2025-04-08T03:45:06+00:00 • Leestijd 9 - 12 minuten

Leerlingen krijgen jaarlijks tientallen cijfers. Maar worden die wel zinnig gegeven? Dat kan beter, met een nieuwe methode om huiswerk te beoordelen. Het is interessant, intuïtief en sommige docenten zweren erbij.

Het is verdraaid moeilijk om in het onderwijs een eerlijk oordeel te vellen. Deze methode kan helpen

▶ Luisteren Beluister in de app

Correspondent Sport



Michiel **DE HOOG**





Foto's door Annegien van Doorn (voor De Correspondent)

Stel, je bent leraar.

Stel, je hebt je leerlingen de opdracht gegeven om een opstel te schrijven.

En stel, nu liggen er 25 betogen op je bureau over de definitie van 'lef', over waarom je wel of niet voor euthanasie, dierenproeven of kernenergie bent, of over het Oedipuscomplex in *De donkere kamer van Damocles*.

Hoe ga je de betogen beoordelen?

Als je geen plan hebt, dan lees je het eerste betoog en denk je misschien: hmm, niet goed, niet slecht, een boel spelfouten. Een 7.

Je pakt het volgende betoog erbij en denkt: sodeju, dit kind kan serieus denken! Een 9. En dan lees je opstel drie. Duidelijk een nog beter stuk dan die 9 van zojuist. Maar moet je, met nog 22 opstellen te gaan, nu al een 10 uitdelen?

Cijfers geven op intuïtie is lastig. Voor je het weet, moet je al die opstellen herlezen en herwaarderen. Daarom hebben docenten vaak een plan. Ze hanteren een checklist met elementen – bijvoorbeeld stijl, structuur, beeldspraak – waarop ze de essays beoordelen.

Wel zo belangrijk, want leerlingen krijgen op school aan de lopende band cijfers. De checklist maakt het de docent makkelijker om de essays te beoordelen op de gewenste elementen. Het helpt ook om de betogen betrouwbaar te beoordelen – dus dat je dezelfde prestaties dezelfde beoordeling geeft.

Probleem opgelost, zou je denken.



Een quiz: welk kattenkwaad vind je erger?



En nu?

Checklists zijn top. Totdat ze een doel op zich worden

Maar een checklist wordt al gauw een doel op zich. Als docenten een checklist hanteren, moeten ze leerlingen ook wijzen op de elementen uit die checklist. Wel zo eerlijk; dat is immers waar ze op worden beoordeeld. Maar gewiekste leerlingen kunnen essays inleveren die compleet voldoen aan de elementen, maar toch matig zijn, of zelfs absurd.

In het Britse voortgezet onderwijs kregen leerlingen extra punten door een zin met een bijwoordelijke bepaling te beginnen – een constructie die als stijlvol gezien werd. En dus kregen docenten superstijlvolle en tevens volstrekt onzinnige zinnen opgediend, zoals: '*Forgettably, he crept through the darkness*'. (Op vergetelijke wijze kroop hij door het donker.)

Een nog groter probleem voor docenten zijn essays die niet voldoen aan de checklist, maar toch erg goed zijn. Waarom zo'n essay fantastisch is, is niet altijd goed uit te leggen. Maar je *weet* het gewoon, al je collega's weten het ook, en toch is het volgens de checklist geen hoog cijfer waard. Dus wat doen docenten vaak? Ze baseren het cijfer op hun gevoel, en gebruiken vervolgens de elementen van de checklist om hun oordeel 'wit te wassen'.

Een beetje belachelijk. En dan hebben de checklists je ook nog niet bevrijd van het probleem dat je halverwege de stapel beseft dat je te streng of te soft bent geweest voor de voorgaande essays.

Stilzwijgende kennis: als je iets weet, maar je het niet kunt uitleggen

Wat nu als je de boel omdraait? Als je je oordeel niet witwast, de goedbedoelde checklists en hun vergetelijk lelijke resultaten opzijlegt, en simpelweg vertrouwt op je expertise? Of zoals de wetenschapsfilosoof Michael Polanyi het noemde: op je stilzwijgende kennis?

Polanyi wist dat er tal van dingen zijn die mensen weten, zonder dat ze kunnen uitleggen waarom ze ze weten. Neem fietsen: je weet hoe het moet, want je doet het foutloos, ook al kan je niet beschrijven hoe. En zelfs als je dat zou kunnen, dan is die beschrijving zo lang en gedetailleerd, dat niemand die wil leren fietsen er iets aan heeft. Denk ook aan je moedertaal: je spreekt 'm perfect, maar je kunt 'm iemand anders niet zomaar leren. (Probeer het maar eens: drama!)

Het is dus letterlijk stilzwijgende kennis. Stel nou dat je die stilzwijgende kennis over essays het uitgangspunt maakt van je beoordeling. Dat je alles wat niet in de checklist staat, maar toch van waarde is, kunt meenemen in het cijfer dat je geeft. Dat is het idee achter een nieuwe methode om cijfers en feedback te geven: comparatief beoordelen.

Niet meten, maar vergelijken

Het idee achter comparatief beoordelen is al zo'n honderd jaar oud – maar raakte in de vergetelheid. Begin deze eeuw werd het

concept herontdekt door enkele Britse wetenschappers, maar pas in de jaren tien ging een stel docenten ermee aan de slag.

Het werkt als volgt: docenten uploaden de essays van de leerlingen naar een softwareprogramma. Het programma schotelt ze telkens twee essays voor. De enige vraag die ze beantwoorden, is welke van de twee beter is. Daarna krijgen ze weer twee essays voorgeschoteld, en weer twee, en weer twee. Steeds draait het om dezelfde simpele vraag: welke van de twee is het betere essay?

Met elke vergelijking wordt duidelijker hoe goed de essays zijn ten opzichte van elkaar; de software vervat de vergelijkingen in een ranglijst. Dit proces gaat door totdat de docenten zoveel paartjes hebben beoordeeld – en elk essay vaker hebben gezien – dat de ranglijst niet meer noemenswaardig verandert.

Het resultaat is miraculeus: een ranglijst die de collectieve wijsheid van docenten over wat goed en slecht is weerspiegelt. Een van de voordelen van de werkwijze is dat leerlingen niet meer afhankelijk zijn van de grillen van een docent. Het oordeel over hun werk is het collectieve oordeel van – bijvoorbeeld – alle docenten aardrijkskunde van 3 havo.
Nu denk je misschien: hartstikke leuk dat het zo eerlijk is, maar kost het niet veel meer tijd als een hele groep docenten elk essay vaker dan één keer bekijkt? Meestal is dat antwoord: nee. Een docent ziet weliswaar meer essays dan normaal voorbijkomen, maar bespaart veel tijd doordat er geen checklist meer nodig is. De enige vraag die de docent hoeft te beantwoorden is immers: welk essay van de twee is beter?

Een versimpelde casus – geleend van een bedrijf dat software voor comparatief beoordelen verkoopt – helpt te begrijpen waarom het werkt. Hieronder zie je een rijtje van acht tinten paars, gesorteerd van licht naar donker, waar de lichtste tint (bijna wit) een één is, en de donkerste een acht. De tinten en hun scores moet je onthouden.

Wat je vaak krijgt: vier, vijf of soms zes goede antwoorden. Dit illustreert dat dit absoluut (geïsoleerd) beoordelen, zoals docenten doen met checklists, heel lastig betrouwbaar te doen is.

Daarna ga je vergelijkend beoordelen. Je krijgt telkens paartjes van twee tinten paars te zien. Je enige taak: het donkerdere blokje aanklikken.

Na zo'n twintig snel gemaakte keuzes blijkt dat je tot een perfecte beoordeling van de tinten bent gekomen.

Zo werkt het met het vergelijken van de essays ook: de opgetelde vergelijkingen resulteren in een betrouwbare ranglijst.

Als je maar mazzel moet hebben met je docent

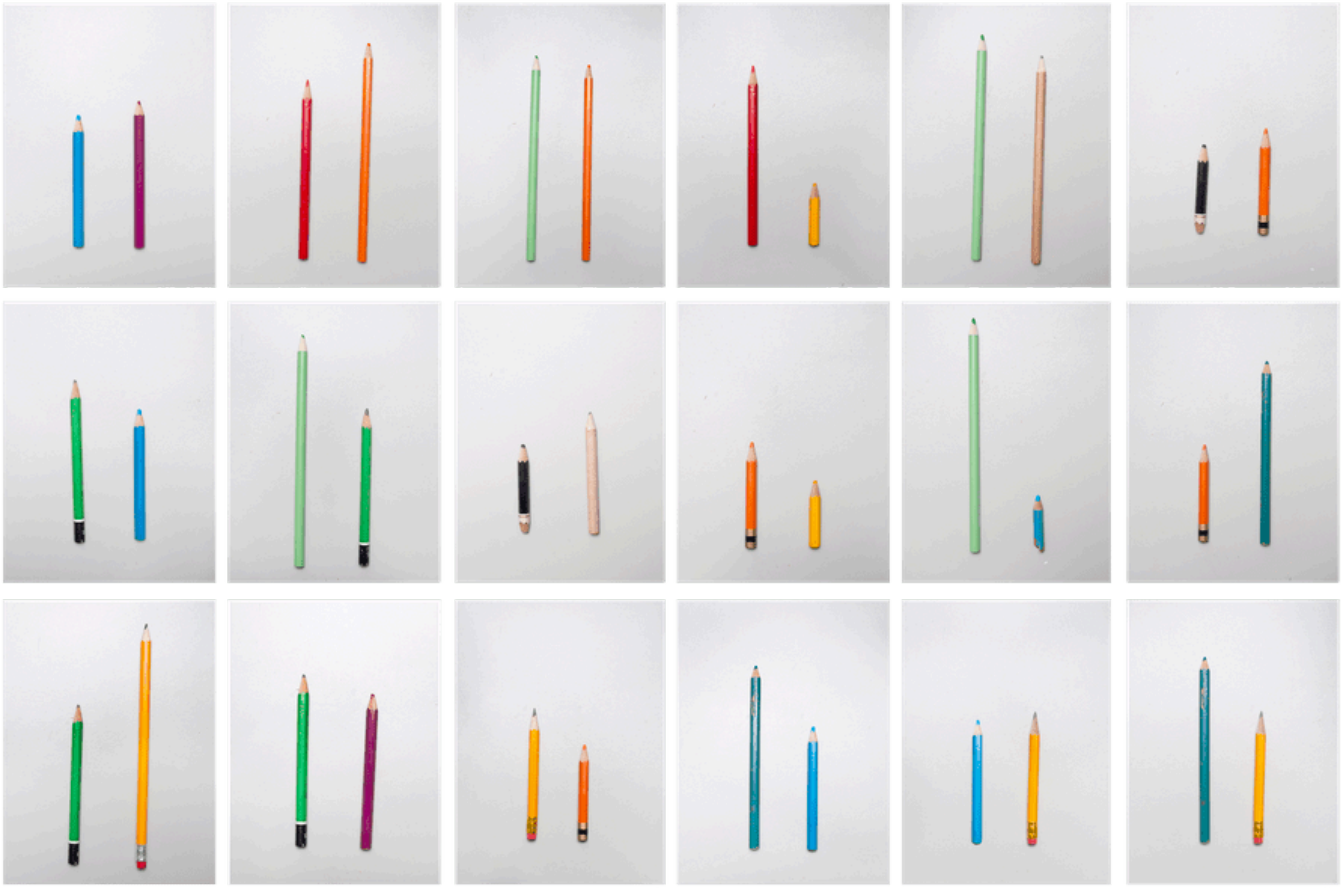
Een van de voorstanders van comparatief beoordelen is Daisy Christodoulou, een voormalig docent Engels uit Londen, die in haar boek *Making Good Progress?* over de methode schrijft.

Als docent kende ze de problemen van de checklists maar al te goed. Ze kosten veel tijd, ze bestraffen de creativiteit van leerlingen, en ze belonen draken van teksten. '*Forgettably, he crept through the darkness*', is het klapstuk van haar collectie Frankensteinzinnen die de checklists hadden opgewekt.

Daarnaast merkte Christodoulou dat ze fouten maakte bij het beoordelen. Ze gaf zevens aan wat achteraf negens bleken, en negens aan zevens. De checklists hielpen, maar ze losten het probleem niet op; ze verkleinden het hooguit. De consequentie: leerlingen waren – en zijn – behoorlijk afhankelijk van de docent die ze toevallig hebben.

Comparatief beoordelen was sneller en eerlijker. De essays stonden in de juiste volgorde van goed naar slecht (net als de tinten paars). En dat kwam vooral omdat de docenten hun stilzwijgende kennis in het oordeel konden meenemen – de kwaliteiten in de essays die niet in de checklist stonden, maar toch waardevol waren.

Christodoulou werd zó enthousiast, dat ze in 2015 toetrad tot No More Marking, een bedrijf dat comparatief beoordelen in software giet. Inmiddels werkt zo'n vijftien procent van de Engelse scholen met de methode, zegt Christodoulou.



Als een docent beseft dat beoordelen verschrikkelijk moeilijk is

De Nederlands-Vlaamse tegenhanger van No More Marking is het bedrijf Comproved (een samentrekking van *compare* en *improve*), voortgekomen uit wetenschappelijk onderzoek naar comparatief beoordelen.

Groot is comparatief beoordelen in Vlaanderen en Nederland vermoedelijk nog niet. Een vroege Nederlandse gebruiker van Comproved is Gijs Leenders, die ruim acht jaar als docent Nederlands werkte op het Amadeus Lyceum in Utrecht, en nu werkt op de lerarenopleiding aan de Hogeschool Rotterdam.

Leenders hoorde enkele jaren geleden van het concept. Hij besefte dat het kon helpen bij problemen die hij bij zichzelf had gediagnosticeerd: hij was onbetrouwbaar als beoordelaar. ‘Objectief kijken naar teksten is ontzettend moeilijk. Als ik bijvoorbeeld eerst een hele goede tekst lees, en daarna een medium tekst, dan ben ik veel kritischer op de medium tekst.’ Comparatief beoordelen neemt dit probleem weg, doordat elke tekst meerdere keren wordt vergeleken, met betere en slechtere teksten.

Ook hij merkte dat hij de checklists misbruikte. Hij paste de punten aan zodat ze overeenkwamen met zijn intuïtieve indruk. ‘Als ik de checklists toepaste, kwam ik veel te vaak uit op scores waarvan ik wist: dit klopt niet.’ Maar hij moest de checklists invullen: dat was nu eenmaal de werkwijze.

Comparatief beoordelen – vermoedde Leenders – zou leiden tot een waarachtiger analyse van de essays. En docenten zouden inhoudelijkere gesprekken met elkaar voeren. Wat maakt het ene werk, dat net iets hoger op de ranglijst staat, beter dan het werk dat eronder staat?





Leerlingen leren veel van vergelijken, maar ze doen het niet zo vaak

Leenders overtuigde zijn collega's van de vakgroep Nederlands aan het Amadeus Lyceum om met de methode te werken. Enthousiaster misschien nog, raakte hij over comparatief beoordelen *met* leerlingen. Als leerlingen elkaars werk vergelijken, stelt Leenders, doorgronden ze de stof beter.

De procedure is dezelfde als voor docenten. Leerlingen krijgen telkens twee essays te zien en moeten zeggen welke ze beter vinden. En nog eens en nog eens en nog eens, tot de software een stabiele ranglijst detecteert. Het proces van vergelijken is het doel; door het vergelijken begrijpt de leerling de stof beter.

Leerlingen, zegt Leenders, krijgen te zien wat andere leerlingen met dezelfde opdracht hadden gedaan. 'Hierdoor zien ze aanpakken die ze anders niet snel hadden gezien. Foute en goede voorbeelden. Ze genereren op deze manier hun eigen feedback.' Dit leidt tot meer nadenken over teksten. Waarom staat mijn essay hoger of lager dan dit andere essay?

Zo leren de kinderen ook dat er niet altijd kraakheldere criteria mogelijk zijn; dat complexe producten als essays of kunstwerken niet zo goed te gieten zijn in definities of regels. 'Ik merkte dat leerlingen dit met wat oefening heel goed snappen.' Niet zo gek, want leerlingen die comparatief beoordelen, beoordelen nauwelijks anders dan docenten, zo viel Leenders op.

Feedback van de docent blijft soms wel nodig. Maar het proces zelf verricht vaak al veel werk, aldus Leenders. Als leerlingen de essays lezen die net wat beter of slechter zijn dan hun essay, dan gaan ze als vanzelf zien waarom dat zo is. Voor de helderheid: de uiteindelijke beoordeling komt voort uit de comparatieve beoordeling door het docentenkorps.


Bonduelle
Crispy Maïs

Een nieuwe methode moet zich bewijzen – maar hoe?

Is er dan nog iets aan te merken op comparatief beoordelen? Jazeker. Er is nog niet zo veel bewijs dat het tot betere beoordelingen leidt. Dat klinkt ernstig, maar het omgekeerde is ook waar: er is ook nog niet zoveel bewijs dat het niet werkt.

De zaak draait in de kern om de waarde van expertise. Het risico van vertrouwen op de expertise van de leraar is dat beoordelen te subjectief wordt – en dat leerlingen oneerlijk worden beoordeeld. Het risico van de checklist is dat het verwordt tot een doel op zich.

Comparatief beoordelen zit er ergens tussenin. Het is geen glorificatie van intuïtie: docenten aaien niet over hun onderbuik en delen vervolgens cijfers uit. Ze vergelijken essays zonder cijfers te geven. Niettemin: het traditionele houvast van de checklist is weg. En dan is de vraag: welke methode vertrouw je meer? De checklist of een nieuwe methode?

Nadat Leenders vertrok bij het Amadeus Lyceum is de sectie Nederlands weer gestopt met comparatief beoordelen. Het geven van uitleg over cijfers aan leerlingen zonder checklists bleek lastig. Bovendien kwamen de scores van veel essays zo dicht bij elkaar te liggen, dat docenten de ranglijst niet meer vertrouwden. De verschillen hoorden groter te zijn, vonden ze.

Het is ergens wel grappig. Een werkwijze die stoelt op de intuïtieve expertise van docenten moest resulteren in betrouwbare oordelen. Maar die oordelen strookten vervolgens niet met de intuïtie van de

docenten – en dus nam het Amadeus weer afscheid van comparatief beoordelen.

Een nieuwe methode voor een oud probleem is altijd in het nadeel; die moet zijn meerwaarde bewijzen. Maar een markt voor een betere manier van beoordelen is er zeker. Zoals een van de docenten aan het Amadeus zei: 'Ik werk nu al vijftien jaar als docent, en in al die tijd ben ik nog nooit tevreden geweest over de methodes waarmee we het werk van leerlingen beoordelen.'

Binnenkort meer over het hoe en wat van het geven van cijfers. Hoe komen cijfers tot stand, hoe weten docenten of cijfers een juiste inschatting van kwaliteit geven, en hoe ziet het proces van comparatief beoordelen er in meer detail uit?

Meer lezen?

De slechtste baas is de baas die de baas wil worden (en meer wetenschappelijke inzichten over de baas)

De slechtste manier om een goede werknemer te vinden? Een sollicitatiegesprek

Hoe 102 cijfers per jaar de motivatie van leerlingen om zeep helpen

decorrespondent.nl

Je las de printversie van dit verhaal. Voor het volledige artikel met links, infocards, eventuele videos en ledenbijdragen, ga naar https://decorrespondent.nl/16016/het-is-verdraaid-moeilijk-om-in-het-onderwijs-een-eerlijk-oordeel-te-vellen-deze-methode-kan-helpen/42c0ac3f-4838-0f9e-2f8a-0992b64bd3bd

De Correspondent is een dagelijks, advertentievrij medium met als belangrijkste doelstelling om de wereld van meer context te voorzien. Door het nieuws in een breder perspectief of in een ander licht te plaatsen, willen wij het begrip 'actualiteit' herdefiniëren: niet om je aandacht te trekken, maar om je inzicht te bieden in hoe de wereld werkt.

Alle verhalen lezen? Dat kan voor € 9,85 per maand op https://decorrespondent.nl