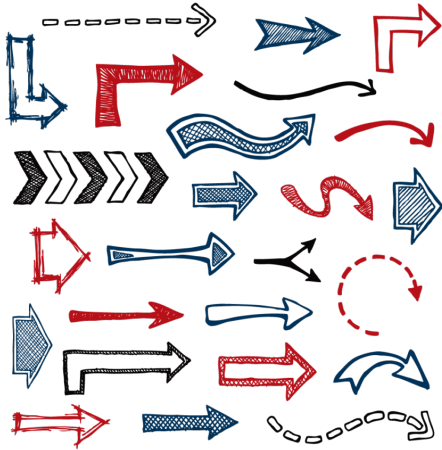


Lezersonderzoek (deel 2): methoden

20 januari 2021 LEO LENTZ



Ik beschrijf tien methoden om te onderzoeken of een document, website of tool begrijpelijk, aantrekkelijk, overtuigend en gebruiksvriendelijk is. Wie verder zoekt, vindt beslist nog andere methoden. Wanneer gebruik je welke methode?

De kernvraag is: Wat is het doel van het document? Wat moet de lezer met het document doen?

Keuzehulp

- Lezers *begrijpen* een tekst
 - Wat is het doel van de test?
 - Optimalisering -> Lezersprotocol, Focusgroep, Plus-en-minmethode eventueel in combinatie met Interview
 - Keuze tussen tekstvarianten of beslissing; welke tekstvariant zullen we gebruiken in de campagne: go - no go -> Cloze-test of Begripsvragenlijst

- Lezers kunnen een document *toepassen* op hun eigen situatie en/of er handelingen mee verrichten
 - Wat is het doel van de test?
 - Optimalisering -> Hardop denken als er handelingen verricht moeten worden + Observatie + Interview achteraf
-> Focusgroep of Plus-en-minmethode als er geen handelingen verricht moeten worden.
 - Beslissing go - no go -> Observatie met taken en Vragenlijst met meerkeuze-vragen
- Lezers voeren acties uit met het document, zoals bij een formulier of een tool/app
 - Wat is het medium?
 - Digitaal op een scherm -> Hardop denken tijdens uitvoering taak + Observatie + eventueel Eye tracking
 - Op papier -> Hardop denken tijdens uitvoering taak + Observatie
- Lezers *navigeren* door het document, een website of tool op zoek naar bepaalde informatie
 - Wat is het doel van de test?
 - Optimalisering -> Plus-en-minmethode + Interview of Focusgroep
 - Beslissing go - no go -> gesloten Vragenlijst naar attitudes en opinies

10 methoden

1. Cloze-test

De cloze-test kent een lange geschiedenis, is een bepaalde periode verhuisd geweest, maar heeft nu toch weer erkenning gekregen in het wetenschappelijk onderzoek. De kern van deze methode is dat de onderzoeker systematisch woorden weglaat in de tekst en dat de proefpersoon moet proberen het juiste woord bij elk gat in de tekst in te vullen. Het belangrijkste kritiekpunt was dat lezers de juiste woorden konden invullen zonder de tekst te begrijpen. Zij zouden ook niet genoodzaakt worden verschillende zinnen in de tekst met elkaar te verbinden en zo op een hoger niveau tot een goede interpretatie te komen. Kamalski (2007) en Kleijn (2018) hebben echter laten zien dat die bezwaren weggenomen kunnen worden als de onderzoeker wat slimmer met deze methode aan het werk gaat. Dit zijn de slimme ingrepen:

- In de titel en de eerste zin van de tekst creëer je geen gaten, zodat de lezer een goede verwachting kan opbouwen over de inhoud van de tekst.

- Vermijd het weglaten van volkomen voorspelbare woorden (*Dat had hij niet goed de gaten*).
- Vermijd het weglaten van volkomen onvoorspelbare woorden (*De wind kwam uit het en wakkerde behoorlijk aan*).
- Vermijd dat hetzelfde woord een paar keer achter elkaar in de tekst ingevuld moet worden.
- Doseer het aantal gaten in de tekst tot ongeveer een op elke tien worden.
- Gebruik bij de beoordeling van de antwoorden je gezonde verstand en reken ingevulde woorden ook goed als ze de *betekenis* van het oorspronkelijke woord goed weerspiegelen.

Suzanne Kleijn heeft in haar proefschrift een bijlage opgenomen met een uitwerking van deze slimme ingrepen. Waaruit bleek nu dat dit een goede methode is om begrip te testen? Het eerste kritiekpunt, dat deze test alleen zicht geeft op lokaal begrip van lezers en dus niet op hogere tekstniveaus is bestreden met een complexe taaltechnologische en statistische analyse, waaruit bleek dat de computer niet in staat was om te voorspellen welk woord ingevuld moest worden. Daaruit mag je concluderen dat er dus wel degelijk enig begrip van de hele tekst nodig is voor goede scores. Maar belangrijker is dat de resultaten van cloze-tests bij grote groepen scholieren heel goed correleerden met andere variabelen, zoals hun leesvaardigheid en hun woordenschat. Van betere lezers met een grotere woordenschat mag je verwachten dat ze een tekst meestal beter begrijpen. Als die hoger scoren op een cloze-toets dan mensen met een geringere leesvaardigheid, dan mag je dat dus als ondersteuning zien voor de stelling dat deze toets wel degelijk begrip meet.

Verder lezen:

Kleijn, S., Pander Maat, H., & Sanders, T. (2019). Cloze testing for comprehension assessment: The HyTeC-cloze. *Language Testing*, 36(4), 553-572.

Kleijn, S. (2018). *Clozing in on readability: How linguistic features affect and predict text comprehension and on-line processing*. Utrecht: LOT.

Kamalski, J. (2007). *Coherence Marking, Comprehension and Persuasion on the processing and representations of discourse* (Doctoral dissertation, Netherlands Graduate School of Linguistics).

2. De vragenlijst: testen van begrip met een vragenlijst

Een vragenlijst met begripsvragen is de klassieke methode om begrip te testen. In het schoolexamen Nederlands wordt deze methode als de standaard gezien om tekstbegrip bij jongeren te testen. Maar ook op deze methode is kritiek. Zo bleek uit het onderzoek van Kamalski (2007) dat de scores van leerlingen op de begripstest nauwelijks samenhangen met andere methoden om begrip te testen en dat de test niet erg geschikt zou zijn om individuele verschillen tussen leerlingen aan het licht te brengen. Een ander kritiekpunt of nadeel van deze methode is dat we er uitsluitend problemen mee op het spoor komen die we zelf al hadden vermoed, anders hadden we er immers geen vraag over gesteld. Over een passage die niet bevraagd is, zullen we geen problemen opsporen. Maar dat is nu eenmaal een kenmerk van alle gesloten methoden: de regie ligt volledig bij de tester.

De begripsvragenlijst is misschien ook wel de moeilijkste methode, niet voor de proefpersoon maar voor de maker van de test. Hoe selecteer je de onderdelen die je wilt bevragen? En wat voor soort vragen stel je dan? Soms verradt de formulering van de vraag al welk antwoord het juiste is. Mag de proefpersoon de tekst erbij houden tijdens de beantwoording van de vragen? Hoe voorkom je dat de vraag zelf interpretatieproblemen creëert? Uit onderzoek naar biologie-examens blijkt bijvoorbeeld dat leerlingen met een geringe leesvaardigheid misschien best goed zijn in biologie maar soms de vraag verkeerd begrijpen en daardoor slecht scoren.

Ook bij deze methode moet de maker van de test vooraf een duidelijk idee hebben van het doel van het document en van de test. Als de tekst een krachtige argumentatie bevat tegen het leenstelsel en het doel van de test is begrip van tekststructuur vast te stellen, dan stel je vragen naar relaties tussen beweringen en typen argumentatie. Maar als de tekst het proces beschrijft waarin virusdeeltjes overgedragen worden en het immuunsysteem aantasten, is het doel misschien inzicht in dat proces te peilen. Dan zijn vragen naar het "mentale model" (Kamalski, 2007) meer gepast, waarbij de proefpersoon bijvoorbeeld vier of vijf kernbegrippen moet plaatsen in soort stroomschema dat het proces symboliseert.

Begrip van de tekst kan op verschillende niveaus gemeten worden. We maken onderscheid tussen drie niveaus en daarmee verschillende manieren van bevragen. Het meest elementaire niveau betreft het *onthouden* van de tekst. Dat meet je uiteraard met een methode waarbij de tekst wordt weggehaald voordat de vragenlijst wordt uitgedeeld. Reproductie kunnen we meten met bijvoorbeeld goed-fout stellingen.

De auteur beweert dat het leenstelsel de solidariteit tussen inkomensgroepen aantast.

Deze bewering is:

- *Goed*
- *Fout*

Maar dezelfde inhoud kan ook bevraagd worden in een vorm met verschillende beweringen in het veld met antwoorden.

Wat beweert de auteur over de relatie tussen het leenstelsel en de solidariteit tussen inkomensgroepen?

- *Hij beweert dat die relatie niet bestaat*
- *Hij beweert dat het leenstelsel de solidariteit bevordert*
- *Hij beweert dat het leenstelsel de solidariteit schaadt*

In deze voorbeelden gaan we ervan uit dat de tekst expliciet vermeldt wat het standpunt van de auteur is. Dan is er dus sprake van een begripsvraag die reproductie test. Maar wanneer de tekst die informatie niet expliciet formuleert, moet de proefpersoon het standpunt afleiden uit verschillende beweringen waartussen hij of zij een verband ziet. Dan vragen we niet naar de reproductie maar naar een *inferentie* en dat is het tweede niveau van tekstbegrip. Daarvoor kunnen we uiteraard precies dezelfde vraag gebruiken, maar nu testen we in hoeverre de proefpersoon samenhang ziet in de tekst.

Het derde niveau vraagt nog weer een stapje extra van de lezer: het *toepassen* van de informatie uit de tekst op een andere situatie of in een context die ruimer is dan de tekst biedt. Als de tekst vooral het perspectief van de studenten biedt, dan zou een vraag op dit niveau betrekking kunnen hebben op docenten of universiteitsbestuurders. Of op andere interventies die de solidariteit bevorderen, zoals hogere belastingen voor de rijken. Nu moet de proefpersoon de eigen kennis van de wereld verbinden met de informatie die de tekst aanbiedt. Met zulke vragen peilen we het *situatiemodel* van de lezer.

Wat zijn de gevolgen voor de financiering van de universiteiten als het leenstelsel wordt afgeschaft?

- *Die wordt verbeterd, omdat studenten dan sneller zullen studeren*
- *Die verslechtert, omdat de universiteiten dan de compensatiegelden van 1 miljard missen*
- *Die blijft hetzelfde, omdat het louter gaat om een regeling tussen overheid en student.*

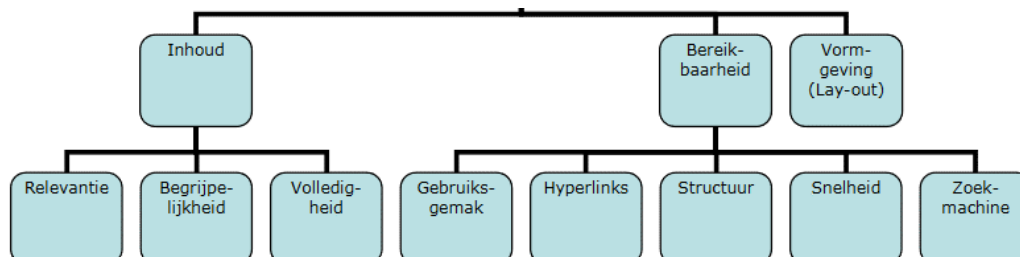
Dit derde niveau is het meest complex, zowel voor de proefpersoon als voor de maker van de test. De proefpersoon kan immers niet meer volstaan met het grondig lezen van de tekst en moet echt verder denken. De maker van de test moet zich extra goed verdiepen in allerlei mogelijke relaties en redeneringen en heeft ruime kennis van de context nodig om zeker te zijn van het juiste antwoord.

Verder lezen:

Kamalski, J. (2007). *Coherence Marking, Comprehension and Persuasion on the processing and representations of discourse* (Doctoral dissertation, Netherlands Graduate School of Linguistics).

3. De vragenlijst: testen van opinies met een vragenlijst

De vragenlijst kan ook gebruikt worden om oordelen en opinies te bevragen over een document of website. Voor het evalueren van websites met een (online) vragenlijst is in Utrecht de WEQ (Web Evaluation Questionnaire) ontwikkeld. Dit instrument is vooral voor overheidswebsites gemaakt. Er worden acht dimensies onderscheiden waarop de oordelen verzameld worden.



Als we willen weten wat de proefpersonen vonden van de inhoud, dan bevragen we hun idee over de relevantie van de geboden informatie, over de begrijpelijkheid en volledigheid van het document of de website. Hier zou een dimensie overtuigingskracht of aanvaardbaarheid aan toegevoegd kunnen worden voor websites die meer persuasief zijn.

In een digitale omgeving is de bruikbaarheid of *usability* sterk afhankelijkheid van de kwaliteit van de navigatie en die evalueren we door te vragen naar het gebruiksgemak, de duidelijkheid van de linklabels in het menu, de overzichtelijkheid van de structuur, de snelheid van de website en de kwaliteit van de zoekmachine (indien aanwezig). Tenslotte vragen we ook oordelen over de vormgeving.

Deze vragenlijst is uitgebreid getest en vergeleken met andere methoden. Over de ontwikkeling ervan vind je meer informatie in Elling (2007).

Er zijn talloos veel vragenlijsten ontwikkeld om oordelen te verzamelen over documenten en websites. Een handig overzicht plus bespreking van de voornaamste kenmerken vind je in Schrepp (2020).

Een vragenlijst kan natuurlijk ook allerlei open vragen bevatten, zoals *Wat vond je van deze tekst?* Dan kunnen proefpersonen zelf kiezen op welk aspect van tekstkwaliteit ze willen reageren. Daarmee komen we bij de tweede groep methoden, die meer open zijn en veelal kwalitatieve data genereren.

Verder lezen:

Elling, S., Lentz, L., & De Jong, M. (2007, September). Website evaluation questionnaire: development of a research-based tool for evaluating informational websites. In International Conference on Electronic Government (pp. 293-304). https://doi.org/10.1007/978-3-540-74444-3_25. Springer, Berlin, Heidelberg.

Schrepp, M., (2020). A Comparison of UX Questionnaires - What is their underlying concept of user experience?. In: Hansen, C., Nürnberger, A. & Preim, B. (Hrsg.), Mensch und Computer 2020 - Workshopband. Bonn: Gesellschaft für Informatik e.V.. DOI: [10.18420/muc2020-ws105-236](https://doi.org/10.18420/muc2020-ws105-236)

4. Het lezersprotocol

Het lezersprotocol wordt ook wel de hardop-leesmethode genoemd. De kern van deze methode is simpel: proefpersonen krijgen een tekst die zij hardop voorlezen. Tussentijds mogen zij het lezen onderbreken om opmerkingen over de tekst te maken. Alle gedachten die zij daarbij hebben zijn welkom. De proefleider maakt aantekeningen: 1) in een kopie van de tekst als er duidelijke haperingen hoorbaar zijn, want elke hapering is een indicatie van een begripsprobleem; 2) en op een apart vel worden alle opmerkingen samengevat, met een verwijzing naar het regelnummer in de kopie. Een audio-opname kan helpen om die notities nog eens goed te checken en misschien ook aan te vullen met nieuwe observaties.

Volgens Pander Maat en Lentz (2003) is dit een bijzonder goede methode om met name begripsproblemen in teksten op te sporen. Anders dan bij een vragenlijst, wordt de proefleider nu regelmatig verrast door lezersproblemen die absoluut niet verwacht werden. Dat is nu eenmaal de kracht van een open methode. Lentz en Pander Maat (2007) vergeleken het lezersprotocol in een onderzoek met twee andere methoden (plus-en-minmethode en een online tool voor website-evaluatie) en concludeerden dat deze methode meer begripsproblemen aan het licht brengt en dat de opbrengst ervan ook beter overeenkomt met een begripsvragenlijst: de correlatie tussen enerzijds het aantal foute antwoorden bij de vragenlijst en anderzijds het aantal probleemmeldingen bij de drie kwalitatieve methoden was het hoogst bij het lezersprotocol.

In een vervolgonderzoek gingen Pander Maat en Lentz na waarom die methode beter scoort. Is het omdat de proefpersonen de tekst hardop voorlezen? Nee, dat bleek niet het geval. Een variant waarbij de proefpersonen de tekst in stilte lezen, maar wel hardop commentaar gaven, liet een vergelijkbare opbrengst zien. De meest aannemelijke verklaring is dat de proefpersonen hun commentaren niet hoeven uit te stellen totdat ze de tekst uit hebben gelezen. Een tweede verklaring is dat met deze methode de proefpersonen geen passages kunnen overslaan. Zij worden dus min of meer gedwongen blootgesteld aan de volledige tekst en leveren zodoende ook commentaren op passages waar ze met andere methoden minder of geen aandacht aan zouden schenken.

Verder lezen:

Pander Maat, H., & Lentz, L. (2003). Waarom het lezersprotocol zo'n goede pretestmethode is [Readersprotocol is a good pretestmethod. Why?]. Tijdschrift voor taalbeheersing, 25, 202-220.

Lentz, L., & Maat, H. P. (2007). Reading aloud and the delay of feedback: Explanations for the effectiveness of reader protocols. *Information Design Journal*, 15(3), 266-281.

5. Hardop denken

Volgens Fan e.a. (2020) is het zogenaamde *Think-Aloud Protocol* in de praktijk de meest gehanteerde methode om problemen in teksten, tools, (digitale) formulieren of websites op het spoor te komen. Korter gezegd: het is de favoriet in *usability research*. De methode is vooral krachtig in het opsporen van problemen met de vindbaarheid van informatie, doordat navigatieproblemen goed zichtbaar worden. Maar ook problemen met het begrijpen en toepassen van de inhoud, met de aantrekkelijkheid of overtuigingskracht kunnen goed zichtbaar gemaakt worden.

Naar deze methode is bijzonder veel onderzoek gedaan, zowel in Nederland (Van den Haak, De Jong en Schellens, 2007), Elling, Lentz en De Jong (2012) als in Denemarken (Nørgaard en Hornbaek, 2006) en de VS (Cooke, 2010; Hertzum, 2009; McDonald, 2012). Een belangrijke vraag in dat onderzoek is of we de proefpersonen moeten vragen om direct commentaar te leveren (concurrent) of dat achteraf te doen (retrospectief). We zagen hierboven al dat Pander Maat en Lentz (2003) pleiten voor het direct leveren van commentaar, omdat dan de opbrengst hoger is. Een nadeel is weliswaar dat het verwoorden van feedback de proefpersoon afleidt van de taak waar hij mee bezig is, maar ook in de praktijk is volgens de survey van Fan e.a. (2020) de concurrent aanpak favoriet. Natuurlijk is die ook efficiënter, omdat er geen tweede sessie nodig is om feedback te verzamelen. Vooral in digitale contexten scheelt dat veel tijd, aangezien er bij retrospectief onderzoek meestal een opname wordt teruggespeeld waarin de proefpersoon zijn eigen navigatieproces ziet en becommentarieert.

Fan e.a. (2020) formuleren tips en richtlijnen voor de uitvoering van deze methode. Die vat ik hieronder samen, met hier en daar een persoonlijke aanvulling.

1. *Creëer een oefensessie*. Het is voor proefpersonen heel ongebruikelijk om tijdens het uitvoeren van een taak voortdurend hardop hun gedachten uit te spreken. Laat ze daar een keer mee oefenen met een andere taak en geef daar feedback op.
2. *Geef neutrale taakinstructie*. Proefpersonen laten zich makkelijk beïnvloeden in het soort feedback dat gewenst wordt. Vermijd daarom in de instructie zulke indicaties en spoor de proefpersonen aan *alles* wat ze bedenken uit te spreken.

3. *Beperk interactie*. Laat de proefpersonen zoveel mogelijk ongestoord werken. Ga uitsluitend in interactie als:
 1. Proefpersonen helemaal stilvallen, herhaal dan de oproep hardop te denken
 2. Proefpersonen volledig vastlopen in de taakuitvoering, voorkom dan frustratie en bied een oplossing
 3. Proefpersonen vastlopen bij een probleem dat je al eerder hebt gezien en goed hebt begrepen, voorkom dan ook frustratie en bied een oplossing.
4. *Analyseer breed*. Besteed in de analyse aandacht aan het *gedrag* van de proefpersonen, aan hun *verbalisaties* van gedachten, en aan de *gevoelens* die blijken uit de manier waarop ze hun commentaar verwoorden. Blijkt daar ergernis of opluchting uit? Noteer dat dan ook.
5. *Zoek een balans tussen efficiëntie en kwaliteit*. Deze methoden kosten meestal veel tijd, vooral wanneer voor de analyse het protocol ook volledig wordt uitgeschreven en de observaties zorgvuldig worden gecategoriseerd. Dat leidt tot een meer systematische analyse die minder afhankelijk is van de toevallige proefleider. Maar die werkwijze is ook kostbaarder. Minder proefpersonen is natuurlijk ook goedkoper dan een grotere groep deelnemers aan de test.
6. *Ontwikkel efficiënte analysesystemen*. Op basis van ervaring kun je een analyseschema ontwikkelen waarmee tijdens de observatie en zeker in de fase daarna problemen direct in *categorieën* herkend en genoteerd kunnen worden. Sommige tools stellen de proefleider in staat om opnames van de sessie direct te scoren in een analysesysteem, waardoor later een bepaalde score gemakkelijk opnieuw beoordeeld kan worden, omdat die verbonden is met een fragment uit de opname.

Er zijn verschillende tools verkrijgbaar waarmee je een opname kan maken van het gedrag van je proefpersonen, terwijl zij opdrachten uitvoeren op een website, in een game of een andere applicatie. Zie [hier](#) voor een overzicht van low budget tools.

Verder lezen:

Cooke, L. (2010). Assessing concurrent think-aloud protocol as a usability test method: A technical communication approach. *IEEE Transactions on Professional Communication*, 53(3), 202-215.

Elling, S., Lentz, L., & Jong, M.D. (2012). Combining Concurrent Think-Aloud Protocols and Eye-Tracking Observations: An Analysis of Verbalizations and Silences. *IEEE Transactions on Professional Communication*, 55, 206-220.

Fan, M., Shi, S., & Truong, K. N. (2020). Practices and Challenges of Using Think-Aloud Protocols in Industry: An International Survey. *Journal of Usability Studies*, 15(2).

Hertzum, M., Hansen, K. D., & Andersen, H. H. (2009). Scrutinising usability evaluation: does thinking aloud affect behaviour and mental workload? *Behaviour & Information Technology*, 28(2), 165-181.

McDonald, S., Edwards, H., & Zhao, T. (2012). Exploring Think-Alouds in Usability Testing: An International Survey. *IEEE Trans. Prof. Communication*, 55, 2-19. 10.1109/TPC.2011.2182569.

Nørgaard, M., & Hornbæk, K. (2006, June). What do usability evaluators do in practice? An explorative study of think-aloud testing. In *Proceedings of the 6th conference on Designing Interactive systems* (pp. 209-218).

Van den Haak, M. J., De Jong, M. D. T., & Schellens, P. J. (2007). Evaluation of an informational web site: Three variants of the think-aloud method compared. *Technical communication*, 54(1), 58-71.

6. Plus-en-minmethode

De plus-en-minmethode is in Nederland ontwikkeld en bekend gemaakt door Ben Vroom. Op zijn site geeft hij tips voor de gebruiksvriendelijkheid van teksten en de uitvoering van tests. De methode is vooral geschikt voor papieren documenten, waar lezers aantekeningen op kunnen maken, en voor het testen van de begrijpelijkheid, waardering en overtuigingskracht van die teksten.

De methode is heel eenvoudig en kent twee fasen. Na een instructie lezen proefpersonen in de eerste fase een tekst en zetten daarbij plussen en minnen in de kantlijn, waarbij zij aan allerlei mogelijke reacties mogen denken. In de tweede fase bespreekt de proefleider de plussen en minnen en vraagt om toelichting. Eigenlijk zijn die plussen en minnen niet veel meer dan een geheugensteun en katalysator om over de tekst concrete feedback te verzamelen. Juist in die toelichtingen komt het waardevolle commentaar naar voren. Tijdens de bespreking heeft de proefleider zelf een kopie van de tekst voor zich met genummerde regels. Op een apart notitievel noteert hij of zij via die regelnummers de kern van de commentaren bij elke passage. Voor elke proefpersoon komt een nieuw notatievel op tafel.

Vaak wordt deze methode gecombineerd met een interview, waarin de proefleider op specifieke punten feedback vraagt, zoals begrip van een bepaalde passage of de geloofwaardigheid van de auteur.

De Jong (1998) promoveerde op een onderzoek naar de waarde van deze methode door teksten op basis van de commentaren te reviseren en opnieuw te testen. Hij concludeert dat de feedback wel vaak tot verbetering leidde, hoewel dat niet bij elke tekst even duidelijk was. En soms kan een revisie ook weer nieuwe problemen veroorzaken.

Meer suggesties voor het werken met deze methode zijn hier te vinden.

Verder lezen:

De Jong, M. (1998). *Reader feedback in text design: Validity of the plus-minus method for the pretesting of public information brochures* (Vol. 13). Rodopi.

Vroom, B. (1994). *De tekst getest. Handleiding voor het pretesten van schriftelijk materiaal*. Van Gorkum, Assen.

7. Interview

In het interview voert de proefleider een gesprek met de proefpersoon, nadat die een tekst gelezen heeft of een website of tool bekeken heeft. Bij de plus-en-min-methode voert de proefleider in de tweede fase ook een interview uit, maar dan zijn de plussen en minnen bepalend voor de aard van het gesprek. Ik richt me nu op het gestructureerde interview met open vragen die vooraf zijn opgesteld. Deze methode laat zich goed combineren met andere methoden waarbij proefpersonen bijvoorbeeld eerst een taak uitvoeren of open commentaar geven op een document. In het interview komen vragen aan de orde die de proefleider beslist aan alle proefpersonen voor wil leggen, bijvoorbeeld omdat hij of zij een beter gefundeerd idee wil hebben van de mate waarin de proefpersonen de tekst aantrekkelijk of overtuigend vinden.

De proefleider heeft vooraf vastgesteld wat de doelen van de tekst zijn, waar mogelijke struikelblokken zijn en daaruit de voornaamste topics voor het interview afgeleid. Elk topic wordt met meerdere interviewvragen aan de orde gesteld, om te voorkomen dat de specifieke formulering van één vraag de test te veel een bepaalde kant op stuurt. De vragen worden vooraf uitgeschreven en in een paar proefinterviews uitgeprobeerd en zo nodig gereviseerd. Van groot belang is dat die vragen niet sturend zijn. Proefpersonen laten zich maar al te gemakkelijk leiden in een richting die door de vraag gesuggereerd wordt. Vergelijk:

- *Wat zou u met deze folder doen als u hem in de wachtkamer van de huisarts op een stapeltje zag liggen?*
- *Zou u deze folder mee naar huis nemen als u hem in de wachtkamer van de huisarts op een stapeltje zag liggen?*

De tweede vraag is sturend en gesloten; proefpersonen vermoeden dat *ja* het gewenste antwoord is en vrezen ook minder lastige vervolgvragen dan bij een ontkennend antwoord.

Wanneer de interviews face to face worden afgenomen, ontstaat de mogelijkheid om de verwerking van het document gefaseerd te laten verlopen. Eerst bekijkt de proefpersoon de folder heel kort, waarna enkele vragen volgen over de aantrekkelijkheid, relevantie en misschien het begrip van de kern van de folder. Is na een halve minuut duidelijk waar dit document over gaat? Daarna wordt de tekst rustig in zijn geheel gelezen, met vervolgens weer een reeks vragen. Tenslotte zou een specifieke passage aan het slot van het interview nog extra bevraagd kunnen worden.

Interviews kunnen natuurlijk ook telefonisch of online afgenomen worden met videobellen. Het grote nadeel bij die varianten is dat de proefleider en proefpersoon niet samen een document voor zich hebben. Praten over een tekst leidt vaak tot aanwijzen van plaatjes of passages, en dat is erg nuttig om te weten over welk onderdeel de proefpersoon feedback geeft. Om die reden zijn telefonische of online interviews vaak globaler van karakter, tenzij proefleider en proefpersoon een scherm delen, waarbij de proefpersoon met de cursor passages aan kan wijzen. Sinds de Corona-crisis zijn proefpersonen meer gewend geraakt aan het werken met tools, zoals Microsoft Teams, waarbij twee of meer personen zowel elkaar als een document kunnen zien.

Verder lezen:

Baarda, B. & Van der Hulst, M. (2017). Basisboek Interviewen. Handleiding voor het voorbereiden en afnemen van interviews. Noordhoff, Groningen.

8. Observatie van gedrag

Wanneer het document, de website of een tool specifiek gedrag vraagt, kan observatie van dat gedrag leerzaam zijn om problemen op te sporen. Dat kan natuurlijk navigatiegedrag zijn als het om een digitale omgeving gaat, zoals de website van een gemeente of de Belastingdienst. Maar het kan ook ander gedrag zijn als het een meer instructieve tekst is, waarbij proefpersonen fysieke handelingen moeten verrichten, of een formulier waar vragen moeten worden beantwoord. Observatie leert ons dan hoe zij de tekst gebruiken bij het uitvoeren van die handelingen, wanneer zij problemen ervaren en of het eindresultaat bevredigend is. Bij het testen van formulieren is observatie nuttig om te achterhalen of proefpersonen vragen overslaan, lang aarzelen bij het invullen van tekstvelden en non-verbale signalen geven van frustratie of onvrede.

In veel commerciële testsituaties, waarbij opdrachtgevers een testbureau inschakelen om een website of tool te laten testen, nodigen de proefleiders de opdrachtgever uit om mee te kijken. Dat gebeurt dan via een blind screen, zodat de proefpersoon er niks van merkt. Maar met tools als Microsoft Teams kan dat nu ook op afstand, waarbij proefleider en proefpersoon aan het werk zijn en de opdrachtgever op de achtergrond meekijkt. Het voordeel van die participatie is dat observatie van gebruiksproblemen een sterke indruk kan maken op de ontwerpers van het systeem. Ontwikkelaars die maanden aan een product gewerkt hebben, kunnen zich vaak niet voorstellen dat proefpersonen het menu niet begrijpen, instructies negeren, cruciale onderdelen overslaan en blijven verbijsterd achter nadat zij tien proefpersonen achter elkaar hebben geobserveerd.

In een digitale omgeving wordt het gedrag van de proefpersonen meestal opgenomen en later voor de analyse teruggespeeld. Daarvoor zijn verschillende technieken beschikbaar, afhankelijk van de testsituatie. Lonneke Theelen beschrijft een vergelijking tussen twee methoden waarbij proefpersonen een mobiele telefoon gebruiken tijdens de test. Net als bij hardop-denksessies is het dan belangrijk dat de proefleiders een goed systeem hebben om gedragingen te scoren. Duidelijk moet worden op welk moment, met welk onderdeel, welk specifiek probleem zich voordoet en waaruit dat blijkt.

Deze methode wordt vaak gebruikt in combinatie met hardop denken, een retrospectief interview en eye tracking.

Verder lezen:

Theelen, L. (2014). Mobile usability testing vergeleken: welke methode kies jij? URL: <https://www.frankwatching.com/archive/2014/08/04/mobile-usability-testing-vergeleken-welke-methode-kies-jij/>

9. Eye tracking

In professionele en academische testlaboratoria is het heel gebruikelijk om niet alleen het gedrag van de proefpersoon vast te leggen, maar ook de oogbewegingen. Dat gebeurt met een zogenaamde *eye tracker*, die ingebouwd is in de monitor en via infrarood-technologie registreert op welk punt van het scherm de ogen zich richten. Dat levert drie soorten data op:

- De volgorde of het patroon van de oogbewegingen, waardoor we dus zien welke onderdelen het eerst aandacht krijgen en welke later;
- De tijd die besteed wordt aan een bepaald onderdeel, wat een indicatie kan zijn van relevantie maar ook van een probleem, omdat proefpersonen lang moeten nadenken over een element;
- Wanneer we de data van meerdere proefpersonen combineren kan een zogenaamde *heat map* gemaakt worden die laat zien welke onderdelen door een groep proefpersonen het best bekeken zijn.

Natuurlijk zijn die oogbewegingen ook een vorm van gedrag, maar het verschil met andere handelingen is dat we veel minder controle hebben over onze manier van kijken. Een proefpersoon die een website voor het eerst bezoekt, neemt geen bewuste beslissing over de locatie waar de ogen het eerst naar toegaan, maar wordt automatisch aangetrokken door een bewegend plaatje of het linksboven geplaatste hoofdmenu van de site.

Hieronder zie je een schermafbeelding van de website van Apeldoorn die bezocht is door een proefpersoon in het onderzoek van Elling (2012), waarbij in één beeld alle fixatiepunten te zien zijn en het pad dat die punten verbindt. Hoe groter de stip, des te langer de fixatie op dat punt duurde. Je kunt hier goed zien dat de proefpersoon de pagina tamelijk goed bekijkt, maar het menu aan de bovenkant volkomen negeert. Zo kunnen we met deze techniek heel goed verklaren waarom sommige proefpersonen zoveel moeite hebben met het vinden van de informatie die zij zoeken.



In het onderzoek van Elling (2012) werd een combinatie gebruikt van eye tracking met hardop denken en observatie. Zij concludeert dat elk van die methoden, juist in combinatie met de andere, zinvolle feedback oplevert. Dat komt onder andere omdat proefpersonen tijdens het hardop denken regelmatig stilvallen. Elling laat zien dat zij dan meestal een pagina aan het scannen zijn of een stuk tekst lezen. Die twee activiteiten worden meestal in stilte uitgevoerd. Het heeft niet zoveel zin om proefpersonen op die momenten aan te sporen hardop te denken.

De combinatie van oogbewegingen en commentaren kunnen relevante problemen aan het licht brengen. Zo zegt een proefpersoon na het langdurig scannen van een pagina: *ik denk dat ik maar terug ga naar de homepage*. Vervolgens klikt zij op een link die niet naar de homepage leidt. Zonder verbalisatie was niet duidelijk geworden wat haar intentie was. En zonder observatie was niet duidelijk geworden dat zij een bepaalde link ten onrechte interpreteert als een link naar de homepage. Tenslotte laten de oogbewegingsdata zien of zij überhaupt de homelink heeft waargenomen. Dat is relevant, want als ze hem niet gezien heeft zou het design aangepast moeten worden. Maar als ze hem wel gezien heeft, is de titel van de link misschien verwarrend. Uiteraard krijgen zulke interpretaties van de data pas kracht wanneer die problemen bij meerdere proefpersonen worden waargenomen.

In een ander deel van haar onderzoek ging Elling na of het zin heeft proefpersonen zelf een filmpje te laten zien van hun eigen zoekproces inclusief de registratie van oogbewegingen. Zij concludeert dat dit niet tot rijkere feedback leidt. Proefpersonen vinden het heel interessant om zo'n filmpje te zien, maar verwoorden dan geen extra problemen. Die conclusie werd later bevestigd in een studie van Alhadreti e.a. (2017).

Verder lezen:

Alhadreti, O., Elbabour, F., & Mayhew, P. (2017). Eye tracking in retrospective think-aloud usability testing: is there added value?" *Journal of Usability Studies*, 12(3), 95-110.

Bergstrom, J. R., & Schall, A. (2014). *Eye tracking in user experience design*. Elsevier.

Elling, S., Lentz, L., & De Jong, M. (2012). Combining concurrent think-aloud protocols and eye-tracking observations: An analysis of verbalizations and silences. *IEEE transactions on professional communication*, 55(3), 206-220.

10. Focusgroepen

In een focusgroep bespreekt de proefleider met maximaal zeven proefpersonen een document of ander product van communicatie. Het doel is feedback te verzamelen over de zwakke en sterke kanten van dat product. Zo'n focusgroep heeft twee voordelen boven andere methoden. Het eerste voordeel is praktisch: deze methode is efficiënter omdat je in plaats van zeven individuele interviews nu in één groepsgesprek feedback verzamelt. Het tweede voordeel is inhoudelijk: als het gesprek goed geleid wordt en de deelnemers constructief meewerken kan de interactie leiden tot een dieper inzicht in mogelijke problemen en oplossingen. Sommigen raden om die reden aan de focusgroep homogeen samen te stellen qua opleidingsniveau of professionele achtergrond en vervolgens in een reeks gesprekken verschillende groepen aan het woord te laten. De methode is zeer geschikt om in vrij korte tijd een indicatie te krijgen van de begrijpelijkheid, aantrekkelijkheid en overtuigingskracht van een document.

Een potentieel gevaar van deze methode is dat proefpersonen elkaar gaan beïnvloeden of dat sommige proefpersonen anderen overvleugelen. Dat gevaar wordt kleiner als de proefleider de volgende richtlijnen hanteert:

- Geef de proefpersonen de tijd om eerst rustig naar het document te kijken en het te lezen.
- Vraag elke proefpersoon vervolgens eerst een paar punten (misschien plussen en minnen) op te schrijven.
- Maak soms een rondje, waarbij elke proefpersoon kort een reactie geeft.

Belangrijk is ook dat de proefleider vooraf de proefpersonen op hun gemak stelt. Het kan voor sommigen intimiderend zijn om in een groep opinies of ervaringen te moeten verwoorden. Een informele omgeving kan helpen, dus liever een wijkgebouw dan het gemeentehuis als locatie. En even koffie drinken vooraf en informeel kennis maken kan ook helpen om de proefpersonen op hun gemak te stellen.

Verder gelden voor de focusgroep dezelfde richtlijnen als voor het interview. De proefleider stelt dus nooit sturende vragen, geeft zelf geen opinies, vraagt door bij vage antwoorden, maakt duidelijk dat een antwoord nooit fout kan zijn, en zorgt dat er niet afgedwaald wordt. We noemen dit niet voor niets een Focusgroep. In Smithson (2020) worden de sterke en minder sterke kanten van deze methode uitgebreid besproken.

Verder lezen:

Smithson, J (2000). Using and analysing focus groups: Limitations and possibilities. *International Journal of Social Research Methodology*, 3(2), 103-119. [DOI](#).

Lees hier de andere bijdragen over onderzoeksmethoden bij lezersonderzoek:

[Lezersonderzoek \(deel 1\): inleiding](#)

[Lezersonderzoek \(deel 3\): onderzoek in stappen](#)

Graag als volgt naar deze bijdrage verwijzen: Lentz, L. (2021). *Methoden. Didactiek Nederlands - Handboek*. Geraadpleegd [datum] via [<https://didactieknederlands.nl/handboek/2021/01/lezersonderzoek-deel-2-methoden/>].

Auteurs:



Leo Lentz

+ posts

Leo Lentz is hoogleraar aan de Universiteit Utrecht op het terrein van taal en communicatie. Hij doet onderzoek naar de begrijpelijkheid van teksten en de gebruiksvriendelijkheid van digitale documenten. Dat zijn meestal alledaagse teksten, zoals bijsluiters bij medicijnen, kookrecepten of formulieren en websites; maar hij doet ook onderzoek naar minder alledaagse documenten die te maken hebben met financiën: hypotheekaktes, pensioendocumenten, aanmaningen en verzekeringen.

Delen:



Gerelateerd:

- Lezersonderzoek (deel 3): onderzoek in stappen
- Lezersonderzoek (deel 1): inleiding
- Schrijfprocessen
- Begrijpend lezen, drie factoren die deze cognitieve processen beïnvloeden (deel 3)