

D-PAC

event 15 november 2017



Welkomstwoord Sven De Maeyer

DANK, HEEL HEEL VEEL DANK!

Professor Sven De Maeyer gaat in zijn openingswoord zelf niet veel inhoud aanbieden, zegt hij. Hij gaat vooral ceremoniemeester zijn, een rol die hem op het lijf geschreven is. Bij de voorbereiding van zijn openingswoord heeft hij –zoals iedereen doet– plaatjes op internet gezocht. “En zoeken naar plaatjes, dat is vooral vergelijken nietwaar? Welk plaatje past me het best?”

Welcome to the future

Het plaatje dat hij het langst op het scherm laat staan heeft als tekst ‘Welcome to the future’. Toepasselijk, volgens De Maeyer, nu de D-PAC-tool na vier jaar project-af is geworden. De inhoud, dat gaan vandaag andere mensen doen. De Maeyer gaat vooral mensen bedanken. En, zo roept hij de deelnemers op, “...we gaan vooral veel grapjes maken over vergelijken en beoordelen!”

Eerste ideetje

Om te beginnen bedankt hij de Universiteit Van Antwerpen. “Voor het meenemen op het pad om te komen waar we nu gekomen zijn...” Dan regent het pasfoto’s van mensen die bij D-PAC betrokken zijn en waren. “Vanaf het eerste ideetje van Jan Ardies in 2011... Die kwam met ‘er bestaat zoiets als paarsgewijs vergelijken...’ Ja, dachten we, tof idee, hoe vinden we geld om het uit te werken?”

Het woord dat hij het meest gebruikt? Heel heel heel veel dank... Voor de mensen die écht het werk hebben gedaan! De Maeyer vraagt om applaus uit

de zaal. De mensen hoeven die vraag niet te beoordelen: ze klappen meteen!

Door: Richard Derks



Over de toekomst: ondernemers gezocht

INSPIRATIE AWARD NAAR ‘SCOREN EN FEEDBACK VOOR 2.700 TEKSTEN’

Dit is de proef op de som. De deelnemers van het event van vandaag worden zelf beoordelaar. En wel van de 10 posters met praktijkvoorbeelden van D-PAC. We beoordelen de posters en kiezen zo de Inspiratie Award van vandaag (een Dopper-fles en eeuwige roem). Een D-PAC over D-PAC!

“Dit is het spannendste moment van de dag. Hoe evolueert de betrouwbaarheid gedurende het vorderen van de stemrondes?” zegt Sven De Maeyer terwijl hij ingespannen naar de uitkomsten op het grote scherm kijkt. “Het zou toch een nachtmerrie zijn als het cijfer van de betrouwbaarheid nu per stemronde zou teruglopen hè?”



Inspiratie Award

Maar dat is niet het geval. Het cijfer van de betrouwbaarheid springt van -0,44 naar 0 en naar +0,20, +0,30 en verder naar boven. Tot een score van ongeveer 0,75 bereikt wordt. En wie staat er bovenaan in de D-PAC-beoordeling? De Maeyer voert de spanning op. Dan roept hij de twee dames van poster nummer 4 naar voren: Nina Vandermeulen en Brenda Van den Broek van de Universiteit Antwerpen, met hun poster over ‘Scoren en feedback voor 2.700 teksten’.

Kom praten!

“Maar dan nu, na het spannendste moment van de dag, het *moeilijkste* moment van de dag. De toekomst. Hoe gaan we verder met D-PAC?” zegt De Maeyer in zijn afsluiting. Hij roept vooral op om te komen praten. “D-PAC is open source, dus je mag er zelf naar hartenlust mee aan de slag. Maar we willen als bouwers en onderzoekers graag betrokken blijven, blijven leren en verder ontwikkelen.”

Een mogelijkheid, “een ander luikje” noemt De Maeyer het, is om de tool te ‘verkopen’ als werkend pakket. “Mocht u ideeën hebben, mocht u zelf ondernemer zijn of beslissingsbevoegdheid hebben: kom vooral met ons praten!”



Door: Richard Derks



Postersessies

INSPIRERENDE PRAKTIJKVOORBEELDEN

Tijd voor actie. In een live assessment mag het publiek nu D-PAC zelf gebruiken. In de hal hangen posters van tien praktijkvoorbeelden. Welke vinden de aanwezigen het meest inspirerend?

Na de laatste presentatie stroomt de hal waar de posters hangen vol. Al snel staan er voor vrijwel elk bord groepjes te luisteren naar de bevlogen gebruikers. D-PAC blijkt zich te lenen voor uiteenlopende toepassingen: van het toetsen van video-animaties tot spraakproductie bij kinderen en een visie over inclusieve klassen.

Met hun mobiele telefoon kunnen de bezoekers met de D-PAC-tool aangeven welke toepassing ze het meest inspirerend vinden. De uiteindelijke winnaar krijgt straks in het laatste plenaire deel eeuwige roem toebedeeld én mag de enige echte D-PAC Inspiratie Award in ontvangst nemen.

Dit zijn de tien inspirerende onderzoeken

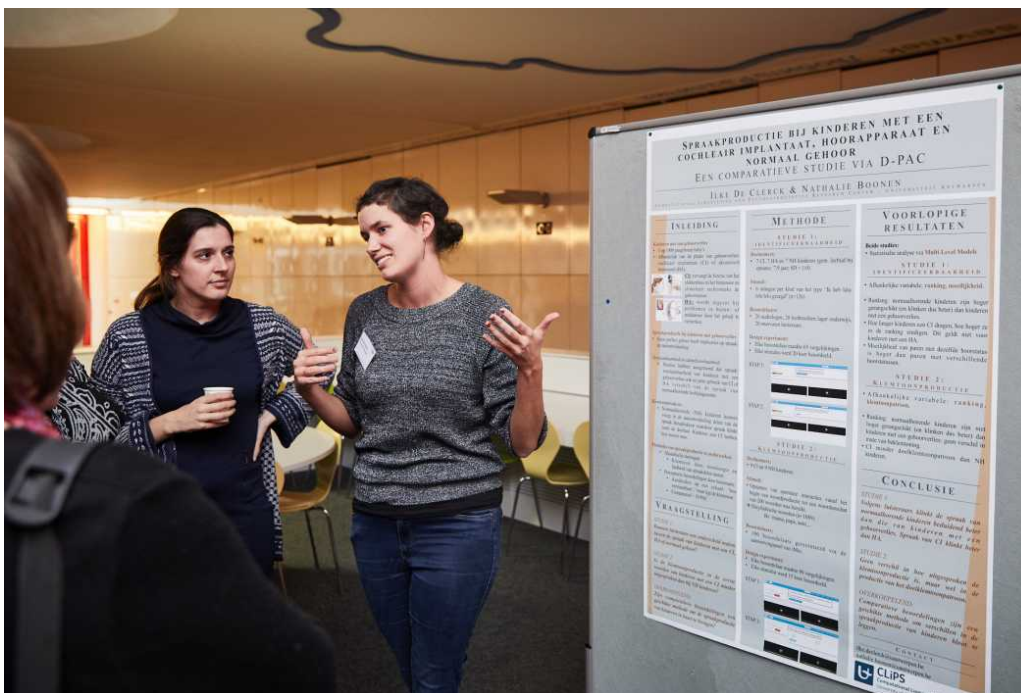
POSTER 1

Spreekvaardigheid in groepswork beoordelen en bespreken

Bij de organisatie van groepswork gaat te weinig aandacht naar het spreken zelf. Praktijkonderzoek waarbij gebruik werd gemaakt van D-PAC toont aan dat 'exploratieve gesprekken' leerlingen beter doen leren, beter doen samenwerken en zowel taalvaardigheid als probleemoplossend vermogen

versterken.

Jan T'Sas, praktijkassistent/PhD student (School of Education, Universiteit Antwerpen)



Ilke De Clerck - Poster 2

POSTER 2

Spraakproductie bij kinderen met een cochleair implantaat, hoorapparaat en normaal gehoor

Kinderen die geboren zijn met (zeer) ernstige gehoorschade krijgen met een Cochlear implantaat (CI) of een akoestisch hoortoestel (HA) de mogelijkheid om alsnog taal en spraak te ontwikkelen. Een CI of HA leidt echter niet tot een perfect gehoor, met verschillen in taalverwerving tot gevolg. In dit onderzoek worden die verschillen blootgelegd door de spraak van kinderen met een CI, HA en normaal gehoor via de D-PAC-tool voor te leggen aan beoordelaars.

Ilke De Clerck & Nathalie Boonen, doctorandi FWO-Aspirant (Universiteit Antwerpen en FWO)



— Jordi Casteleyn - Poster 3

POSTER 3

Het beoordelen van spreekopdrachten op video

Veel leerlingen ervaren angst als ze voor een groep moeten spreken, maar vaak zal de les Nederlands dit niet expliciet behandelen. Daarom hebben deze onderzoekers een innovatieve en compacte training voor het secundair onderwijs gecreëerd die de voordelen van improvisatietheater centraal zet en spreekangst van leerlingen aanpakt. Voor en na de training vulden de leerlingen een vragenlijst in en legden ze een spreekopdracht af die de onderzoekers op video opnamen. De beoordelaars gebruikten D-PAC om deze video's te scoren.

Jordi Casteleyn, Docent Didactiek Nederlands, Didactiek Nederlands aan anderstaligen & Talenbeleid (School of Education, Universiteit Antwerpen)

POSTER 4

Score & feedback voor duizenden teksten

Op deze poster is te zien hoe D-PAC werd ingezet voor het beoordelen en geven van feedback op 2700 teksten. De onderzoekers presenteren hoe een D-PAC-assessment met een klein aantal teksten de basis vormde voor de ontwikkeling van een breed inzetbaar beoordelings- en feedbackinstrument.

Nina Vandermeulen & Brenda van den Broek, PhD-onderzoekers (Universiteit Antwerpen)

POSTER 5

Live beoordelen van fysieke ontwerpproducten in hoger beroepsonderwijs

Om transparantie te bevorderen zijn binnen het hoger beroepsonderwijs analytische beoordelingen met rubrics de laatste jaren sterk in opkomst. Maar

hoe staat het met de validiteit en generaliseerbaarheid van dergelijke beoordelingen? Deze onderzoekers zijn gestart met het beoordelen van door studenten gemaakte ontwerpproducten met D-PAC.

Marion Tillema, docent en onderzoeker (Avans Hogeschool – Academie voor Industrie en Informatica & lectoraat Brein & Leren)



—— Iris Roose - Poster 6

POSTER 6

De professionele visie van (student)leraren op inclusieve klassen

Is een leraar in opleiding in staat om te herkennen wat inclusieve interacties zijn en welke werkvormen daarbij horen? Voor het onderzoek van 'POTENTIAL-Power to Teach All' beoordeelden student-leraren filmpjes van klassen. Hun oordeel werd afgezet tegen dat van ervaren leerkrachten.

Esther Gheysens, onderzoeker Potential (Vrije Universiteit Brussel) & Karolien Keppens & Iris Roose, onderzoekers Potential (Universiteit Gent)

POSTER 7

Videoanimaties beoordeeld door een jury met verschillende expertises

Een animerende poster met veel beeld. Voor het vak Animation 1 van de opleiding Multimedia Technology aan de Karel de Grote Hogeschool kregen de studenten een fragment uit het hoorspel De Bokrokker van het Geluidshuis. Het was hun opdracht het fragment te voorzien van animatie. De resultaten werden beoordeeld door animatie-professionals, mensen van het Geluidshuis en docenten van de opleiding.

Ivan Waumans, Lector Multimedia Technology & Creative Technologies and Entrepreneurship (Karel de Grote hogeschool)

POSTER 8

Peer feedback met D-PAC: ervaringen van studenten en docenten in professionele communicatie

Een populaire poster, zo op het eerste oog. Veel mensen zijn benieuwd naar de studenten van de master Meertalige Professionele Communicatie. Zij gebruikten D-PAC om elkaar feedback te geven op een schriftelijke opdracht over klachtencommunicatie. Die input vormde de basis voor een feedbackcollege.

Suzy Stals, praktijkassistent Professionele Communicatie & Luuk Van Waes, hoogleraar Professionele Communicatie, TEW – Departement Management (Universiteit Antwerpen)

POSTER 9

Papers op een bestaande schaal plaatsen met een nieuw algoritme

Een poster met wat vouwlijnen (ontroerend). Dan weet je dat hij waarschijnlijk een hele route heeft afgelegd, net als de onderzoekers die hem gemaakt hebben. In dit geval bestond het parcours uit het maken van een rangorde van essays. Een nieuw algoritme werd vervolgens ontwikkeld om te bepalen hoeveel vergelijkingen nodig zijn om nieuwe essays aan die bestaande rangorde toe te voegen.

Antony Furlong, Manager Assessment Research and Design (International Baccalaureate)

POSTER 10

Welke tools voor technische innovatie?

Hoe bepaalt een team dat tools ontwikkelt welke tools het zelf gebruikt? Het zou een dief zijn van zijn eigen innovaties als het daar zijn eigen tools niet voor gebruikte. D-PAC dus!

Tanguy Coenen, Application prototyping team manager (imec) & Camille Reynders, programmeur

Door: Mirjam Streefkerk & Elske van Lonkhuijzen



Presentatie Renske Brouwer

“DEZE WIL IK NIET MEER ZIEN!”

Dit praktijkgeval zou eigenlijk worden gepresenteerd door een duo: Anneleen Mortier en Renske Bouwer. Maar Anneleen is zojuist moeder geworden. En ze is gepromoveerd. Dus Renske staat er eventjes alleen voor. Ze slaat zich in rap tempo door een aantal slides.

Uit haar presentatie blijkt dat de D-PAC-tool ook gebruikt kan worden om betrouwbare selecties te houden. Bijvoorbeeld bij sollicitaties, waarbij je goede CV's wil selecteren, of bij langere portfolio's of projectvoorstellen.

Experts zijn sneller betrouwbaar

Om te bekijken of dat D-PAC een betrouwbare selectie oplevert, is gekeken naar het selecteren van CV's bij een echte vacature, voor een specialist bij het bedrijf Hudson. De beoordelingen werden gedaan door 7 experts van het bedrijf zelf, en als vergelijking door een groep van 57 'novices', studenten. Daar kwam vooral uit naar voren dat de experts minder vergelijkingsrondes nodig hebben (10 rondes) dan de studenten (23 rondes). Experts kijken vooral naar relevante ervaring en studenten naar zaken als de vormgeving van het CV.

Top-5

Opvallende uitkomst van het comparatief beoordelen van CV's was dat zowel de studenten als de experts uiteindelijk een hoge mate van betrouwbaarheid bereiken. De top-5 van de sollicitanten, die ook op gesprek zijn geweest bij de

werkgever, zaten in de top van de curve zoals aangewezen door de D-PAC-tool. Er is ook gekeken naar efficiëntie, daar ontlopen de traditionele manier van selecteren en de D-PAC methode elkaar niet veel.

Iedereen ziet alles

Hetzelfde patroon is te zien bij de selectie van (grote) portfolio's en projectvoorstellen. Ook bij deze omvangrijke casussen komt de D-PAC-tool al vrij snel tot een betrouwbare rangorde. Bijkomende voordelen zijn dan dat de beoordelaars de methode ervaren als iets minder tijdsintensief dan traditionele manieren van selectie. Als ander voordeel wordt genoemd dat iedereen alle portfolio's en projectvoorstellen onder ogen heeft gehad, en dat er een grote transparantie naar de indieners van de plannen bereikt wordt.

'Flaggen'

De ontwikkelaars van D-PAC zijn op zoek naar manieren om de efficiëntie van de D-PAC-tool te verhogen. Renske Bouwer en het team van D-PAC kijkt naar twee punten waar winst valt te halen: het inzetten van een filteralgoritme en van een plaatsingsalgoritme. Dat filteralgoritme plaatst een 'flag' bij een voorstel of CV als dat voorstel redelijk snel als aantoonbaar onder de maat wordt gezien. De flag is dan 'deze willen we niet meer zien!'. Een plaatsingsalgoritme kan zorgen dat een nieuw te beoordelen casus sneller 'instroomt' in een bestaand beoordelingstraject. Ook dat betekent tijdswinst.

Is dat wel eerlijk?

Het plaatsen van de flag levert een interessante vraag uit de zaal. Is dat wel eerlijk? Zorgt het niet voor een 'vervuiling' van de resultaten om een casus bij voorbaat uit te sluiten? "Dat klopt..." zegt Bouwer. "Daar moet je inderdaad goed naar kijken." Eerdere en betere feedback kan hiervoor de oplossing zijn. "Een tool als D-PAC is nooit af, het kan altijd beter!"

Door: Richard Derks



Presentatie Roos Van Gasse

HOE D-PAC-DATA KAN HELPEN BIJ VERDERE PROFESSIONALISERING

Dat de data die met D-PAC wordt vergaard ook gebruikt kan worden bij verdere professionalisering van assessments en organisaties, laat onderzoeker Roos Van Gasse zien met een mooi praktijkvoorbeeld bij AHOVOKS, de Vlaamse Examencommissie.

Organisaties lopen bij hun assessments regelmatig tegen betrouwbaarheids- en validiteitsproblemen aan, legt Van Gasse uit aan haar publiek.

"Beoordelaars hebben dan bijvoorbeeld niet naar dezelfde dingen gekeken of er zijn verschillen in hoe streng ze zijn. Dit soort problemen hoort erbij, beoordelen is gewoon moeilijk. Het grootste probleem is eigenlijk dat gangbare methoden die deze problemen onderzoeken veel tijd en geld kosten."

Maar, gelukkig is er D-PAC. Daarmee kun je snel inzicht krijgen in de kwaliteit van assessments. D-PAC levert namelijk veel data op. Denk aan de prestatiescore, de eensgezindheidsscore, de mate waarin beoordelaars het product gelijkwaardig beoordelen, maar ook aan gegevens over de beoordelaar: hoeveel tijd is die ermee bezig geweest en in hoeverre wijkt hij af van de consensus?

In de praktijk

Bij AHOVOKS onderzocht Van Gasse hoe deze data in de praktijk kon worden gebruikt om assessments te verbeteren. Hun vraag: in hoeverre verschillen de beoordelingen van interne beoordelaars en freelancers? Bij de afdeling Engels werd een try-out gedaan, waarbij 33 informele brieven werden beoordeeld. Intern waren er geen afwijkende beoordelaars, extern wel: 6 van de 16.

"We hebben toen aan beide groepen gevraagd welke aspecten ze belangrijk vinden bij een informele brief. Daarbij bleken de verschillen tussen de internen en externen behoorlijk groot." De consensus was dat inhoud een belangrijk criterium is maar de misfits vonden dat veel minder belangrijk.

Er bleek dus ruimte voor verbetering. In een vervolgssessie gingen de beoordelaars met elkaar in paren en later in groepen in gesprek over beoordelingen. Er werden vragen gesteld en er werd kennis gedeeld. Met succes. In een volgend assessment waren alle misfitters erop vooruitgegaan.

Nuttige data

"Voor D-PAC betekent het dat het zin heeft om die misfitmaat te hanteren, en dat gerichte interventies het mogelijk maken om ze eruit te halen. D-PAC levert dus voor gebruikers nuttige data aan om assessmentprocessen bespreekbaar te maken."

Van Gasse benadrukt dat het door haar geschetste voorbeeld slechts een van de vele functionaliteiten van D-PAC is. "Denk aan leerlijnonderzoek of 360 gradenfeedback. Helaas hadden we te weinig tijd om dat verder te onderzoeken."

Door: Mirjam Streefkerk



Presentatie Maarten Goossens

LEREN VAN LEEFTIJDGENOTEN

Daar duikt dan toch nog een Steve Jobs-coltrui op. Maarten Goossens betreedt het podium met een missie: ons overtuigen van de waarde en haalbaarheid van D-PAC als tool voor peer assessment.

Peer assessment: mission impossible?

Laten we maar eens beginnen met de dingen waar docenten tegen aanlopen wanneer ze studenten elkaars werk willen laten beoordelen. Hoe leer je studenten werken met het beoordelingsinstrument dat je gebruikt (bijv. een Rubric's Cube)? Zetten studenten elkaar niet onder druk (peer pressure) en nemen ze elkaars feedback wel aan?

Bottleneck voor de docent: tijd en kwaliteit

Nog twee bezwaren van docenten: een peer assessment kan al snel verschrikkelijk veel tijd kosten om te organiseren en dan is het nog maar de vraag of de kwaliteit van de oordelen wel goed is. Reden voor velen om er maar niet aan te beginnen. En dat terwijl het zoveel kan opleveren.

Mission possible! (D-PAC to the rescue)

Eerst maar eens zien of studenten elkaars werk via comparatief vergelijken eerlijk kunnen beoordelen. Onderzoek met D-PAC wijst uit van wel. In één case vergeleken zowel studenten als docenten, van een opleiding tot architect, de kwaliteit van moodboards. Wat bleek: de rangordes van beide groepen waren even betrouwbaar. De validiteit verschilde nog wel wat:

studenten lieten zich afleiden door wat minder relevante aspecten (bijvoorbeeld lay-out en taalgebruik). Maar, voorspelt Goossens, met meer oefening zouden ze daar nog in kunnen groeien.

Time is on your side (met D-PAC)

Dan dat andere bezwaar: de tijd die het opzetten van een peer assessment kost. Ook daar blijkt comparatief vergelijken dé oplossing voor. Studenten hebben geen training nodig, ze kunnen hun eigen werk uploaden, het algoritme van D-PAC verdeelt de taken (zodat uitval wordt opgevangen) en verwerkt de resultaten. De docent kan via de tool monitoren hoe het gaat en de input en feedback van studenten gebruiken in zijn of haar eigen oordeel. In een rangorde is dan vooral nog aandacht nodig voor het omslagpunt van onvoldoende naar voldoende.



— Maarten Goossens

Studenten leren van comparatief vergelijken

Het ultieme doel is natuurlijk dat studenten door peer assessment van elkaar leren. En dat doen ze gelukkig ook. Het geven van feedback laat hen zien hoe medestudenten een opdracht hebben aangepakt, waar ze staan ten opzichte van die medestudenten en het dwingt hen om te formuleren wat ze vinden (en zo op een dieper niveau informatie te verwerken). Het krijgen van feedback met D-PAC zorgt voor input op hoger niveau (complexe aspecten). Studenten ervaren de feedback die ze via de tool krijgen als betrouwbaar, relevant en eerlijk.

En wat die peer pressure betreft: de tool verwerkt alle beoordelingen anoniem.

‘Je kan je mening geven zonder afrekening,’ aldus Goossens.

De dwarsliggers filter je er gewoon uit

Twee vragen uit het publiek. ‘Kun je outliers uit de resultaten verwijderen,’ vraagt iemand, ‘bijvoorbeeld als een student de opdracht niet serieus neemt en zo de rangorde beïnvloedt?’ Ja, dat kan, zegt Goossens. De tweede vraag is of de PowerPoint beschikbaar komt. Waarschijnlijk omdat er een handzaam schema in staat voor hoe een docent door een jaar heen peer assessments zou kunnen organiseren. ‘Uiteraard’, zegt Goossens, ‘de PowerPoint komt later online.’

Door: Elske van Lonkhuijzen



Inleiding door Renske Bouwer

MAAR DAN NU... DE PRAKTIJK

Oke. Prima. We hebben in het ochtendprogramma de theorie van comparatief beoordelen voor de kiezen gehad. Maar wat kunnen we er in de praktijk nou precies mee? Veel, blijkt uit de inleiding van Renske Bouwer. Ze geeft aan hoe de D-PAC-tool in de afgelopen vier jaar in de praktijk is ingezet, en door wie.

Het begon in 2014 met een proef in het onderwijs. Een eerste praktijktoets van de D-PAC-tool. Op 10 scholen werd een schrijfpdracht van de studenten –‘argumentatief schrijven’– met de tool bekeken. Er deden 135 leerlingen mee, en 83 beoordelaars. De eerste echte studies in de praktijk vonden in 2015 plaats. 189 deelnemers deden mee, er waren 120 beoordelaars betrokken.

Vleugels

In 2016 kreeg D-PAC echt vleugels. In 33 assessments deden 1.249 deelnemers mee die werden beoordeeld door 1.236 beoordelaars. In dit jaar ging D-PAC ook internationaal, zegt Bouwer lachend: “Als je de Universiteit van Amsterdam tenminste ziet als ‘internationaal’. Maar ook Duitsland en Cyprus”. Ze laat op een dia een grote hoeveelheid deelnemende partijen zien. Door de zaal gaat een golf van herkenning: de futuristische stoeltjes draaien en wiebelen als de bezoekers van het event van vandaag hun logo herkennen.

Cijfers... punten...

In 2017 vindt er bijna een verdubbeling van de deelnemers aan de D-PAC-tool plaats. In 93 assessments worden 4.000 deelnemers comparatief beoordeeld. In dit jaar doen voor het eerst ook HBO-scholen mee. “De D-PAC-tool wordt

vooral gebruikt in het onderwijs...” zegt Bouwer “En vooral veel door onderzoekers. We zien dat 60% van de assessments worden gebruikt voor onderzoek.” De methode kan ook voor andere dingen worden gebruikt. Om bijvoorbeeld complexe groepsbeslissingen in organisaties te nemen legt Bouwer uit. Het is ook in te zetten voor het bepalen van cijfers. “Of ‘punten’, zoals de Vlaming zegt...” zegt de (Nederlandse) Renske Bouwer.

Feedback voor de ontwikkelaars

“We hebben in de afgelopen vier jaar veel geleerd. Niet in de laatste plaats over de methode zelf.” zegt Bouwer. Hoe meer gebruikers er komen, hoe meer informatie er terugkomt uit het proces van comparatief beoordelen zélf. Dat komt ten goede aan de tool, legt Bouwer uit. Feedback voor de ontwikkelaars zelf!

In de rest van de middag diepen we drie praktijkgevallen verder uit, vertelt Bouwer tot slot. Zijn we er klaar voor? “Ready mensen?” vraagt ze. “Go!” zegt de zaal.

Door: Richard Derks



Presentatie Marije Lesterhuis

“HET RECHTERPOPPETJE HEEFT VOETJES.”

Marije Lesterhuis zet ons meteen voor het blok. Twee kindertekeningen. Links een woelig gezelschap van kopvoeters, rechts een keurig ingekleurd poppetje met geel haar. Het publiek is verdeeld. De linkertekening is organisch, creatief, experimenteel. De rechter blinkt uit in kleurgebruik en details. ‘De anatomie klopt beter.’ Het poppetje heeft zelfs voetjes. Zie hier het probleem met validiteit. Wat beoordeel je eigenlijk als je beoordeelt? En verschillen beoordelaars daarin niet enorm?

Een betrouwbare rangorde is peanuts als iedereen op hetzelfde let

Inhakend op de presentatie van San Verhavert over betrouwbaarheid weet Lesterhuis ons te vertellen dat een hoge betrouwbaarheid prima te bereiken is, mits alle neuzen dezelfde kant op staan. Maar beoordelaars vinden soms heel verschillende dingen belangrijk. Waarin verschillen ze precies en krijg je ze met comparatief beoordelen op één lijn? Dat heeft zij de afgelopen vier jaar onderzocht.

Met de convergente validiteit zit het wel snor

De wat? De convergente validiteit. Die drukt uit of een methode overeenkomt met andere methoden die hetzelfde meten. Lesterhuis (& co) zetten resultaten van comparatief beoordelen (van teksten) af tegen drie andere meetmethoden: ankerteksten, rubrieken en criterialijsten. De correlatie was hoog. Waarom dan toch kiezen voor comparatief beoordelen, vraagt iemand in het publiek zich af. Is dat niet een beetje ‘wij van wc-eend adviseren u wc-

eend'? In principe zou je voor een andere methode kunnen kiezen, maar wat comparatief beoordelen uniek maakt, is dat het snel en makkelijk op te zetten is. Aan een effectieve criterialijst, schaal of rubriek wordt soms wel jaren gewerkt.

Beoordelen we wat we willen beoordelen?

Letten beoordelaars op alle aspecten in een complexe vergelijking? Letten ze op de meest relevante aspecten en waarin verschillen ze? Uit een assessment van argumentatieve teksten bleek dat de meeste beoordelaars letten op de argumenten en de tekststructuur. Redelijk veel consensus over wat belangrijk is dus. Er waren echter ook verschillen. De ene beoordelaar vindt taalgebruik belangrijk, de ander focust op het gebruik van bronnen. Dit blijken echter nuanceverschillen die zeker overkomelijk zijn. Overall letten beoordelaars vooral op inhoud en dan pas op vorm.

Hoe complex is comparatief beoordelen?

Soms ligt de kwaliteit van twee producten heel dicht bij elkaar. Dan is kiezen lastig. Wat een beoordelaar ziet (het duo) bepaalt waar hij op let. Is in twee kindertekeningen het kleurgebruik in orde, dan zal de afweging dus eerder over anatomie gaan, bijvoorbeeld over voetjes.

Validiteit optimaliseren in de praktijk

Hoe krijg je al je beoordelaars in het gareel? Lesterhuis sluit af met een paar praktische take-aways. Bepaal voorafgaand aan het assessment al waar je je rangorde voor wilt gebruiken, welke beoordelaars in staat zijn op relevante criteria te letten en geef ze een heldere instructie mee. Na afloop kun je kijken of de scores uit jouw assessment overeenkomen met scores uit andere methodes, of je beoordelaars daadwerkelijk op één lijn zitten en bij hen navragen hoe ze het ervaren hebben.

De muis kan terug naar de eigenaar

De presentatie begon met een toetsenbord dat niet werkte. 'Is er een muis in de zaal?' In de kleine battle die volgde tussen toetsenbord en muis won: de muis. Het publiek heeft aandachtig geluisterd en er is genoeg om op te kauwen tijdens de lunch (niet alleen al die grappige broodjes).

Door: Elske van Lonkhuijzen



Presentatie San Verhavert

HOE BETROUWBAAR IS COMPARATIEF BEOORDELEN?

Verhavert begint met een kwinkslag. “Alle taalfouten die in deze presentatie staan zijn louter voor uw amusement!” Met de taalfouten valt het best mee, blijkt al snel, want Verhavert laat vooral veel cijfers zien. Nee, niet meteen wegklikken mensen: cijfers zijn interessant!

Vraagtekens...

Verhavert laat ons even schrikken... Hij gaat het hebben “over de letter alpha... over interne consistentie... of over SSR... Maar...” zegt hij droogjes, “Als ik dat doe zit u er al snel ZO bij...” Hij laat een plaatje zien van een verward poppetje met een vraagteken boven zijn hoofd, staand op een berg vraagtekens... “Maar dat ga ik niet doen” gaat hij verder. “Ik hoop dat ik u met mijn presentatie kan overtuigen van het feit dat comparatief beoordelen zowel betrouwbaar als efficiënt is.”

Rangorde: Sliert van streepjes

Dat overtuigen doet hij met argumenten. Het is namelijk wetenschappelijk onderzocht. Hij laat een plaatje zien dat we vandaag vaak tegenkomen. Een sliert van verticale streepjes op een horizontale grafiek. De schaal is de score uitgezet tegen de representaties. Je ziet zo in één oogopslag de uitkomst van een groot aantal comparatieve beoordelingen van een bepaalde taak of product. De streepjes drukken kortweg uit hoe ‘zeker’ je kunt zijn van een beoordeling. Hoe korter de streepjes, hoe betrouwbaarder de uitkomst.

Verhavert noemt een cijfer. “Mijn resultaten zijn 0,87 zeker... Ik hoor u denken... Ja? Nou? 0,87?”

Cijfermatige betrouwbaarheid

Dan duikt Verhavert de diepte in. Je kunt een vergelijking laten doen door twee groepen van beoordelaars. Daar komen verschillende slierten uit, met andere gemiddelde getallen. Die getallen kun je middelen. Je kunt het proces van beoordelen laten doen door een grotere groep beoordelaars. En dan de uitkomst weer middelen. Uiteindelijk wordt de betrouwbaarheid groter en groter. Op een schaal van 0 naar 1 komt er bij het voorbeeld dat Verhavert noemt 0,87 uit. Hoe hoger dat getal, hoe minder de betrouwbaarheid afhangt van de beoordelaars. Hoe hoger het getal, hoe hoger de betrouwbaarheid van het gevolgde proces.



Optimum voor efficiëntie

Volgende punt. *Hoeveel* beoordelaars heb je eigenlijk nodig om een betrouwbare schaal te krijgen? En hoe veel werk moeten ze eigenlijk verzetten? Hoeveel expertise hebben ze nodig? Kortom: hoe efficiënt is het eigenlijk om het op deze manier te doen? Nou. Dat heeft Verhavert ook bekeken. Hij tovert een curve op het scherm. Een lijn die langzaam afvlakt... In gewone mensentaal: na een bepaald aantal beoordelingen per representaties wordt een maximale betrouwbaarheid bereikt, bij de hoeveelheid werk of bij de expertise van de beoordelaars.

Expertise van minder belang

En het opvallende is: als je naar die curve kijkt zijn er eigenlijk maar twee

factoren van belang voor de betrouwbaarheid van een beoordeling: het aantal vergelijkingen per representatie en het aantal vergelijkingen per beoordelaar. Vreemd genoeg heeft de expertise van de beoordelaar nauwelijks invloed op de betrouwbaarheid. Als je hetzelfde product (bijvoorbeeld een tekst of een mailtje) door een groep peers laat beoordelen en door een groep docenten, kan de uitkomst (de vorm van de sliert) iets verschillen, maar de betrouwbaarheid –het gemiddelde cijfer dat er uit komt rollen op die schaal van 0 naar 1– is uiteindelijk gelijkwaardig.

Vraag uit de zaal

“Nou...” Sluit San Verhavert af... “Ik hoop dat ik jullie niet het gevoel heb gegeven van een ‘system overload’... Dat valt reuze mee. Er komen uit de zaal eigenlijk maar twee vragen. De belangrijkste: waar zit ‘m precies de efficiëntie van het comparatief vergelijken? De hele aanpak lijkt toch behoorlijk arbeidsintensief... “Dat valt reuze mee” zegt San Verhavert “Het is contra-intuïtief, maar analytisch kijken naar zeg 100 producten, bijvoorbeeld aan de hand van een uitgewerkte lijst met criteria, kost even veel tijd als comparatief beoordelen met dezelfde betrouwbaarheid in de uitkomst.”

Door: Richard Derks



Demo van de tool

D-PAC IN A NUTSHELL

Een gloednieuwe tool demonstreren. Hoe doe je dat met klasse? Tanguy Coenen, application prototyping team manager bij imec, pakt zijn Steve Jobs-moment. Geen zwarte coltrui maar een kleurrijke houthakkersblouse. De zelfverzekerde maar bescheiden toon van iemand die zijn werk door en door kent.

De beoordelaar en zijn missie

Coenen neemt ons mee de tool in. Op het scherm verschijnt de openingspagina waar nieuwe gebruikers een tutorial kunnen volgen. Die is vandaag niet nodig: Coenen legt live uit hoe het werkt. De belangrijkste functie is natuurlijk comparatief beoordelen. Beoordelaars kiezen een assessment. Bijvoorbeeld van argumentatieve teksten over orgaandonatie. De tool laat zien welke opdracht studenten precies kregen, maar ook welke opdracht de beoordelaar heeft.

A of B? A of B? A of B?

De te beoordelen producten (in deze demo teksten) verschijnen steeds in tweetallen in het beeld. Welke tekst is beter? Klik op de blauwe knop. De tool houdt bij hoeveel vergelijkingen de beoordelaar al gemaakt heeft. Na de keuze welke beter is, kan de beoordelaar aanvullende feedback geven in aparte tekstvelden.



— Tanguy Coenen

Inzicht in de statistieken

Tientallen teksten, meerdere beoordelaars, een hele reeks vergelijkingen. Reken maar dat daar veel data uit rolt. Aan Coenen en zijn team de taak dit behapbaar te presenteren. Tijdens het event aan het publiek, maar natuurlijk ook aan de gebruikers in de tool. Voor elk assessment zijn onder andere de rangorde van beoordeelde producten en de betrouwbaarheid te zien, met nog een keur aan andere cijfers.

De tool in de praktijk

‘Ik zal jullie de technische details besparen van hoe de tool is gebouwd,’ besluit Coenen zijn demo. ‘Zijn er nog vragen?’ De futuristische stoeltjes in Aula 1 komen in beweging. Ja, meer dan genoeg. ‘Kunnen we veel Wordbestanden tegelijk uploaden?’ ‘Kan er commentaar in de tekst gegeven worden?’ ‘Kunnen studenten hun tekst zelf uploaden?’ Het is duidelijk: men popelt om D-PAC in te zetten in de praktijk. Mooi iets om tijdens de koffiebreek over door te praten.

Door: Elske van Lonkhuijzen



Presentatie Liesje Coertjens

DE KRACHT VAN PAARSGEWIJS BEOORDELEN

De titel van de presentatie van professor Liesje Coertjens is ‘De kracht van paarsgewijs vergelijken voor het beoordelen van competenties’. Er staat inmiddels ‘Prof.’ voor haar naam. Ze gaat in een presentatie van een minuutje of twintig een helicopterview over het onderwerp geven. Dat doet ze aan de hand van een leuk mailtje...

Moeilijke zaken

“Moeilijke zaken beoordelen. Hoe doe je dat? Een portfolio, een CV, hoe beoordeel je dat? Is iemand klantgericht? Spreekvaardig? En waarom is dat precies zo moeilijk?” Zo begint Coertjens. Verschillen we bijvoorbeeld als beoordelaars niet teveel van elkaar? bijvoorbeeld. De tweede belangrijke factor: kijken we wel naar de belangrijke elementen met een juist gewicht? ...Dit noemt de professor de ‘validiteit’.

Betrouwbaar beoordelen

Liesje gaat verder met het benoemen van ‘families’ van manieren van beoordelen. Bij de eerste grote familie wordt ‘holistisch’ beoordelen tegenover ‘analytisch’ beoordelen gezet. Ze laat een voorbeeld zien. Een mailtje van een leerling met een vraag. *Ik heb niet genoeg smikkelpunten, maar ik wil toch een CD...* Door de fouten en de kromme zinnen heen kan een beoordelaar gericht kijken naar dit mailtje. Naar een aantal ‘grote’ punten aan de hand van een criteriumlijst. De inhoud, de structuur, de stijl... Maar het hangt erg van de beoordelaar af hoe precies die beoordeling verloopt.

De betrouwbaarheid van een beoordeling is binnen deze familie een lastig ding. Door analytisch of holistisch naar een mailtje als dit te kijken is het probleem van de betrouwbaarheid in ieder geval niet opgelost. Al besteed je er 20 pagina's met beoordelingscriteria aan (let op, aan één mailtje!): het probleem van de betrouwbaarheid en validiteit blijft bestaan.



Bestaand onderzoek

De professor duikt vervolgens –daar is ze professor voor, nietwaar– in het onderzoek en de literatuur rond beoordelen en vergelijken. Yeates bijvoorbeeld. Die heeft gekeken naar het beoordelen van video's door expert-beoordelaars. De volgorde van beoordelen blijkt daarbij een rol te spelen, en dat is toch eigenlijk onwenselijk. Gemiddelde video's bijvoorbeeld, worden hoger beoordeeld als ze na een slechte video volgen dan na een goede video.

Wat te doen? Paarsgewijs vergelijken, zegt Liesje Coertjens. De tweede grote familie van beoordelen. Ze keert terug naar het mailtje over de smikkelpunten. Neem twee mailtjes en vergelijk ze. Welke is beter? Het is handiger (geen getob over de precieze criteria) en sneller (tot 5x sneller!) om het beoordelen

paarsgewijs te doen. Als een groepje van 4 beoordelaars 4 tekstjes beoordeelt (dat is 24 beoordelingen op 'goed' of 'slecht') komt daar een mooie wolk als resultaat uit. Een rangorde die betrouwbaarder is dan alleen holistisch of alleen analytisch beoordelen.

Flirten binnen de families...

Er zijn binnen de familie van 'paarsgewijs beoordelen' weer een aantal takken te benoemen: paarsgewijs vergelijken kan random, of rekening houdend met een 'voorlopige score', met benchmarks of op andere manieren. En jawel, met de verwantschappen tussen de families kun je 'een beetje flirten', besluit Coertjens... Een beetje meer holistisch... Een beetje meer analytisch... En dat gaan we vandaag doen!

Dan stapt Sven De Maeyer naar voren. Hij laat Liesje Coertjens niet met lege handen gaan. Hij heeft –jawel– twee cadeautjes voor haar. Ze moet kiezen! Want als iemand het kan, dan is het degene die aan de wieg van het project (en de tool) D-PAC stond. En wat doet Liesje? Jawel: ze pakt BEIDE cadeautjes aan!

Door: Richard Derks



Voorwoord

HET SUCCES VAN D-PAC

Het is dringen bij de koffietafel. Al die bekertjes! Welke te kiezen? Welke is het lekkerst, welke smaakt het best? Als u de afgelopen vier jaar goed heeft opgelet moet dit een makkie voor u zijn. U heeft inmiddels de tools in handen om succes meetbaar te maken!

Spreken, luisteren, vergelijken...

Vandaag maken we het ongrijpbare grijpbaar. Het is de afsluiting van het project D-PAC, of 'Digital Platform for the Assessment of Competences'. Hoe meten we zoiets ongrijpbaars als 'succes'? En hoe bepalen we het succes van D-PAC zélf?

Daar gaan we vandaag eens lekker over vergelijken. Vandaag houden we een dag van spreken en luisteren, morgen gaan we zelf aan de slag met workshops en concrete opdrachten!

Door: Richard Derks

