

Leerdoelgericht feedbackvol (dus) cijferloos toetsen, in de onderbouw

Onne van Buuren,
Luuk Nelissen,
Madelief de Jong
+ andere collega's



NRO
NATIONAAL RECHTERSDIENST

METIS
MONTESSORI
LYCEUM

VU
VRIJE
UNIVERSITEIT
AMSTERDAM
Faculteit der
Gedrags- en
Bewegings-
wetenschappen

Vraag 1

Welk cijfer geef je aan een leerling
die op een toets
slordig en onaf formuleert,
nauwelijks berekeningen opschrijft,
vaak eenheden vergeet
maar toch inzicht toont
en alle antwoorden goed heeft?

Vraag 2

Welke cijfer geef je aan leerling
die in een toets
waarin de wet van Ohm een hoofdrol speelt
bij alle vragen consequent rekent met $R = I / U$
en verder geen fouten maakt?

Cijferloos beoordelen?

1. Effecten van cijfers
2. Formatief, summatief en cijfers?
3. Situatie op Metis en kernvragen
4. Randvoorwaarden voor de pilot
5. Voorwaarden werkzaamheid feedback systeem
6. Leerdoelentabel als basis voor feedbacksysteem
7. Triage
8. Flankerende didactiek
9. Resultaten
10. Stand van zaken



1. Effecten van cijfers, Metis

“Ik ben blij, meester!”

Zei een leerling uit 3H die een 8,0 had gehaald voor zijn eerste toets
.... toen hij zijn cijfer kreeg voor zijn tweede toets: een 3,4.

“Ik sta gemiddeld 5,7!”

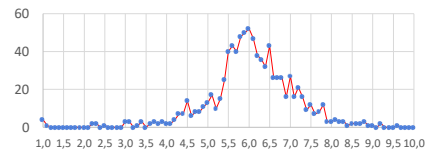
Eind van het schooljaar → hele lijst met 5,6-en en 5,7-ens
Hij mikte daar bewust op
N = 1 leerling?

METIS
MONTESSORI
LYCEUM



1. Effecten van cijfers op het Metis

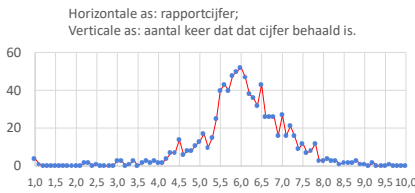
Horizontale as: rapportcijfer;
Verticale as: aantal keer dat dat cijfer behaald is.



- Alle rapportcijfers bij alle schoolvakken 3H eind 2018.
- Ruim de helft van de cijfers ligt tussen 5,5 en 6,4 → 6
- ..,3 en ..,4: ondervetegenwoordigd,
- ..,5 en ..,6: overvetegenwoordigd

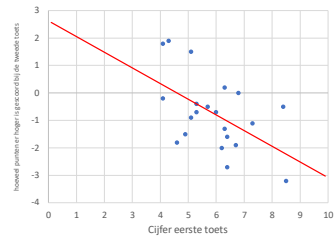


1. Effecten van cijfers, Metis eind 2018



- Direct alle cijfers middelen: → 5,94
- Afronden op gehelen, dan middelen → 6,13
- Afrondingsbonus: 0,19 gemiddeld per leerling per vak
- 12 vakken → bonus van meer dan 2 punten per rapport

1. Effecten van cijfers, Metis 3H



- Effect van het cijfer voor de eerste toets op de tweede toets?
- N = 1 Eén 3H-klas natuurkunde (20 leerlingen), twee toetsen (2019)

1. Effecten van cijfers

Verklaring?

- Afrondings- en middelingscultuur stimuleert leerlingen om onze norm (5,5) te halen (en niet veel meer)

Waarom?

- Leerlingen zijn niet per se lui, vaak 'efficiënt' → moeten keuzes maken
- Leerlingen passen zich aan **ONZE normen** aan
- Docenten werken waarschijnlijk mee aan 5,5 -cultuur:
 - geven bij twijfel liever een 5,5 dan een 5,4
 - durven soms niet om veel onvoldoendes te geven

Consequenties:

- Primaire doel: het **gemiddelde** eindcijfer, niet het leren van het vak
- Doorgaande leerlijnen? Leerlingen komen 'ver', ondanks grote hiaten



2. Formatief, summatief en cijfers



"Het initiatief van LAKS en de VO-raad richt zich op het verminderen van toetsdruk in het voortgezet onderwijs. Door de focus te verleggen van summatieve naar formatieve toetsen, wordt de nadruk gelegd op persoonlijke ontwikkeling in plaats van cijfers."

- Hoofdargument: te veel toetsen (is dat echt zo?)
- Maar toetsen helpt je kennis beter op te slaan, doordat je er meer moeite voor moet doen... (Spitzer, 1939; Sins 2024)
- Toetsconcurrentie tussen secties..
- Niet genoemd: het ontstaan van hiaten

2. Formatief, summatief en cijfers?

Pedagogische blik: DOELEN VAN TOETSEN



2. Formatief, summatief en cijfers



- Veel scholen willen 'minder summatief' en 'meer formatief' gaan toetsen Maar...
- Overblijvende toetsen gaan over (te) veel stof → stress! of slechte dag?
- Leerlingen leren vaak niet voor toetsen die niet 'meetellen'
- ... of geven voorrang aan summatief georganiseerde toetsen (verdringing)
- Resultaten vallen (dus) tegen

2. Formatief, summatief en cijfers



Een paar misverstanden rechtzetten:

1. **Summatief en formatief vormen geen tegenstelling, maar zijn twee verschillende aspecten van toetsing**
2. **Betere naam: formatief evalueren of formatief handelen**
3. **.... en dat deed elke goede docent altijd al**
4. **Lettercodes (o, z, v, g; A t/m F-level) zijn geen kenmerk van formatieve toetsing maar zijn vermomde cijfers**

2. Formatief, summatief en cijfers?



Onderzoek van Butler (1988):

- Twaalf klassen, dezelfde lessen. Schriftelijk werk ingenomen en nagekeken.
 - A. 4 klassen: schriftelijk commentaar op het werk,
 - B. 4 klassen: cijfer,
 - C. 4 klassen: zowel schriftelijk commentaar als cijfer.

Pre- en posttest.

- Groep A: grote vooruitgang
- Groep B: geen vooruitgang
- Groep C: **ook geen vooruitgang**

Conclusie:

het geven van cijfers werkt contraproductief

Herhaaldelijk bevestigd door ander onderzoek, ook al bij Montessori (1912)

2. Formatief, summatief en cijfers?



De focus op cijfers verklaart waarom leerlingen:

- commentaar van docenten op toetsen nauwelijks lezen
- na publicatie van het cijfer op Magister de toets niet meer in willen zien
- vaak niet opletten bij het nabespreken van toetsen

Uitzondering: als dezelfde stof op korte termijn opnieuw getoetst wordt (bijvoorbeeld voor een herkansing) zijn leerlingen wél geïnteresseerd.

2. Formatief, summatief en cijfers?



- Herkansing is een positief formatief element, maar...
- Leerlingen geven voorrang aan toetsen die niet herkanst mogen worden
- Is een 5,5, behaald na herkansing, eigenlijk wel een voldoende?

2. Formatief, summatief en cijfers?



- Butler e.a.: cijfers leiden tot 'ego-' of 'prestatie-' gerichte houding
 - Focus van leerling op vergelijking met klasgenoten
 - Leerlingen vermijden risico's en uitdagingen (gevaar afbreuk eigenwaarde)
- In tegenstelling tot taakgerichte houding:
 - 'je best doen helpt'
 - 'je kunt leren van je fouten'
- Commentaar i.p.v. cijfers kan deze houding stimuleren
- cijfer ≈ beloning, en beloningen blijken intrinsieke motivatie te verminderen. (Lillard, mede op basis van Montessori):

2. Middelen van cijfers?



- primaire doel van leerlingen is gemiddeld eindcijfer, dus leerlingen mikken vooral op scores, minder op begrijpen
- leerlingen proberen hun werk niet te verbeteren omdat een matige volgende toets meer kan bijdragen aan hun eindcijfer bij dezelfde tijdsinvestering

2. Samengevat



Conclusie:

- **Toetsen op zichzelf bevordert leren**
- **Formatieve evaluatie/ goede feedback draagt bij aan leren**
- **Cijfers ondermijnen effect formatieve evaluatie**
- **Niet 'meetellende' toetsen leggen het af tegen wel 'meetellende' toetsen**
- **Middelen van cijfers kan doorgaande leerlijnen ondermijnen**

3. Situatie op Metis en kernvragen



Metis in transitie van klassikaal → Montessori

- Vooral klassikaal geroosterde vaklessen
- Inhaaluren voor toetsen
- Sinds 2019 M-uren (keuzewerktijd) → iets meer eigen tempo
- Meeste secties gebruiken gewone lesmethodes
- Meeste secties geven cijfers → Magister, gemiddelden
- 2021-2022: start in klas 2V(H), natuurkunde
- 2022-2023: uitbreiding naar klas 3V en 3H, natuurkunde, en '2023-2024: verbeterpunten invoeren en testen
- 2024-2027: start NRO-project

3. Situatie en kernvragen



Hoe kun je overgaan op een systeem

- waarin formatieve mogelijkheden (**feedback**) optimaal worden benut,
- dus zonder cijfers en gemiddelden,
- als de meeste andere secties wel werken met cijfers, gemiddelden en minder herkansingsmogelijkheden? (o.a. wegdrukeffect)

- Kunnen leerlingen nog wennen?
- Welke flankerende didactiek is nodig?
- Onder welke organisatorische omstandigheden kan het werken?
- Hoe benutten de leerlingen de verschillende elementen van zo'n systeem, en in het bijzonder de feedback bij de toetsen? Optimalisatie?

4. Randvoorwaarden

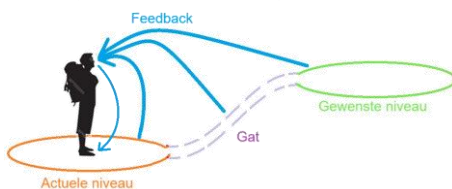


1. **Werkbaar voor docenten**
2. **Werkbaar voor docenten**
3. **Werkbaar voor docenten**
4. Niveau voldoende borgen, lat niet te laag
5. Begrijpelijk voor leerlingen
6. Werkbaar voor leerlingen
7. Risico's minimaliseren

In feite: hoe kun je feedback optimaliseren in de (les- en) toetspraktijk?

5. Voorwaarden werkzaamheid feedback

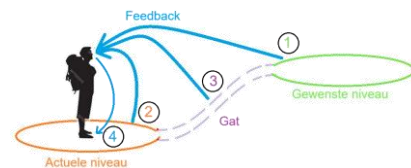
“Feedback is informatie over het gat tussen het actuele niveau en het gewenste niveau die gebruikt wordt om dat gat te verkleinen”



5. Voorwaarden werkzaamheid feedback

Feedback binnen een leerproces is alleen effectief als de leerling:

1. informatie krijgt over het gewenste niveau,
2. informatie krijgt over zijn actuele niveau,
3. informatie krijgt over het gat tussen actueel en gewenst niveau, en
4. actie onderneemt om het gat te verkleinen





5. Voorwaarden werkzaamheid feedback

Essentiele kenmerken voor werkzaamheid feedback:

- leerling moet **weten wat er mee te doen**
- leerling moet de informatie **inhoudelijk begrijpen**
- leerling moet **gemotiveerd zijn om de feedback te gebruiken**



6. Leerdoelentabel als basis voor feedbacksysteem

Geïnspireerd door het 'stoplichtsysteem' van Dylan Wiliam.

- **Tabel in boek** (of werkwijzer) met genummerde leerdoelen, met koppeling aan §§
- **Verkorte versie van de tabel OP de toets**
- Leerlingen maken de toets **OP het opgavenformulier**
- Docent:
 - Beoordeelt per leerdoel in tabel op de toets: **in orde, bijna in orde, nog niet in orde**
 - Inhoud: geeft fouten aan in de kantlijn met nummers van gemiste leerdoelen
 - Leerproces: leertips / toetstips bij tabel omcirkelen
 - Beoordeelt toets als geheel: **in orde, bijna in orde, nog niet in orde** → triage

6. Tabel in boek

Werkboek bevat per hoofdstuk leerdoelentabel met verwijzingen naar §§.

Leerdoel	Ja, bijna, nog niet? (Wat denk je zelf?)
1. Ik gebruik grootheden, eenheden en het SI correct (Onderzoeken), ik weet dat eenheden in een (woord)formule bij elkaar moeten passen en ook te de zo nodig aan. (§1) en ik schrijf de juiste eenheid achter elk antwoord (§1)	
2. Ik kan omrekenen tussen en binnen het liter- en metrieke stelsel (Onderzoeken)	
3. Ik kan werken met de woordformules voor: <ul style="list-style-type: none"> - het oppervlak van een rechthoek (§0, wiskunde), - het volume van een balk (§0, wiskunde), - het volume van een 'taagjes'-voorwerp (§0). 	
4. Ik kan een formule opstellen voor dichtheid, massa en volume (§1, §3) Ik ken de symbolen voor deze grootheden (§3)	
5. Ik weet hoe ik rekenopgaven moet aanpakken (§1): <ul style="list-style-type: none"> a. eerst de (woord)formule(s) opschrijven, b. als een stof gegeven is: de dichtheid opzoeken, c. dan alle gegevens invullen d. laatste uitwerken. 	
6. Ik kan een formule omwerken zodat alle bekende grootheden aan één kant van het =-teken komen (§3)	
7. Ik kan aan de dichtheid zien of een voorwerp zinkt, zweeft of drijft in een vloeistof (§2)	
8. Ik kan de (woord)formules van leerdoel 3 en 4 met elkaar combineren (§3)	
9. Ik kunt met een woordformule de middelde dichtheid uitrekenen van een voorwerp dat uit meerdere stoffen bestaat (§2 en §3)	
10. Ik heb de practica naar behoren gedaan (§2 en §3), en kan: <ul style="list-style-type: none"> - het volume van een voorwerp meten met een maatcilinder (§1) - massa meten met een elektronische weegschaal (§1) 	

6. Corresponderende tabel op toets, inclusief leerprocesgerichte feedback

- Eindbeoordeling toets op basis aantal behaalde leerdoelen (Magister)

Gehele toets voor:	in orde	bijsa	nog niet in orde
		→	herkansingen

- Lat ligt hoog!
- Gemaakte toets = administratie
- Direct zichtbaar verband toetsmatrijs ↔ resultaat
- Ook leerproces-gerichte feedback (sinds vorig jaar)

Toets 'Dichtheid' 2V en 2VH

	in orde	bijsa nog niet		in orde	bijsa nog niet
1. SI grootheden en juiste eenheden, omschrijven eenheden			6. formule omwerken:		
2. Omrekenen liter → metriske stelsel			7. zinken, zweven, drijven		
3. Wordformules: Volume en oppervlak bereken en gebruiken			8. combineren van formules		
4. Formule voor dichtheid, volume en massa bereken en gebruiken			9. gemiddelde dichtheid (V)		
5a. aanpak rekenopgaven en gegevens opzoeken			10. Practicum dubbeltoets		
leerdoelformules opschrijven invullen			Toetstips: leer goed / houde bestmelingen, werk opvakkend / houde niet te veel.		
leestheid niet vergeten			Leer tips: leer opvakkend / maak het hoofdstuk af / maak genoeg opgaven / kijk goed na		

Start → 0. Naam: Klas 2..... Datum

• (Bijna) in orde rekenopgaven



7. Triage

Beoordeling op basis van aantal behaalde leerdoelen en gevolgen



Figuur 1: werkstroombiagram per hoofdstuk met triage.

8. Flankerende didactiek

- Leg het systeem goed uit aan leerlingen (per toets, per rapport)
- Gebruik de tabel in de les, bijv.:
 - Wat gaan we vandaag doen?
 - Welke leerdoelen beheers je al?
 - Oefenopgaven met klas nakijken met leerdoelentabel erbij
- Doe dit vaak
- Begeleid 'in orde maken' en herkansen (stimuleren, gevareknop in Magister)
- Informeer de mentoren/coaches

8. Rapport



(Voorlopig) compromis 🚫:

- voorlopig rapportcijfers op basis van aantal toetsen in orde (veel)! want
- ..eerder was gebleken (informatica): geen rapportcijfer → minder serieus genomen door leerlingen én docentenvergadering*
- Breuk door periodisering en overgang?
 - Extra ondersteuning volgend leerjaar
 - Overgang naar C&M en E&M?
 - Alsnog calculerend gedrag bij bekend worden normering

9. Resultaten enquête 21-22 klas 2

- Alle drie 2V-klassen
- (helaas) één gewone 2VH-klas
- Special 2VH-klas

9. Eerste resultaten (kerst 2021)



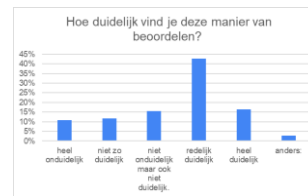
Interviews:

- 'Je kunt veel beter zien wat je wel en niet het snapt, een cijfer zegt je niet zoveel'
- 'Als je voor wiskunde een 4 hebt moet je maar zien of en hoe je de stof alsnog leert'
- 'Als je voor wiskunde een 4 hebt moet je voor de volgende toets een 8 halen, dat geeft veel stress'
- 'Herkansmogelijkheid is heel fijn'
- "Nog niet" klinkt fijner dan een 4'

Leerling die anders misschien een 10 had gehad:

- '8, 9 of 10 maakt me niet zoveel uit'
- 'Wel wat stress, dat je alles in orde moet hebben'

9. Resultaten 21-22 alle geënquêteerden, klas 2



- VH- iets minder positief
- Verband duidelijkheid ↔ ervaren moeilijkheid
- "Cijfer zegt iets duidelijker hoe ze de toets hebben gemaakt."
- Relatie beoordeling ↔ cijfer moet enigszins duidelijk worden gemaakt .

9. Resultaten 21-22 alle geënquêteerden, klas 2

- Dat een toets die 'bijna in orde' is alsnog in orde gemaakt moet worden vinden leerlingen (heel) goed.
- De mogelijkheid om een beetje uit te stellen wordt ook gewaardeerd
- De mogelijkheid tot herkansen wordt zeer gewaardeerd.
- Ook over de verplichting om te herkansen zijn leerlingen niet negatief.
- De manier van beoordelen leidt er volgens de leerlingen niet toe dat ze harder of minder hard hoeven te werken dan bij andere vakken.

9. Resultaten 21-22 alle geënquêteerden, klas 2



- Kijk je vaker of minder vaak naar wat je fout hebt gedaan vergeleken met andere vakken?
- Linker kolommen: vinden natuurkunde moeilijk

9. Resultaten 21-22 alle geënquêteerden, klas 2

Ik leer door de manier van beoordelen bij natuurkunde vergeleken met de meeste andere vakken:



Hierbij is geen eenduidige relatie met moeilijk vinden.

9. Resultaten 21-22 alle geënquêteerden, klas 2

	Percentage
Bij natuurkunde moeten er voortaan weer cijfers gegeven worden	23%
Bij andere vakken zouden ze ook zonder cijfers moeten gaan werken	20%
Het maakt me niet zoveel uit of er cijfers worden gegeven of niet	45%
Anders:	12%

“Anders” bevat vooral steunbetuigingen (10%), maar met nuances, bijv.:
- aspecten, zoals de herkansingsmogelijkheid.

- ‘bijna’ –variant is niet geschikt is voor reproductietoetsen.
(daarvoor is [slim stampen](#), van [Memorylab](#), waarschijnlijk nuttiger).

9. Overige resultaten tot nu toe

- Veel leerlingen ervaren minder stress
- Deel van de leerlingen: leerdoelentabel is nuttig bij het leren
- Niet-werkers vallen door de mand
- Leerlingen die het hoofdstuk niet af hebben ook
- Toetsen enigszins aanpassen aan beoordelingssysteem → betere afstemming.
- Minder gemarchandeer over 0,1 punt
- Meer ruimte om eisen te stellen
- **Meer contact met leerlingen over leren**
- **Soms betere diagnostisering**

9. Resultaten (onderwijsontwerper)

- Tabel = toetsmatrijs → groter bewustzijn & betere afstemming leerdoelen ↔ lesmateriaal ↔ toets (constructive alignment)

Bijvoorbeeld: ‘leerlingen kunnen formule omwerken’

- maar leerlingen kunnen vormen van formule ook uit hun hoofd leren
- Dus: expliciete vraag met omwerken onbekende formule

Leerdoelen naar inhoud en niveau: hoe ga je om met complexere vragen?
Leerdoelen → in content standards naarmate het niveau hoger wordt

10. Stand van zaken



- Resultaten overwegend positief → onze sectie gaat door
- Eerste poging niet serieus (?) → valt erg mee
- We verliezen soms de concurrentieslag met andere vakken
- Flexibiliteit maximaliseren → M-uren en toetsinhaaluren
- Bij eindrapportcijfer soms toch wat calculerend gedrag
- Soms uitstelgedrag → coachen, gevaar knop, eisen aan tempo herkansen
- Voorwaarden bij herkansing
- Docententeam is positief
- O&O, Informatica en Kunst & Co werken ook zonder cijfers
- Andere secties experimenteren met elementen van het systeem

10. Stand van zaken

- **Cijferloos** (+ toetsen, ± rapporten)
- **Feedbackvol** (+ feedbacksysteem, ++± stimuleren van feedbackgebruik)
- **Leerdoelgericht** (+ bij voorbereiding en deels verwerking, ± deels verwerking)

Randvoorwaarden

1. **Werkbaar voor docenten** +
2. Niveau voldoende borgen, lat niet te laag +
3. Begrijpelijk voor leerlingen +
4. Werkbaar voor leerlingen +
5. Risico's minimaliseren +

10. Stand van zaken



- Kunnen leerlingen aan zo'n systeem wennen als ze het niet gewend zijn? +
- Onder welke organisatorische omstandigheden kan het werken? +
- Bij welke flankerende didactiek kan het werken? +
- Hoe verlopen hierbij leerprocessen van leerlingen: hoe benutten ze de elementen van het systeem?
- Verschillen in aanpak en opbrengsten bij deelnemende vakken?
- Kijkkader voor VO-scholen die ook cijferloos willen toetsen?



11. Als je dit ook wilt doen:

- Zorg voor heldere leerdoelen (tabel)
- Alignment: stem toetsen en leerdoelen goed op elkaar af
- Meerdere toetsversies
- Besteed in de klas regelmatig aandacht aan het systeem
- Leg de lat hoog
- Er moet ruimte zijn voor differentiatie
- Coaching is soms belangrijk!

o.p.m.vanbuuren@VU.nl

O.Buuren@msa.nl



Referenties en verdiepingsliteratuur

- Black, P., & Harrison, C. (2010). Formative assessment in science. In J. Osborne & J. Dillon (Eds.), *Good Practice In Science Teaching: What Research Has To Say* (p. 183). McGraw-Hill, Open University Press.
- Black, P., & William, D. (1998). *Inside the black box: Raising standards through classroom assessment*. Granada Learning.
- Butler, R. (1988). Enhancing and undermining intrinsic motivation: the effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, 58(1), 1–14.
- Clymer, J.B. and William, D. (2007). Improving the Way we grade Science: Standards based Grading Systems can improve how we communicate Learning Expectations to Students. *Educational Leadership*, 63(4), 36-42.
- Lillard, A. (2007). *Montessori: The Science Behind the Genius*. Oxford University Press, USA.
- Lipnevich, A. A., & Smith, J. K. (2009). Effects of differential feedback on students' examination performance. *Journal of Experimental Psychology: Applied*, 15(4), 319.
- Montessori, M. (1912) *The Montessori Method*. Translated by Anne Everett George. New York: Frederick A. Stokes Company. Retrieved from <https://digital.library.upenn.edu/women/montessori/method/method.html>
- Spitzer, H. F. (1939). Studies in retention. *Journal of Educational Psychology*, 30(9), 641.
- Sins, P. (2021). Leerpowerrrrr!. *Didactief*, retrieved from <https://didactiefonline.nl/blog/blonz/leerpowerrrrr-03-11-2024>
- William, D., & Black, P. (1996). Meanings and consequences: a basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, 22(5), 537–548.