

“Meneer, hier kan ik toch niets mee?”

Een praktijkgericht onderzoek naar percepties over klassikale mondelinge feedback.

Laurens Janssen, Anco Hoogeveen, Alje van den Bosch, Jesse van de Hulsbeek

Begeleiders: Felix van Vugt, Jan Vermunt

Abstract:

Onderzoek heeft aangetoond dat een docent zijn leerlingen verder kan helpen in hun ontwikkeling door goede en gerichte feedback. Zijn docenten en leerlingen het in de praktijk echter wel met elkaar eens wat betreft de effectiviteit van gegeven feedback en de overdracht daarvan? En hoe zit dat met een externe observator? De resultaten van een door docenten, leerlingen en observatoren ingevulde vragenlijst suggereren dat deze drie partijen het onder andere oneens zijn over de duidelijkheid, de terechtheid en de leerzaamheid van feedback. De wijze waarop de percepties verschillen is echter verrassend.

Inleiding

Relevantie onderzoek

Academisch onderzoek naar feedback heeft al verschillende inzichten opgeleverd over de effectiviteit van feedback. Enkele studies, hieronder beschreven in het theoretisch kader, tonen het belang van goede feedback aan. Daarnaast doen onderzoekers uitspraken over wat wel en niet effectief is. Er is ook aandacht geweest voor hoe leerlingen gegeven feedback ervaren: positieve feedback motiveert en bevordert leerprestaties.

Er zijn echter nog een aantal interessante invalshoeken waarin deze onderzoeken niet voorzien. Hoewel er onderzoeken zijn geweest die de manier in kaart hebben gebracht waarop leerlingen (Thurlings et al., 2010) en docenten feedback ervaren, zijn die resultaten niet met elkaar vergeleken. Een vergelijking zou interessante implicaties kunnen hebben, zeker wanneer een externe observator ook zijn ervaring van de feedback beschrijft en in de vergelijking wordt meegenomen. Wanneer leerlingen, docenten en observatoren het met elkaar eens zijn wat betreft de feedback zou dat suggereren dat er een eenduidig beeld is over de kwaliteit van feedback en de overdracht daarvan. Bovendien zou het in verder onderzoek niet meer nodig zijn de verschillende groepen allemaal te ondervragen en/of analyseren. In plaats daarvan zou bijvoorbeeld een analyse van een externe observator genoeg kunnen zijn om iets te zeggen over de manier waarop docenten in de praktijk feedback geven. Dit zou het vervolgonderzoek vergemakkelijken.

Het is ook mogelijk dat de percepties maar deels of niet overeenkomen. Ook dit draagt bij aan eventueel verder onderzoek: de onderzoeker is zich ervan bewust dat hij er rekening mee moet houden dat leerlingen, observatoren en docenten feedback verschillend ervaren. Daarnaast werpt het de vraag op in hoeverre het genoeg is om alleen te kijken naar academische bronnen over feedback wanneer deze bijvoorbeeld exclusief vanuit het perspectief van een externe observator is geschreven. Bovendien zouden de verschillende resultaten een mooie basis voor vervolgonderzoek zijn. Het is bijvoorbeeld mooi om vast te stellen dat de percepties van de groepen uit elkaar lopen, maar nog veel interessanter om te analyseren waarom dat precies zo is.

Theoretisch kader

Docenten gebruiken feedback in het onderwijs om het leerresultaat van de leerling te vergroten (Sol en Stokking, 2009). John Hattie noemt in zijn grote metaonderzoek feedback geven één van de meest effectieve onderwijsstrategieën (Hattie en Timperley, 2007). Feedback werkt, mits op de juiste manier gegeven, motiverend op het schoolwerk en bevordert leerprestaties (Van Rijs, 2011).

Feedback is commentaar op een handeling, zij het in positieve of negatieve zin. Effectieve feedback geeft de ontvanger een handvat om de kloof tussen hetgeen hij wél begrijpt en hetgeen hij zich nog niet eigen gemaakt heeft te overbruggen. Recent onderzoek onderbouwt het idee dat kwalitatieve feedback meer kan opleveren dan het geven van een cijfer. Sterker nog, leerlingen leren beter en zijn gemotiveerder als ze effectieve feedback krijgen dan wanneer ze alleen maar een cijfer krijgen. Kwalitatief goede feedback is dus erg belangrijk in het onderwijssysteem (Brookhart, 2008; Hattie en Timperley, 2007).

Iedere leerling zal een specifieke vorm van begeleiding en feedback vragen. Daar moet een docent flexibel op kunnen insprijgen. Feedback moet voor alles positief en constructief zijn. De docenten willen graag rekening houden met het niveau, het zelfvertrouwen en de persoonlijke situatie van de leerling. Ze vinden het belangrijk om leerlingen er actief bij te betrekken. Ze zijn verheugd als leerlingen hun handelen positief waarderen (Sol, Stokking en Vermunt, 2009). Docenten zijn bij het

geven van feedback sterk gericht op de leerling. Ze zijn betrokken met hun leerlingen en willen maatwerk leveren. Feedback moet volgens hen worden gegeven in persoonlijk contact, in een situatie van vertrouwen en met voldoende tijd en aandacht. Ook moet er rekening worden gehouden met het niveau en het zelfvertrouwen van de individuele leerling, en is het belangrijk dat de docent bevestigt wat goed gaat en wat de leerling kan.

In het concrete feedbackgedrag van de docenten is hun gerichtheid op het geven van positieve feedback duidelijk zichtbaar. Ze geven zowel resultaat - als procesgerichte feedback en ook aanvullende instructie en uitleg, maar gaan naar verhouding niet zo veel in op de inhoudelijke voorwaarden waar het werk aan moet voldoen. Zij reageren op problemen van leerlingen die iets niet begrijpen en daardoor niet verder kunnen, waarbij hun feedback erop gericht is dat de leerlingen de stof begrijpen en weer verder kunnen. (Sol en Stokking, 2009).

Uit onderzoek blijkt dat leerlingen over het algemeen tevreden zijn over de feedback die ze van hun docenten krijgen. Zij vinden deze positief, duidelijk, terecht, leerzaam en bruikbaar. Zij signaleren wel dat het accent vaak ligt op wat goed is of gaat, maar bij sommige docenten juist op wat niet goed is of gaat. Daarbij zien ze meer resultaatgerichte feedback (doet de leerling het goed?) dan procesgerichte (hoe doet de leerling het?). Mogelijk zijn de leerlingen zelf wat meer gericht op resultaatgerichte feedback, en wellicht is procesgerichte feedback ook een meer natuurlijk onderdeel van de verdere interactie en wordt het daardoor minder opgemerkt. De leerlingen willen verder dat hun docenten meer of vaker verschillende vormen van feedback geven dan nu gebeurt, maar dit verschilt wel tussen docenten (Sol en Stokking, 2009).

Uit onderzoek blijkt verder dat leraren nog onvoldoende zijn voorbereid op het aanleren van zelfsturende vaardigheden (bijvoorbeeld het zelfstandig aan de slag gaan of het plannen) aan leerlingen. Het ontbreekt volgens onderzoekster Helen Jossberger (2011) leerlingen nog aan voldoende zelfsturende vaardigheden. Goede feedback biedt volgens haar mogelijkheden leerlingen uit te dagen na te denken over hun eigen leerproces om zodoende zelfsturende vaardigheden te vergroten).

Net als leerlingen zitten ook leraren in een ontwikkelingsproces. Leraren blijven als leden van een lerende schoolorganisatie ook zelf leren. Om greep te krijgen op dit proces blijkt zelfevaluatie een effectieve aanpak. Door leerlingen te betrekken in het proces van evaluatie (bijvoorbeeld bij het geven van feedback) wordt voor veel leraren bijzonder duidelijk wat er belangrijk is in de manier waarop ze in de klas handelen en hoe de feedback overkomt bij de leerling, zoals hierboven aangegeven is (Van Petegem en Deneire, 2006).

Van Petegem en Deneire stellen dat leraren op het gebied van het geven van feedback nog volop in ontwikkeling zijn. Een deel van de ontwikkeling van leraren op het gebied van het geven van feedback bestaat uit het bewustzijn van hoe een leerling over de gegeven feedback denkt. Thurlings et al. verrichten daarom onderzoek naar hoe leerlingen mondelinge feedback percipiëren (2010). Daarbij is een referentiekader noodzakelijk. Dit onderzoek voorziet daarin door zowel de leerling als een docent als een externe observator exact dezelfde vragen te stellen.

Vraagstelling

Dit onderzoek richt zich op de perceptie van mondelinge klassikale feedback. Hierbij worden docentpercepties, leerlingpercepties en percepties van observanten met elkaar vergeleken. De hoofdvraag is als volgt:

In welke mate komen de percepties over mondelinge klassikale feedback bij het bespreken van een werkstuk of huiswerk overeen bij een docent, zijn leerlingen en een externe observant?

Daarbij is het interessant ook verschillende groepen te vergelijken. Op basis hiervan zijn de volgende deelvragen relevant voor het onderzoek:

- 1) Is er een verschil in perceptie van de feedback tussen leerlingen in de bovenbouw en leerlingen in de onderbouw in vergelijking met de perceptie van de feedback van docenten en observanten?
- 2) Is er een verschil in perceptie van de feedback tussen leerlingen op het VWO en leerlingen op de Havo in vergelijking met de perceptie van de feedback van docenten en observanten?
- 3) Is er een verschil in perceptie van de feedback tussen vrouwelijke docenten en mannelijke docenten in vergelijking met de perceptie van de feedback van leerlingen en observanten?
- 4) Is er een verschil in perceptie van de feedback tussen docenten met meer ervaring en docenten met minder ervaring in het onderwijs in vergelijking met de perceptie van de feedback van leerlingen en observanten?
- 5) Is er een verschil in perceptie van de feedback tussen docenten die alfavakken doceren en docenten die bètavakken doceren in vergelijking met de perceptie van de feedback van leerlingen en observanten?

Variabelen

Er zijn op verschillende niveaus variabelen te onderscheiden in dit onderzoek. In feite wordt gemeten in hoeverre de variabelen 'docent', 'leerling', 'observant' overeenkomen. Het is belangrijk om te benadrukken dat de docent in dit onderzoek altijd degene is die de feedback geeft, de leerling degene die de feedback ontvangt en de observant degene die de feedback waarneemt.

Deelnemers aan het onderzoek kunnen de effectiviteit van feedback op verschillende domeinen percipiëren. Het is daarom belangrijk om een aantal kernvariabelen te splitsen uit de overkoepelende term 'perceptie'. Wij onderzoeken de perceptie van deelnemers wat betreft:

- A.** De kwaliteit van de overdracht van de feedback.
- B.** De mate van focus van de feedback op de kwaliteit van het gemaakte werk.
- C.** De aanwezigheid van feedback op essentiële randvoorwaarden voor een succesvolle uitvoering van de leertaak.
- D.** De invloed van de feedback op het vervolgproces.

Daarbij horen de volgende indicatoren:

- A.**
 1. Duidelijkheid
 2. Terechtheid
 3. Leerzaamheid
 4. Positiviteit
 5. Negativiteit
- B.**

1. Focus op goed werk
2. Focus op niet goed werk

C.

1. Focus op werkhouding
2. Focus op samenwerking

D.

1. Mate van motivatie
2. Mate van verheldering leertaak
3. Mate van ruimte voor verbetering leertaak
4. Mate van suggesties voor verbetering leertaak
5. Mate van bruikbaarheid
6. Mate van duidelijkheid over doorgaan op dezelfde weg of verandering van invalshoek

Hypothesen

- de perceptie van de feedback tussen docent en observant komt meer overeen dan de perceptie van de feedback tussen docent en leerling, en observant en leerling.

De reden hiervoor is dat docenten en observatoren over het algemeen meer kennis hebben over de effectiviteit en kwaliteit van feedback. Ze zouden daarom de feedback accurater moeten kunnen inschatten.

- de perceptie van de feedback van een VWO-leerling komt meer overeen met de perceptie van de feedback van docenten en observanten dan de perceptie van de feedback van een HAVO-leerling.

De reden hiervoor is dat VWO-leerlingen over het algemeen een 'denker'-leerstijl (Kolb, 1984) hebben, waarbij reflectie en abstractie een grote rol spelen. Het invullen van deze enquête is een vorm van reflectie, en de indicatoren die daarin voorkomen zijn vrij abstract. Havo-leerlingen zijn over het algemeen 'doeners' en kunnen beter overweg met actieve en concrete situaties.

- de perceptie van de feedback van leerlingen in de bovenbouw komt meer overeen met de perceptie van de feedback van docenten en observanten dan de perceptie van de feedback van leerlingen in de onderbouw.

Er is een grotere kans dat leerlingen in de bovenbouw beter hebben leren reflecteren dan leerlingen in de onderbouw. Bovendien hebben leerlingen in de bovenbouw een uitgebreidere en bredere ervaring met feedback dan leerlingen in de onderbouw.

- de perceptie van de feedback van ervaren docenten komt meer overeen met de perceptie van de feedback van leerlingen en observanten dan de perceptie van de feedback van docenten met minder ervaring.

Ook hier geldt het argument van ervaring. Docenten met meer ervaring hebben vaker feedback gegeven, meer kans gehad om op de effectiviteit van eerder gegeven feedback te reflecteren, en meer kans gehad om hun feedback te herijken en verbeteren.

- Er is geen verschil in de mate waarbij de perceptie van de feedback tussen mannelijke docenten en vrouwelijke docenten overeenkomt met de perceptie van de feedback van leerlingen en observanten.

Het is niet aangetoond dat geslacht een rol speelt wat betreft de perceptie van feedback. Er zijn ook geen logische redenen waarom dat wel zo zou zijn.

- Er is geen verschil in de mate waarbij de perceptie van de feedback overeenkomt tussen docenten die alfavakken doceren en docenten die bètavakken doceren met de perceptie van de feedback van leerlingen en observanten.

Het is niet aangetoond dat het domein van onderwijs een rol speelt wat betreft de perceptie van feedback. Er zijn ook geen logische redenen waarom dat wel zo zou zijn.

Context

Dit onderzoek heeft plaatsgevonden op een openbare school voor VMBO, HAVO en VWO. De school heeft zowel eerste- als tweedegraads opgeleide docenten en biedt elk jaar veel stagiaires de mogelijkheid om ervaring op te doen in het onderwijs. Dit komt omdat de school namelijk nauwe contacten met de eerstegraads lerarenopleiding in Utrecht. Het onderzoek is uitgevoerd op HAVO en VWO.

Methode

Respondenten

In totaal hebben 183 middelbare scholieren (96 meisjes, 77 Jongens, 10 geslacht niet ingevuld) deelgenomen aan het onderzoek, verdeeld over 9 klassen. De vragenlijsten zijn uitgedeeld in 1havo/vwo, 2vwo (3x), 3vwo, 4vwo (2x), 5vwo en 4havo bij de vakken biologie, Engels, mens en natuur, Nederlands, scheikunde en wiskunde B. We hadden graag ook VMBO-leerlingen geënquêteerd. De reden dat het VMBO niet is meegenomen in het onderzoek is het feit dat geen docenten van de VMBO-locatie bereid waren mee te werken aan het onderzoek. Van de betrokken docenten zijn er 3 man en 6 vrouw en ze hebben tussen de 1 en 10 jaar onderwijservaring. Ook hier konden we niet kieskeurig zijn. We hebben iedereen die bereid was mee te werken geënquêteerd. Dit betekent dat de conclusies van dit onderzoek alleen gelden voor havo/vwo.

Door middel van e-mail contact en persoonlijke benadering zijn contacten gelegd en bleek het mogelijk om uiteindelijk bij 9 verschillende docenten te enquêteren. Drie personen die betrokken zijn bij dit onderzoek hebben de rol van observator op zich genomen. De feedback is gemeten bij zowel het bespreken van een werkstuk als het bespreken van huiswerk, omdat één van beide meten niet haalbaar bleek.

Door het grote aantal leerlingen dat in dit onderzoek gebruikt is, zullen de conclusies in de enquêtes over net gegeven mondeling feedback tussen docent en leerling en tussen observant en leerling, betrouwbaarder zijn dan tussen observant en docent. Er zijn immers maar negen respondenten in die beide groepen, waar in principe 30 het minimum is voor statistisch significante uitspraken. Graag hadden we 30 enquêtes afgenomen, maar daarvoor bleek binnen de school niet genoeg draagvlak te zijn. Hoewel de gegevens dus niet statistisch significant genoeg zijn om harde uitspraken te doen over de resultaten, kunnen wat betreft het overeenkomen van de enquêtes tussen observant en docent wel trends worden geconstateerd en daarom relevante uitspraken gedaan worden.

Instrumenten

De enquêtes die gebruikt worden voor dit onderzoek zijn bijlage 1 t/m 3 te vinden. De uitspraken in de enquêtes worden voorgelegd aan docent, leerlingen en observant. Elke enquête bevat vijftien uitspraken over feedback. Deze uitspraken zijn een selectie uit de vragenlijst van het onderzoek van Yvette Sol en Karel Stokking (2009) naar mondelinge feedback bij zelfstandig werken. De geselecteerde vragen sluiten goed aan bij de kernvariabelen beschreven in de paragraaf over variabelen. Alle vijftien zijn ze een indicator van deze vier kernvariabelen. Dit onderzoek maakt gebruik van deze uitspraken omdat ze goed aansluiten bij de kernvariabelen ervan uitgegaan mag worden dat de uitspraken die in dit onderzoek gesteld zijn eenduidig en betrouwbaar zijn.

Opzet

Het centrale element van dit onderzoek is een enquête. De uitspraken in deze enquête zijn zo geformuleerd dat alle groepen de uitspraken kunnen beantwoorden. Bij elke uitspraak geeft de invuller op een schaal van 1 t/m 5 aan of hij het eens is met de uitspraak. Dit levert kwantificeerbare cijfers op, op basis waarvan geanalyseerd kan worden in hoeverre de percepties van de groepen met elkaar overeenkomen.

De verzameling van data heeft plaatsgevonden over de periode van mei en juni 2012. De observatoren hebben docenten benaderd en hebben vervolgens (een deel van) een lesuur bijgewoond waarin de docent mondelinge klassikale feedback gaf op huiswerk of een werkstuk.

Direct na de feedback deelde de observator de vragenlijsten uit en vulden alle drie de groepen de vragenlijst in. Dit was essentieel om te garanderen dat de perceptie geen herinnering werd, maar zo accuraat mogelijk kon worden gevangen in de respons.

Na de dataverzameling zijn de data verwerkt in verschillende tabellen die de relatie tussen de drie groepen weergaven (zie resultaten). Hiertoe heeft één van de observatoren een programma in Excel ontwikkeld. Vervolgens zijn de data geanalyseerd op basis van de hoofd- en deelvragen.

Analysemethode

In deze paragraaf zal exact beschreven worden hoe de resultaten verwerkt zijn om de resultaten reproduceerbaar te houden. Het technische gedeelte staat hier, de interpretatie en intuïtie achter de berekende waarden staat bij resultaten.

De resultaten van de vragenlijsten worden als volgt gecodeerd. $D_{i,j}$, $O_{i,j}$, $L_{i,j}$. Hierbij staat de D voor de antwoorden van de docent, de O voor de antwoorden van de observator en de L voor het gemiddelde van de antwoorden van de leerlingen. i staat voor het lesnummer en j staat voor het vraagnummer. Dus $O_{2,7}$ geeft aan wat de observator heeft beantwoord bij de 2^e les in het onderzoek op vraag 7 (dus 1, 2, 3, 4 of 5).

Er is voor gekozen om het gemiddelde van de antwoorden van de leerlingen te gebruiken als score voor leerlingen.

Omdat we naar de verschillen in perceptie zullen kijken, definiëren we de volgende variabelen: De belangrijkste indicatoren voor het beantwoorden van de hoofdvraag zijn: gemiddelde afwijking en gemiddelde absolute afwijking tussen twee groepen. Voor de leesbaarheid wordt er gebruikgemaakt van een voorbeeld voor het verschil tussen docent en leerling.

Gemiddelde afwijking (het gemiddelde verschil tussen twee groepen per vraag):

$$(D - L)_j \stackrel{\text{def}}{=} \frac{1}{9} * \sum_{i=1}^9 (D_{i,j} - L_{i,j})$$

Gemiddelde absolute afwijking (het gemiddelde van het absolute verschil tussen twee groepen, per vraag):

$$|D - L|_j \stackrel{\text{def}}{=} \frac{1}{9} * \sum_{i=1}^9 |D_{i,j} - L_{i,j}|$$

Totaal gemiddelde afwijking (het gemiddelde verschil tussen twee groepen, ook gemiddeld over alle vragen):

$$\overline{(D - L)} \stackrel{\text{def}}{=} \frac{1}{9} * \sum_{j=1}^n \frac{1}{n} * (O - D)_j$$

Totaal gemiddelde absolute afwijking (het gemiddelde van het absolute verschil tussen twee groepen, ook gemiddeld over alle vragen):

$$\overline{|D - L|} \stackrel{\text{def}}{=} \frac{1}{9} * \sum_{j=1}^n \frac{1}{n} * |O - D|_j$$

Om te testen of de verschillen statistisch significant zijn wordt een gepaarde tweezijdige t-test uitgevoerd met ongelijke variantie. Hierbij wordt aangenomen dat t_j de Student-T verdeling volgt:

$$t_j = \frac{(D - L)_j}{S_{DL,j}/\sqrt{9}}$$

met 8 vrijheidsgraden. Hierbij staat n voor het totaal geobserveerde lessen. Hierbij staat $S_{DL,j}$ voor de zuivere schatter van de standaard deviatie van $D - L$:

$$s_{DL,j} = \sqrt{\frac{1}{8} * \sum_{i=1}^9 ((D_{i,j} - L_{i,j})^2 - (D - L)_j^2)}$$

Het voordeel van deze methode is dat er per vraag een p waarde wordt gegeven die de kans aan geeft dat de afwijking door toeval ontstaat. Deze waarden zijn met elkaar te vergelijken en dit geeft dus een indicatie van welke vragen beter en welke vragen slechter zijn.

Het nadeel van deze aanpak is dat omdat alles per vraag wordt genomen, het aantal observaties (9) erg laag is. Hierdoor is de spreiding in de berekening hoger waardoor de resultaten aan de voorzichtige kant zijn. Immers er is veel meer data beschikbaar die gelijke spreiding heeft en gelijke conclusies met zich mee draagt.

Een ander nadeel aan deze aanpak is dat omdat alles per vraag wordt genomen de kans bestaat dat er één vraag per toeval een lage p-waarde krijgt. Hierdoor kan het voorkomen dat er onterecht geconcludeerd wordt dat de hypothese verworpen dient te worden.

In alle testen zijn niet ingevulde variabelen niet meegenomen, dus het gemiddelde wordt bepaald over de wel ingevulde vragen.

Een validatie van de statistische tests op basis van een normaalverdeling bevindt zich in bijlage 4.

Analyse validiteit vragenlijst

Ook is een verantwoording nodig voor het verwijderen van vier vragen uit de enquête op basis van de data. De resultaten laten namelijk zien dat niet alle uitspraken uit de enquêtes binnen de context van dit onderzoek even effectief, relevant of werkbaar zijn. Hierbij spelen het percentage niet beantwoorde uitspraken, de standaarddeviaties en aantal ingevulde 1-en een rol (zie tabel 5 in het hoofdstuk resultaten). Vijf uitspraken verdienen op dit gebied specifieke aandacht: uitspraken 3, 4, 7, 8 en 9.

Uitspraak 3 en 4 vertonen hetzelfde probleem. In principe zou het antwoord op deze uitspraak relatief eenduidig en objectief moeten zijn: de feedback gaat nu eenmaal wel of niet over iets wat wel of niet goed ging. De resultaten laten echter zien dat de leerlingen het helemaal niet met elkaar eens zijn. Dit is te zien in de rij *gemiddelde standaarddeviatie tussen leerlingen*, die meet in hoeverre leerlingen van elkaar afwijken. Deze twee uitspraken tonen zelfs de allergrootste standaarddeviaties van de vragenlijst. Het gaat hierbij om een afwijking van respectievelijk 1,05 en 1,004. Het is verrassend dat bij een dergelijke uitspraak de antwoorden van de leerlingen zover uit elkaar liggen. Het is daarom mogelijk dat de leerlingen de uitspraak onduidelijk vonden of niet goed begrepen hebben. De uitspraken zouden daarom om de verkeerde redenen kunnen suggereren dat de percepties van de drie partijen niet overeenkomen.

Uitspraak 8 vertoont geen problemen wat betreft de standaardafwijking. Hierbij valt echter op dat 108 van de in totaal 198 leerlingen bij deze vraag een 1 hebben ingevuld. Daarnaast blijkt uit de relatief kleine standaarddeviatie dat de overige 90 leerlingen ook niet veel hogere cijfers hebben ingevuld. Dit suggereert dat de feedback simpelweg niet vaak genoeg over werkhouding ging om een relevante bijdrage te kunnen leveren aan ons onderzoek. Bovendien zouden de resultaten kunnen suggereren dat de perceptie van de drie partijen veel meer overeenkomt dan daadwerkelijk het geval is bij de relevantere uitspraken.

Datzelfde geldt voor uitspraak 9, maar om een andere reden. Bij uitspraak 9 was in de enquête expliciet aangegeven dat leerlingen deze uitspraak niet hoefden in te vullen wanneer er geen sprake was van samenwerking. Uit de resultaten blijkt dat slechts 28% van de leerlingen deze uitspraak

heeft ingevuld. Geen enkele andere uitspraak is door minder dan 98% van de leerlingen beantwoord. Voor uitspraak 9 geldt dus ook dat hij de resultaten onwenselijk zou kunnen beïnvloeden. Er is nog een andere uitspraak die aandacht verdient. Uitspraak 7 is door 77 leerlingen met een 1 beantwoord. Dit is echter goed verklaarbaar: het is goed mogelijk dat leerlingen feedback over het algemeen als helemaal niet negatief ervaren. Dat is een relevant resultaat. Het feit dat de leerlingen hier 1 invullen is dan ook geen indicatie dat de uitspraak niet van toepassing is, wat bij uitspraak 8 wel het geval is.

Om de voorgaande redenen hebben we ervoor gekozen de uitspraken 3, 4, 8 en 9 niet mee te nemen in de analyse van de resultaten. Uitspraak 7 is wel meegenomen.

Resultaten

Resultaten hoofdvraag

Hieronder worden kort de data gepresenteerd die betrekking hebben op de conclusie over de hoofdvraag in het volgende hoofdstuk. De data omvatten 9 lessen, waarin de groepen docent (D), leerling (L) en observator (O) 15 vragen hebben beantwoord. De score van de leerling (L) is bepaald door het gemiddelde te nemen van alle leerlingen.

Tabel 1 gaat over afwijkingen. Deze afwijkingen hebben direct betrekking op de hoofdvraag, aangezien het meten van de mate van afwijking van de verschillende groepen ten opzichte van elkaar ons onderzoeksdoel is.

In de tabel geven de bovenste horizontale rijen het type afwijking aan. Deze begrippen worden nader toegelicht in de tekst onder de tabel. Onder de afwijkingstypes staat welke groepen met elkaar vergeleken worden: D-L betekent dat het getal in de tabel de docent vergelijkt met de leerling. De getallen 1 t/m 15 in de eerste kolom geven aan over welke vraag de vergelijking gaat.

Tabel 1: Afwijkingen

j	Gemiddelde afwijking			Gemiddelde absolute afwijking			p waarde		
	(D-L) _j	(O-L) _j	(O-D) _j	D-L _j	O-L _j	O-D _j	D-L	O-L	O-D
1	-0,228	0,439	0,667	0,327	0,64	0,667	0,191	0,109	0,022
2	0,525	0,636	0,111	0,619	1,014	0,556	0,031	0,079	0,681
3	-0,507	0,382	0,889	0,848	0,652	1,111			
4	-0,059	-0,504	-0,444	0,633	0,625	0,667			
5	0,049	-0,507	-0,556	0,787	1,017	1,222	0,888	0,238	0,302
6	-0,106	0,227	0,333	0,423	0,442	0,556	0,635	0,296	0,282
7	0,090	-0,466	-0,556	0,48	0,734	1	0,698	0,102	0,179
8	-0,068	-0,179	-0,111	0,807	0,455	0,556			
9	X	X	X	0	0,25	0			
10	-0,220	-0,442	-0,222	0,737	0,988	0,889	0,525	0,273	0,512
11	-0,378	-0,044	0,333	0,554	0,683	0,778	0,081	0,873	0,397
12	-0,118	-0,118	0,000	0,563	0,708	0,444	0,581	0,686	1
13	-0,199	-0,088	0,111	0,699	0,791	0,778	0,55	0,806	0,76
14	-0,521	-0,982	-0,375	0,989	1,539	1,625	0,256	0,128	0,61
15	0,098	-0,124	-0,222	0,785	1,301	1,414	0,732	0,793	0,665

In de tabel geeft de **gemiddelde afwijking** (O-D)_j het gemiddelde verschil aan tussen twee groepen. Als bij D-L staat -0.228 geeft dat bijvoorbeeld aan dat docenten op die vraag gemiddeld 0.228 lager antwoordden dan het gemiddelde van de leerlingen. Het kan dus zijn dat sommige docenten te hoog zaten en anderen veel te laag; gemiddeld zaten ze in ieder geval lager.

De **gemiddelde absolute afwijking** geeft aan hoe dicht de groepen bij elkaar zaten. Een D-L van 0.327 geeft bijvoorbeeld aan dat de docent en leerling gemiddeld over de 9 lessen 0.327 uit elkaar zaten. Dit kan dus zowel 0.327 te hoog zijn als te laag. 4 docenten 0.327 te hoog en 5 docenten 0.327 te laag zou dit als resultaat opleveren.

Het verschil tussen de gemiddelde- en absolute gemiddelde afwijking kan als volgt gezien worden. Wanneer je naar je werk fietst heb je mensen die gewoon altijd 5 minuten te vroeg komen. Je hebt

mensen die altijd 5 minuten te laat komen. Dit is de gemiddelde afwijking (-5, +5 respectievelijk). Je hebt ook mensen die om wat voor een reden een hele grote spreiding hebben, soms zijn ze veel te vroeg en soms veel te laat, deze hebben een grote absolute afwijking.

Persoon	Gemiddelde afwijking	gemiddelde absolute afwijking
Altijd 5 minuten te vroeg	-5	5
Altijd 5 minuten te laat	+5	5
Soms 10 minuten te vroeg (50%) soms 10 minuten te laat (50%)	0	10

De **p-waarde** geeft aan hoe groot de kans is dat het gemiddelde verschil per toeval ontstaat. Een lage p-waarde geeft dus aan dat het onwaarschijnlijk is dat de groepen overeen komen. Normaal gesproken wordt de grens van $p=0.05$ aangehouden, en in dit geval betekent dat dus dat bij p waarden kleiner dan 0.05 het verschil statistisch significant is. De statistisch significant verschillende cijfers zijn dik gedrukt.

Resultaten deelvragen

Bij negen verschillende docenten is de feedback geobserveerd en hebben de docent, de leerlingen en de observator een vragenlijst ingevuld. Per docent is aangegeven hoeveel leerlingen in de klas de vragenlijst hebben ingevuld, het schooltype, of het een onderbouw of bovenbouw klas is, het aantal jaren dat de docent werkzaam is in het onderwijs, of het een alfa- of bètavak betreft en het geslacht van de docent. Van alle leerlingen uit een klas is per vraag het gemiddelde antwoord berekend. Voor elke vraag zijn de absolute verschillen tussen de groepen berekend. Het gemiddelde van de 11 geselecteerde vragen (zie voor een verklaring van de vier weggestreepte vragen de paragraaf *toetsing vragenlijst* in het hoofdstuk **conclusie**) geeft aan hoever de twee groepen gemiddeld per vraag uit elkaar liggen. Deze waarden zijn per docent weergegeven in Tabel 3.

Tabel 3. Absolute verschil in perceptie per observatie (gemiddelde van 11 vragen). De resultaten zijn weergegeven op een kleurenschaal (rood-geel-groen)

Observatie bij	Docent vs Leerling	Observator vs Leerling	Observator vs Docent	aantal leerlingen	Schooltype	Onderbouw / bovenbouw	aantal jaren werkzaam	Alfa/bèta	Geslacht
Docent 1	0,31	0,79	1,00	15	vwo	onderbouw	10	alfa	V
Docent 2	1,03	0,85	0,73	4	vwo	onderbouw	9	bèta	V
Docent 3	0,78	0,80	0,73	25	vwo	onderbouw	10	bèta	M
Docent 4	0,63	0,50	0,55	30	vwo	bovenbouw	4	bèta	M
Docent 5	0,71	0,96	1,09	18	vwo	onderbouw	7	alfa	V
Docent 6	0,44	0,64	0,73	23	vwo	bovenbouw	1	alfa	V
Docent 7	0,55	1,07	1,33	24	havo/vwo	onderbouw	8	bèta	M
Docent 8	0,50	0,86	0,46	22	havo	bovenbouw	1	alfa	V
Docent 9	0,71	1,43	1,27	22	vwo	bovenbouw	1	bèta	V

In de tweede kolom is weergegeven hoever de docent en de leerlingen uit elkaar liggen; deze waarden variëren van 0,31 tot 1,03. In de derde kolom is weergegeven hoever de observator en de leerlingen uit elkaar liggen; deze waarden variëren van 0,50 tot 1,43. In de vierde kolom is weergegeven hoever de observator en de docent uit elkaar liggen. Deze waarden variëren van 0,46 tot 1,33.

Om antwoord te geven op de deelvragen is voor verschillende testgroepen berekend wat het absolute verschil per vraag gemiddeld is. Hierbij is rekening gehouden met het verschillende aantal leerlingen per klas. De resultaten van deze testgroepen zijn weergegeven in tabel 4.

Tabel 4. Absolute verschil in perceptie per testgroep (gemiddelde van 11 vragen). De resultaten zijn weergegeven op een kleurenschaal (rood-geel-groen)

Testgroep	Docent vs Leerling	Observator vs Leerling	Observator vs Docent	aantal docenten	aantal leerlingen
Totaal	0,63	0,88	0,88	9	183
Onderbouw (klas 1-3)	0,68	0,90	0,97	5	86
Bovenbouw (klas 4-6)	0,57	0,86	0,75	4	97
Havo	0,50	0,86	0,46	1	22
Vwo	0,66	0,85	0,87	7	137
havo/vwo brugklas	0,55	1,07	1,33	1	24
1-4 jaar werkzaam	0,57	0,86	0,75	4	97
7-10 jaar werkzaam	0,68	0,90	0,97	5	86
mannelijke docenten	0,65	0,75	0,83	3	69
vrouwelijke docenten	0,62	0,92	0,88	6	114
Alfa	0,49	0,81	0,82	4	78
Bèta	0,74	0,92	0,92	5	105

Resultaten data-evaluatie vragenlijst

Om de effectiviteit van alle vragen te toetsen is het belangrijk te kijken naar een aantal factoren die een indicatie kunnen zijn van een zwakke vraag binnen de context van het onderzoek. Deze factoren zijn hieronder in tabel 5 weergegeven.

Tabel 5: Indicaties zwakke vraag

	Aantal keer niet beantwoord	Gemiddelde standaarddeviatie tussen leerlingen	Gemiddelde afwijking van 3	Aantal keer antwoord 1
Vraag 1	1	0,626379	1,31995	1
Vraag 2	1	0,741634	1,18693	2
Vraag 3	0	1,050303	1,18261	36
Vraag 4	1	1,004855	1,203767	8
Vraag 5	3	0,856725	1,006681	2
Vraag 6	2	0,705687	1,195405	3
Vraag 7	2	0,751354	1,376128	77
Vraag 8	3	0,774615	1,594319	108
Vraag 9	142	Niet beschikbaar (delen 0)	1,686349	25
Vraag 10	0	0,919221	0,99496	20
Vraag 11	0	0,758104	1,231387	2
Vraag 12	0	0,697303	1,267309	2
Vraag 13	3	0,824892	0,99554	11
Vraag 14	2	0,999595	1,295915	12
Vraag 15	0	0,960776	0,985394	15

Conclusie en discussie

Conclusie hoofdvraag

De hoofdvraag in dit onderzoek is als volgt: *In welke mate komen de percepties over mondelinge klassikale feedback bij het bespreken van een werkstuk of huiswerk overeen bij een docent, zijn leerlingen en een externe observant?*

Er zijn verschillende manieren om de hoofdvraag te beantwoorden. Tabel 4 biedt een interessante eerste inzicht: de docent en leerlingen zijn het absoluut gemiddeld meer met elkaar eens (0,68 verschil) dan docent en observator (0,88) en observator en leerling (0,88). Dit staat haaks op de hypothese dat de percepties van de docent en observator het meest overeen zouden komen. Er zijn verschillende verklaringen mogelijk. Het is bijvoorbeeld mogelijk dat de docent rekening houdt met wat de leerlingen zullen antwoorden en dat goed in kan schatten. Ook is mogelijk sprake van een politiek correcte onderschatting van het eigen handelen, dan wel een sociaal wenselijke overschatting door de observator ('t Hart 2003). De docent vult inderdaad gemiddeld lagere cijfers in dan de leerling, die op zijn beurt lagere cijfers invult dan de observator (tabel 1). Daarnaast is het mogelijk dat het kleine aantal observatoren (3) een rol speelt: wanneer deze drie allemaal een natuurlijke neiging tot bovenmatige positieve beoordeling hebben, kan dit de resultaten onevenredig beïnvloeden. Een andere mogelijkheid is dat de leerlingen door praktijkervaring met feedback veel beter de kwaliteit en effectiviteit van feedback kunnen inschatten dan in de hypothese werd verondersteld.

De p-waarden geven aan dat de percepties van de partijen in 2 van de 32 gevallen significant van elkaar verschillen. De docent en leerling verschillen significant wat betreft de terechtheid van de feedback, en de observator en de docent wat betreft de duidelijkheid. Het kleine aantal significante waardes geeft aan dat de conclusies van dit onderzoek voorzichtig gesteld moeten worden; er zijn wel trends zichtbaar, maar die kunnen niet altijd voldoende worden ondersteund om een volwaardige conclusie te waarborgen. De verwachting is wel dat bij een grotere dataset, en dus een grotere steekproef, een aantal vragen (4) met een relatief lage p-waarde eveneens richting een p-waarde van 0,05 of lager zou gaan.

Conclusie deelvragen

In onze onderzoeksgroep werken alle ervaren docenten in de onderbouw en alle beginnende docenten in de bovenbouw. De perceptie van feedback door docenten met weinig ervaring in de bovenbouw komt volgens de resultaten van dit onderzoek meer overeen met de leerlingen dan bij ervaren docenten in de onderbouw. Het is onduidelijk of dit verschil wordt veroorzaakt door ervaring van de docent of de leeftijd van de leerlingen. Wat zou kunnen meespelen is dat docenten met weinig ervaring nog in een fase zitten waarin ze (nog) erg bezig zijn met de kwaliteit van het overbrengen van de feedback. Deze docenten zijn net begonnen en besteden veel aandacht aan het geven van effectieve feedback. Docenten met meer ervaring kunnen (al) meer op routine werken, waardoor het geven van feedback minder prioriteit kan krijgen. Hierdoor kunnen leerlingen deze feedback als kwalitatief minder beschouwen, waardoor er een groter verschil in perceptie van de feedback tussen docenten met meer en minder ervaring met de perceptie van leerlingen zou kunnen ontstaan. De hypothese dat de perceptie van de feedback van ervaren docenten meer overeenkomt met de perceptie van de feedback van leerlingen wordt dus niet ondersteund door de resultaten in dit onderzoek.

We hebben te weinig onderzoeksresultaten in de groep Havo om een betrouwbare conclusie te trekken over een verschil in perceptie van de feedback van leerlingen van de verschillende

schooltypes (Havo en VWO) met de perceptie van de feedback van docenten. Daardoor is het ook moeilijk om een uitspraak te doen over de hypothese dat de perceptie van een VWO-leerling meer overeenkomt met de perceptie van de feedback van docenten dan de perceptie van de feedback van een Havo-leerling. Wel is het opvallend dat de resultaten in dit onderzoek laten zien dat de perceptie van de feedback van leerlingen van de Havo meer overeenkomt met de perceptie van de feedback van docenten dan de perceptie van de feedback van leerlingen van het VWO. Wanneer deze resultaten bij een grotere steekproef stand zouden houden, zou dit verklaard kunnen worden door het feit dat leerlingen op de Havo minder kritisch kunnen zijn op de feedback gegeven door docenten. Leerlingen op het VWO kunnen kritischer tegen de feedback gegeven door docenten aankijken, waardoor er een groter verschil in perceptie van de feedback tussen leerlingen op het VWO en de Havo met de perceptie van docenten zou kunnen ontstaan. Het is echter goed mogelijk dat een ander beeld ontstaat bij een grotere steekproef.

Wat betreft mannelijke en vrouwelijke docenten zijn in dit onderzoek geen grote verschillen waarneembaar tussen percepties van hen en van leerlingen. Uit onderzoek is nog niet gebleken dat er een verschil bestaat op de manier waarop mannen en vrouwen hun feedback overbrengen naar leerlingen. Op dit moment is het dus ook nog niet bewezen dat leerlingen de feedback van mannelijke en vrouwelijke docenten anders ervaren. De hypothese dat er geen verschil in de mate waarbij de perceptie van de feedback tussen mannelijke docenten en vrouwelijke docenten overeenkomt met de perceptie van de feedback van leerlingen is dus op basis van de onderzoeksresultaten in dit onderzoek niet ontkracht. De verwachting is niet dat dit zal veranderen met een grotere steekproef.

Opvallend is dat in dit onderzoek de perceptie van docenten die lesgeven in alfavakken meer overeenkomt met de perceptie van leerlingen dan bij docenten die lesgeven in de bètavakken. Dit verschil wordt mogelijk veroorzaakt door het beperkte aantal docenten in dit onderzoek. Ook zou kunnen meespelen dat de antwoorden die gegeven worden bij alfavakken concreter en minder diepgaand zijn dan bij bèta vakken. Alfaleerlingen kunnen deze antwoorden dus als duidelijker beschouwen waardoor de feedback van docenten met betrekking tot deze antwoorden ook duidelijker over kan komen. Dit kan ervoor zorgen dat er een kleiner verschil in perceptie van de feedback tussen docenten die lesgeven in alfavakken en leerlingen ontstaat dan tussen docenten die lesgeven in bèta vakken en leerlingen. De hypothese dat er geen verschil in de mate waarbij de perceptie van de feedback overeenkomt tussen docenten die alfavakken doceren en docenten die bètavakken doceren met de perceptie van de feedback van leerlingen wordt op basis van de onderzoeksresultaten dus niet bevestigd, en zou getoetst moeten worden in een onderzoek met een grotere steekproef.

Validiteit en betrouwbaarheid

Bij de beschouwing van deze conclusies is het belangrijk rekening te houden met een aantal problemen in het onderzoek. Een groot probleem is de kleine dataset. In eerste instantie was het de bedoeling om veel meer data te verzamelen om een representatieve steekproef te kunnen presenteren. In de praktijk bleek dit echter niet haalbaar. Na herhaaldelijke oproepen bleken slechts negen docenten bereid mee te werken. Er zijn in de resultaten trends zichtbaar; deze kunnen helaas echter niet statistisch significant onderbouwd worden.

Deze beperking bracht andere problemen met zich mee. In eerste instantie was het de bedoeling alleen enquêtes af te nemen in groepen die een werkstukbespreking deden. Dit had de voorkeur omdat er dan een rijke input aan gerichte feedback beschikbaar zou zijn. Bovendien zou er geen sprake van zijn dat emotie de resultaten zou beïnvloeden, wat een risico was geweest bij het observeren van toetsbesprekingen. Dit bleek echter onhaalbaar: slechts twee docenten gaven aan dat de observatoren welkom waren om hun werkstukbespreking bij te wonen. Het compromis om

ook huiswerkbesprekingen bij te wonen was daarmee onvermijdelijk. Bij de huiswerkbesprekingen was de feedback vaak minder gericht en uitgebreid en daarmee minder geschikt voor dit onderzoek.

Er zijn verschillende mogelijkheden voor het uiteenlopen van de percepties. Zo kunnen de verschillende percepties worden toegeschreven aan het feit dat één of meer partijen de feedback accurater kan of kunnen inschatten, maar ook aan verschillende meningen over wat goede feedback nu eigenlijk is. Dit onderzoek is er niet in geslaagd één mogelijkheid te isoleren om daarover een sluitende uitspraak te kunnen doen.

Een aantal vragen bleek niet geschikt voor dit onderzoek. Dit is al beschreven onder het kopje 'analyse validiteit vragenlijst' in het hoofdstuk 'analysemethode'.

Vervolgonderzoek

De meningen over wat goede feedback is verschillen van persoon tot persoon. Dit kan voor een deel de verschillen in percepties van feedback die in dit onderzoek zijn waargenomen verklaren. Als docenten en observatoren vooraf een inleiding zouden krijgen met daarin een theoretisch kader over feedback, dan is de verwachting dat de verschillen in perceptie van feedback tussen deze partijen kleiner zou worden. Ook de verschillen in perceptie tussen docenten onderling zou kleiner kunnen worden. De verwachting is dat de verschillen in perceptie van feedback tussen mannelijke docenten en vrouwelijke docenten met hun leerlingen zal verdwijnen; dit verschil was in dit onderzoek immers al zeer klein. Ook als dit onderzoek zou worden uitgebreid met meer metingen is de verwachting dat deze verschillen zullen verdwijnen.

We verwachten dat er geen verschil zal zijn in perceptie van docenten die op havo en docenten die op het VWO lesgeven. Er zijn namelijk veel docenten die op zowel het VWO als op de havo lesgeven. De leerlingen zouden nog wel voor een verschil in perceptie van feedback kunnen zorgen; onze hypothese is nog steeds dat de perceptie van de feedback van een VWO-leerling meer overeenkomt met de perceptie van de feedback van docenten en observanten dan de perceptie van de feedback van een HAVO-leerling, omdat de dataset te klein was om op dit gebied veel belang te hechten aan de data.

We hebben in ons onderzoek niet kunnen waarnemen dat de perceptie van ervaren docenten meer overeenkomt met die van leerlingen dan bij onervaren docenten; de drie docenten die pas een jaar werkzaam zijn in het onderwijs verschillen minder van hun leerlingen dan de ervaren docenten. Onderwijservaring lijkt hierbij geen rol te spelen, maar dat moet definitief blijken bij een grotere dataset.

Het grootste verschil hebben we gevonden tussen docenten in de alfavakken en docenten in de bètavakken. Dit zou kunnen worden veroorzaakt door de aard van de vakken en de besproken opdrachten. De verwachting is dat dit verschil bij meer metingen zal blijven bestaan.

Er zijn een aantal suggesties voor vervolgonderzoek die uit dit onderzoek volgen.

1. Er is onderzoek gedaan naar de percepties van de verschillende groepen. Het zou echter ook interessant zijn om een zelfde soort proces toe te passen waarbij gemeten wordt hoe de docent en observator de perceptie van de leerling inschatten, en of deze overeenkomt met de perceptie van de feedback van de leerling zelf. Dit brengt in kaart in hoeverre docenten en observatoren kunnen inschatten hoe een leerling denkt over feedback.
2. Hetzelfde onderzoek kan worden toegepast nadat de drie groepen een uitgebreide introductie over academisch onderzoek naar effectieve en goede feedback hebben gekregen. Dit elimineert het probleem waarmee de vorige paragraaf werd afgesloten.
3. Het strekt tot aanbeveling de leerling in onderzoeken naar feedback te blijven betrekken, omdat deze groep feedback anders ervaart dan de docent zelf doet. Voor verder onderzoek naar feedback is het niet genoeg om één van de groepen te ondervragen: alle verschillende

groepen (docent, leerling, eventueel observant) moeten worden meegenomen om een compleet beeld van de gegeven feedback te krijgen.

Suggesties praktijk

1. Tijdens het dataverzamelingsproces bleek dat een deel van de docenten erg benieuwd was naar wat de leerlingen hadden ingevuld. Dit bevestigt de aanbeveling in het onderzoek van Van Petegem en Deneire (2006), waarin zij stellen dat het betrekken van leerlingen in het proces van evaluatie docenten inzicht geeft over de effectiviteit van feedback. Het lijkt dan ook aan te bevelen dit in de praktijk vaker toe te passen.
2. Uit de relatief hoge standaarddeviatie, weergegeven in tabel 5, blijkt dat leerlingen het zelf ook niet altijd met elkaar eens zijn wat betreft de effectiviteit van feedback. Houd bij het geven van feedback rekening met het feit dat verschillende leerlingen verschillende feedback als effectief ervaren.

Reflectie proces en product

Er zijn een aantal zaken in het proces van het onderzoek die beter hadden gekund. Bij de dataverzameling hadden de observatoren eerder en actiever op zoek kunnen gaan naar deelnemers. Misschien had dat uiteindelijk meer data opgeleverd. Toch is het belangrijk om te vermelden dat het schoolklimaat zich niet optimaal leende voor een dergelijk onderzoek. Weinig docenten waren bereid constructief mee te helpen of denken. Onderzoek doen alleen is dus niet genoeg: belangrijker is misschien nog wel dat een docent-onderzoeker zijn omgeving kan enthousiasmeren. Dat had in dit geval beter gekund.

Het product heeft dan ook een te kleine dataset om sluitende conclusies en aanbevelingen anders dan trends en voorzichtige indicaties te presenteren. Wel toont het een kritische en constructieve houding ten opzichte van de rol van feedback in het onderwijs. Bovendien is het onderzoek een aanvulling op de rol als docent gebleken; ook buiten de klas is een kritische houding ten opzichte van het mechaniek van de school en het leerproces belangrijk.

Literatuurlijst

Brookhart, S. M. (2008). Feedback: An overview. *How to give effective feedback to your students*, ASCD.

Hart, H. 't. (2003) *Sociale wenselijkheid*. Jaarboek Markt Onderzoek Associatie.

Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, p. 81-112.

Kolb, D.A. (1984). *Experiential Learning, experience as the source of learning and development*. Englewood-Cliffs, NJ: Prentice Hall.

Open Universiteit. Leraren beroepsonderwijs moeten begeleiden in plaats van controleren, *Persbericht 21 juni 2011*, Opgevraagd op 13-1-2012 van <http://www.ou.nl/eCache/DEF/2/34/331.html>

Sol, Y., & Stokking, K. (2009). Mondelinge feedback bij zelfstandig werken: interactie tussen docenten leerlingen in het Voortgezet Onderwijs, *Volume 74 van Vormgeving van leerprocessen*, IVLOS, Universiteit Utrecht, 170 p.

Sol, Y., Stokking, K. & Vermunt, J. Mondelinge feedback bij het verwerken aan een taak: docentgedrag, docentcognities en leerling percepties, *Paper voor de Onderwijs Research Dagen op 27-29 mei 2009 te Leuven*, Thema Lerarenopleiding en Leraarsgedrag, 23 p.

Thurlings, M., Wassink, W., Bastiaens, T., Stijnen, S. & Vermeulen, M. *Feedback als cruciale factor voor leerresultaat*, Ruud de Moor Centrum, 10 p.

Van Petegem, P., & Deneire, A. (2006). Leraren leren van leerlingen; de ontwikkeling van een theoretisch kader voor een gecombineerd instrument voor zelfreflectie en feedback, *Impuls*, Nr. 4, 10 p.

Van Rijs, A. *Onderzoek Feedback in de klas*, Calvijn College, Goes, Opgevraagd op 12-1-2012 van <http://www.leraar2020.nl/pdf/samenvattingen%20onderzoeken/Samenvatting%20Onderzoek%20Arno%20van%20Rijs.pdf>

Bijlage 1: Enquête voor de leerling

Naam:

Klas:

M / V

Datum:

Docent:

De volgende vragen gaan over de feedback die je zojuist hebt gekregen. Kloppen deze uitspraken? Omcirkel het cijfer dat het best bij je antwoord past.	Helemaal niet			Helemaal wel	
	1	2	3	4	5
1. De feedback was duidelijk	1	2	3	4	5
2. Ik vond de feedback terecht	1	2	3	4	5
3. De feedback ging over wat goed ging of goed was gedaan	1	2	3	4	5
4. De feedback ging over wat nog niet goed was of nog niet goed ging	1	2	3	4	5
5. De feedback gaf heel precies aan wat er goed ging of wat er anders moest	1	2	3	4	5
6. De feedback kwam positief over	1	2	3	4	5
7. De feedback kwam negatief over	1	2	3	4	5
8. De feedback ging over de werkhouding	1	2	3	4	5
9. De feedback ging over de samenwerking (niet invullen als je het alleen moest doen)	1	2	3	4	5
10. De feedback motiveerde me om verder te werken	1	2	3	4	5
11. De feedback was leerzaam	1	2	3	4	5
12. De feedback was bruikbaar	1	2	3	4	5
13. Door de feedback wist ik wat ik moest doen om verder te kunnen	1	2	3	4	5
14. De feedback gaf me de kans mijn werk te verbeteren	1	2	3	4	5
15. Met de feedback gaf de leraar mij suggesties om mezelf te verbeteren	1	2	3	4	5

Bijlage 2: Enquête voor de docent

Naam:

Vak:

M / V

Aantal jaren werkzaam in het onderwijs:

Datum:

De volgende vragen gaan over de feedback die u zojuist aan de leerling(en) heeft gegeven. Kloppen deze uitspraken? Omcirkel het cijfer dat het best bij uw antwoord past.	Helemaal niet		Helemaal wel		
	1	2	3	4	5
1. De feedback was duidelijk	1	2	3	4	5
2. Ik vond de feedback terecht	1	2	3	4	5
3. De feedback ging over wat goed ging of goed was gedaan	1	2	3	4	5
4. De feedback ging over wat nog niet goed was of nog niet goed ging	1	2	3	4	5
5. De feedback gaf heel precies aan wat er goed ging of wat er anders moest	1	2	3	4	5
6. De feedback kwam positief over	1	2	3	4	5
7. De feedback kwam negatief over	1	2	3	4	5
8. De feedback ging over de werkhouding van de leerling	1	2	3	4	5
9. De feedback ging over de samenwerking (niet invullen als het niet van toepassing is)	1	2	3	4	5
10. De feedback motiveerde de leerling om verder te werken	1	2	3	4	5
11. De feedback was leerzaam	1	2	3	4	5
12. De feedback was bruikbaar	1	2	3	4	5
13. Door de feedback wist de leerling wat hij/zij moest doen om verder te kunnen	1	2	3	4	5
14. De feedback gaf de leerling de kans zijn/haar werk te verbeteren	1	2	3	4	5
15. Met de feedback gaf ik de leerling suggesties om zichzelf te verbeteren	1	2	3	4	5

Bijlage 3: Enquête voor observant

Naam:

Datum:

Klas:

Docent:

De volgende vragen gaan over de feedback die je zojuist hebt gekregen. Kloppen deze uitspraken? Omcirkel het cijfer dat het best bij je antwoord past.	Helemaal niet			Helemaal wel	
	1	2	3	4	5
1. De feedback was duidelijk	1	2	3	4	5
2. Ik vond de feedback terecht	1	2	3	4	5
3. De feedback ging over wat goed ging of goed was gedaan	1	2	3	4	5
4. De feedback ging over wat nog niet goed was of nog niet goed ging	1	2	3	4	5
5. De feedback gaf heel precies aan wat er goed ging of wat er anders moest	1	2	3	4	5
6. De feedback kwam positief over	1	2	3	4	5
7. De feedback kwam negatief over	1	2	3	4	5
8. De feedback ging over de werkhouding van de leerling	1	2	3	4	5
9. De feedback ging over de samenwerking (niet invullen als het niet van toepassing is)	1	2	3	4	5
10. De feedback motiveerde de leerling om verder te werken	1	2	3	4	5
11. De feedback was leerzaam	1	2	3	4	5
12. De feedback was bruikbaar	1	2	3	4	5
13. Door de feedback wist de leerling wat hij/zij moest doen om verder te kunnen	1	2	3	4	5
14. De feedback gaf de leerling de kans zijn/haar werk te verbeteren	1	2	3	4	5
15. Met de feedback gaf de leraar de leerling suggesties om zichzelf te verbeteren	1	2	3	4	5

Bijlage 4: Validatie statistische testen op basis van normaalverdeling

Voor de testen die uitgevoerd gaan worden is het noodzakelijk dat het gemiddelde van de variabelen bij benadering normaal verdeeld is. Hoewel vanwege de centrale limietstelling de gemiddelden van variabelen met eindige variantie altijd bij benadering normaal verdeeld zijn, is onze steekproef dermate klein dat dit niet noodzakelijk het geval is. Om deze reden is het belangrijk te kijken of de onderzochte variabelen normaal verdeeld zijn.

In tabel 2 staan de eigenschappen tussen verschillende groepen. Hiervoor zijn alle vragen en lessen samengenomen.

Tabel 2: Eigenschappen tussen verschillende groepen

	D-L	O-L	O-D	Normale verdeling
Gemiddelde	-0,0916952	-0,162402311	-0,07071	μ
Standaarddeviatie	0,82180939	1,145261123	1,247467	σ
Scheefheid	-0,7402951	-0,401628803	-0,15351	0
Kurtosis	1,05842713	0,693946986	1,243763	0

De **scheefheid** is zit tussen de -1 en 1 en is dus gemiddeld tot laag. De **kurtosis** is positief maar niet overmatig. De variabelen lijken ook voor het bepalen van het gemiddelde op een normale verdeling. Hierdoor kan het gemiddelde van de variabelen goed benaderd worden met een normale verdeling.

Voor de testen is ook onderlinge onafhankelijkheid nodig. Dit betekent dat de metingen elkaar niet mogen beïnvloeden, of samen mogen hangen. Een mogelijke reden die voor deze samenhang zou kunnen zorgen is dat enkele observatoren meerdere lessen hebben geobserveerd.

Opvallend in de dataset is dat er bij een aantal vragen het verschil tussen de groepen heel groot is, en bij een aantal vragen dit verschil erg klein is. Omdat de waarden bij benadering onderling onafhankelijk normaal verdeeld zijn kan de tweezijdige, gepaarde Student-t test met ongelijke variantie worden uitgevoerd. Voor vraag 2, het verschil tussen docent en leerling, geeft dit een p van 0.016 dat de docent en leerling overeenkomen. Voor het verschil tussen observator en leerling geeft dit een p van 0.054 bij de nulhypothese dat docent en leerling overeenkomen. Voor vraag 1, het verschil tussen docent en observator geeft dit een p van 0.011 dat de docent en leerling overeenkomen.