoordeling door docenten en studenten. Soms was er sprake van een verschil tussen perceptie van de feedback-c.q. beoordelingskwaliteit door de docenten (zijzelf vinden die verbeterd) en die van de studenten (die uiteraard niet kunnen vergelijken), die er het rapportcijfer 6,2 aan toekennen. Of hier werkelijk een verbeteringsslag heeft plaatsgevonden is lastig te zeggen, daarvoor zouden we de studentevaluaties van diezelfde cursus in voorgaande jaren moeten raadplegen.

De terminologie in de vragenlijst kan voor verwarring gezorgd hebben. Het komt voor dat studenten de vragen over de rubric hebben overgeslagen, terwijl ze in de desbetreffende pilot wel met een rubric werden beoordeeld. Ze hadden geen rubric herkend of ze hadden die misschien wel gezien maar waren vergeten wat het was of dat het 'rubric' heette.

Aanbevelingen

Uit de resultaten blijkt dat digitaal beoordelen en feedback geven met dit soort systemen een efficiënte werkwijze kan zijn die kwalitatief voldoende kan zijn. Uit de reacties van studenten blijkt wel dat de normale kwaliteitscriteria voor het geven van feedback en beoordelingen ook nu aandacht behoeven, zoals: het geven van positieve feedback, het zorgen voor voldoende evenwicht tussen specifieke, persoonlijke feedback en feedback die duidelijk maakt waar de student staat ten opzichte van het verwachte eindniveau. Met name het gebruik van de rubrics verdient aandacht. De docentevaluatie suggereert dat met name door rubrics te gebruiken, tijdsinstroom te behalen valt. Om de rubric goed te benutten, ook in formatieve zin, is er meer nodig dan het simpel overzetten van beoordelingscriteria. Om studenten te informeren over hun actuele prestatieniveau en de weg die ze nog moeten afleggen naar het gewenste niveau is het nodig om te investeren

in het invullen van de rubric. Deze functie kan immers niet door de QuickMarks worden vervuld. Dit is dus nog een weg te gaan: hoe zet je rubrics in? Hoe uitgewerkt moeten ze zijn gegeven het doel? Welke combinaties van tools zijn hiervoor effectief?

Een nieuwe faciliteit van Turnitin, die onlangs is toegevoegd, geeft de mogelijkheid om de QuickMarks te relateren aan de rubric: bij het aanklikken van een QuickMark wordt het bijbehorende criterium uit de rubric ook zichtbaar. Dit lijkt een welkome verbetering, die wellicht kan helpen om de QuickMark- en rubricfunctie in Turnitin beter op elkaar af te stemmen.

5. Literatuur

- Chanock, K. (2000). Comments on Essays: do students understand what tutors write? *Teaching in Higher Education*, 5(1),95-105.
- Gibbs (1999) Using assessment strategically to change the way students learn. In: S. Brown, & A. Glasner (Eds.), *Assessment Matters in Higher Education*. SRHE/Open University Press.
- Gibbs (2010). *Using assessment to support student learning*. Leeds: Leeds MetPress.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112.
- Ramsden, P. (2003). *Learning to Teach in Higher Education*. Routledge Farmer, Abingdon.
- Sainsbury, & Walker (2008). Assessment as a vehicle for learning: Extending collaboration into testing. *Assessment and Evaluation in Higher Education*, 33(2), 103–117.
- Shute, V.J. (2008). Focus on formative feedback. *Review of educational research*, 78(1), 153-189.
- Weaver, M.R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education*, 31(3), 379–394.

6. Bijlagen

- 6.1 Docentvragenlijst**
- 6.2 Studentvragenlijst**
- 6.3 Interviewleidraad**
- 6.4 Resultaten VU**
- 6.5 Resultaten UU**
- 6.6 Resultaten HU**
- 6.7 Resultaten HvA**