

Gaat de AD Haringtest om meer dan de haring? Een update

Ben Vollaard, Universiteit Tilburg, november 2017¹

Samenvatting

Een hoge plaats in de AD haringtest hangt sterk samen met het inkopen van haring bij een leverancier waar het testpanel zakelijke belangen in heeft. Op een schaal van 0 tot 10 ligt het eindcijfer voor vishandels met deze leverancier gemiddeld 3,6 punten hoger dan voor vishandels met een andere leverancier. Dit blijkt uit een analyse van de resultaten van de AD haringtest van dit en vorig jaar, gekoppeld aan gegevens over leveranciers van de geteste haring.

Bijna de helft van het verschil in oordeel is te herleiden tot een subjectieve inschatting door het panel van de kwaliteit van schoonmaken van de haring (op een schaal van zeer goed, goed, matig tot slecht) en de mate van rijping van de haring (op een schaal van licht, gemiddeld, sterk tot bedorven). De 1,5 extra punten die dit gemiddeld genomen oplevert voor vishandels met de bewuste leverancier, betekent 25 tot 30 plaatsen hoger eindigen in de ranking van de haringtest.

Van minder belang voor het verklaren van het verschil in oordeel zijn de temperatuur van de geserveerde haring en of de haring vers van het mes is of niet. Onbelangrijk blijken het gewicht, het vetpercentage en de microbiologische gesteldheid van de haring – alle objectief meetbare factoren.

Aanleiding

Het Algemeen Dagblad (AD) test elk jaar de kwaliteit van Hollandse Nieuwe in een aantal verkooppunten. Een kleine 10 procent van de verkooppunten wordt hiertoe bezocht², voornamelijk in het westen van het land.³ Deze zogenoemde AD Haringtest kan nuttig zijn voor consumenten. Als de test meer is dan alleen een momentopname, dan weten consumenten immers waar zij goede haring kunnen verwachten. Ook brengt de test wettelijke eisen voor de verkoop van haring onder de aandacht, waaronder een maximum verkooptemperatuur van vis.

Tegelijk brengt de haringtest voor het AD verantwoordelijkheid met zich mee. Zowel lezers als ondernemers mogen verwachten dat de test eerlijk is. De beoordeling trekt veel aandacht, anders zou die immers ook

¹ Ben Vollaard is als hoofddocent verbonden aan de vakgroep economie van de Universiteit Tilburg. De auteur heeft voor dit onderzoek geen gelden ontvangen van een betrokken partij. Dit onderzoek is een vervolg op een in juli jl. verschenen analyse van de resultaten van de AD haringtest – vandaar het woord ‘update’ in de titel. Uit dat onderzoek bleek dat vishandels in Rotterdam en omstreken hoger scoren op de haringtest dan uit de kwaliteit van de geserveerde haring is te verklaren. Zie <https://www.tilburguniversity.edu/nl/actueel/nieuws/persbericht-haring-haringtest-Vollaard-vis/>

² Het gaat om iets minder dan 150 bezochte verkooppunten tegen een totaal van ongeveer 1.800; zie NRC Handelsblad <https://www.nrc.nl/nieuws/2015/07/21/jongens-ik-gooi-ze-weg-eeen-nul-1517301-a584524>.

³ Van alle in 2016 en 2017 bezochte verkooppunten ligt driekwart binnen 100 kilometer van Rotterdam.

niet interessant zijn voor de krant, en heeft naar verluidt grote invloed op het aantal klanten dat een verkooppunt bezoekt. Consumenten baseren er dus hun oordeel op en ondernemers kunnen er flink voordeel van hebben of nadeel van ondervinden.

De haringtest is omstreden. In de discussies komen de volgende kritiekpunten op de test vaak naar voren:

- *Smaken verschillen.* Allereerst zou volgens sommigen de beoordeling niet veel zeggen, omdat smaken nu eenmaal verschillen. Zo zou een typische Amsterdammer een andere haring lekker vinden dan een typische Rotterdammer.⁴ Het testteam zou bijvoorbeeld vooral heel koude haring willen: de top drie van 2017 serveerde haring van 3 graden celcius of minder.⁵ Daarmee is de test subjectief en weinig informatief. Daar valt tegenin te brengen dat de test méér is dan alleen een smaaktest. Zo laat het team ook de microbiologische gesteldheid van de haring bepalen.
- *Belangenconflict.* Een ernstiger klacht is dat het oordeel van het AD testteam vertekend is door zakelijke belangen.⁶ Het team beoordeelt namelijk ook ondernemers die een cursus hebben gevolgd bij één van de panelleden.⁷ Het gaat om panellid Aad Taal die vanuit zeevisleverancier Atlantic de cursussen 'Engels haring snijden' en 'Combinatiecursus Kennis haring B' geeft.⁸ Op de webpagina van Atlantic waar de cursussen worden aangeprezen staat te lezen: "Desgewenst kunnen [met panellid Aad Taal] de resultaten van de AD Haringtest worden besproken." De verdenking is dat verkooppunten die voor Atlantic als leverancier kiezen en een cursus bij Atlantic volgen daarmee een hoge score op de test veiligstellen. Vooral vishandels die een zware onvoldoende scoren op de haringtest zou dit in het vooruitzicht worden gesteld. Er zijn verschillende voorbeelden van vishandels die eerst laag scoorden en na overstap naar Atlantic als leverancier opeens een 9 of een 10 op de test behalen.⁹ Dit puntenvoordeel zou een vorm van klantenbinding zijn. Opvallend is dat zeven vishandels in de top 10 van de haringtest 2016 Atlantic als leverancier hebben; in 2017 gold dat voor vijf vishandels. Het aandeel viszaken met Atlantic als leverancier binnen de haringtest ligt lager dan 10 procent; deze zaken zijn dus zwaar oververtegenwoordigd in de top 10 van de haringtest.

⁴ Zo zegt een Haagse ondernemer in reactie op de haringtest: "Dit vind ik een lekkere haring die ik in mijn winkel kan verkopen. Maar dezelfde haring, als ik die in Rotterdam ga verkopen, dan zeggen ze: joh, die is niet te vreten, daar houden ze niet van. De ene wil een harde, de ander wil een zachte haring. Die wil een zoute haring, die wil een beetje flauwe haring. Er zijn natuurlijk zoveel verschillende soorten daarin." Bron is een uitzending op Den Haag FM: <http://denhaagfm.nl/2017/07/24/sprake-van-belangenverstrengeling-bij-ad-haringtest/>

⁵ Zie verder ook een recent blog op het webforum Jongebazen.nl <http://www.jongebazen.nl/trage-stromen/binnenkort-ad-haringtest-is-wel-betrouwbaar>.

⁶ Zie het hierboven aangehaalde artikel uit NRC Handelsblad, de eerder genoemde uitzending op Den Haag FM en een veelgelezen blog op Jongebazen.nl <http://www.jongebazen.nl/lifestyle-kunst/haringtest-ad-daar-zit-luchtje-aan>.

⁷ Zie het hierboven aangehaalde artikel uit NRC Handelsblad. Daarin staat ook vermeld dat AD vishandels die zijn getest zou benaderen of ze een advertentie in de krant willen plaatsen.

⁸ Zie de website van Atlantic: <http://vishandelatlantic.nl/opleidingen/>

⁹ Eén voorbeeld is de firma Simonis in Scheveningen. In 2009 scoorde Simonis nog een 0, in 2010 een 9,5 (derde plaats) en in 2011 een 10- (eerste plaats). "Een nieuwe haringleverancier speelde daarbij een belangrijke rol", was in het AD te lezen: <https://www.ad.nl/economie/ad-haringtest-veel-geklungel-en-gratenbakken~a2936efb/> "Daarna is het personeel bijgeschoold en de inkoop aangescherpt", is te lezen op <http://www.denhaagdirect.nl/dhd-haringtest-wat-is-er-mis-met-haring-van-vishandel-bezuidenhout/>. De nieuwe leverancier is Atlantic.

- *Testscore als dreigmiddel.* Een andere klacht is dat vishandels die kritiek hebben op de haringtest met opzet slechte cijfers krijgen als een soort straf.¹⁰ En ondernemers die uit protest het testteam de deur wijzen, worden over de hekel gehaald in het AD.¹¹ Op deze manier zou het testteam ondernemers willen afschrikken niet met hun kritiek naar buiten te treden.
- *Vertekening door sociale relaties tussen panelleden en ondernemers.* Tenslotte zou de test niet eerlijk zijn, omdat de wijze van testen dat bij voorbaat al in de weg staat. Beoordelingen zijn vaak vertekend als experts de te beoordelen personen (in dit geval de verkopers) kunnen zien, laat staan hen persoonlijk kennen, zo is bekend uit wetenschappelijk onderzoek.¹² Dit zou kunnen verklaren waarom vishandels in de regio Rotterdam/Den Haag (inclusief Scheveningen) – los van de kwaliteitskenmerken zoals door het testteam genoteerd en het lab gerapporteerd – een streepje voor hebben, omdat het team uit die regio komt of althans daar werkzaam is.¹³ Inderdaad blijkt dat als we het testcijfer voorspellen aan de hand van de door het testpanel gegeven kwaliteitskenmerken van de haring, zaken binnen de regio Rotterdam niet langer dit voordeel genieten en ook niet langer oververtegenwoordigd zijn in de top 10 van de test.

Dit onderzoek richt zich exclusief op het bovengenoemde belangenconflict. Centraal staat de volgende vraag: *heeft het belangenconflict binnen het testpanel invloed op hun oordeel?* Alvorens op de aanpak in te gaan, geven we hieronder eerst kort aan hoe de haringtest volgens de makers ervan tot stand komt.

Hoe werkt de haringtest?

Het team van de AD haringtest bestaat uit drie personen: twee mensen uit de viswereld, de ‘keurmeesters’, en een redacteur die de test aan elkaar schrijft. Het team bezoekt een kleine 150 verkooppunten in de start van het seizoen van de verkoop van eerste, jonge haring. Het kan gaan om een klassieke vishandel, maar ook om een viskraam of een supermarkt. Dit kost hen zo ruim twee weken. Rond de 90 procent van de verkooppunten bezoeken zij niet.

Bij elk verkooppunt worden vier haringen gekocht. Van twee van deze haringen meten zij direct na aankoop de temperatuur. Ook beoordeelt het team deze twee haringen op de kwaliteit van het schoonmaken, geur, uiterlijk buitenzijde, uiterlijk binnenzijde, zoutgehalte, mate van rijping en smaak. Tezamen resulteert dit in een voorlopig cijfer. De twee andere haringen worden in een steriele zak verpakt, gewogen en ingevroren, en doorgestuurd naar het laboratorium. Daar wordt het vetgehalte gemeten en de microbiologische gesteldheid van de haringen bepaald. Het cijfer wordt hierop aangepast. De verkooppunten die in de top 10 eindigen worden een tweede maal bezocht. Ook hier wordt het cijfer op aangepast. De uiteindelijke ranking is bepaald op basis van het eindcijfer, waarin al deze zaken zijn verwerkt.

¹⁰ Zie het eerder genoemde blog op het webforum Jongebazen.

¹¹ Zie het artikel ‘Kwaliteit is geen toeval’ van Paul Hovius in Algemeen Dagblad van 22 juli jl.

¹² Claudia Goldin & Cecilia Rouse, 2000, Orchestrating impartiality: the impact of ‘blind’ auditions on female musicians, *American Economic Review*, 90 (4), 715-741; Danielle Li, 2017, Expertise versus bias in evaluation: evidence from the NIH, *American Economic Journal: Applied Economics*, 9 (2), 60-92.

¹³ Zie <https://www.tilburguniversity.edu/nl/actueel/nieuws/persbericht-haring-haringtest-Vollaard-vis/>

Aanpak

De voor dit onderzoek gebruikte gegevens zijn afkomstig van de website van het AD (ad.nl/haringtest), aangevuld met gegevens van leveranciers van de betrokken verkooppunten. De website van het AD geeft voor 2016 en 2017 per bezocht verkooppunt alle kenmerken weer die het testteam in het oordeel meeneemt. We hebben leveranciers benaderd met de vraag aan te geven welke vishandels die mee zijn genomen in de haringtest zij bevoorraden. Alle benaderde leveranciers hebben opening van zaken gegeven, met één uitzondering: groothandel Atlantic. Een substantieel aantal vishandels is daarom direct benaderd om te vragen welke leverancier zij hebben. Op een enkele vishandel na hebben we de leverancier op deze manier kunnen achterhalen.

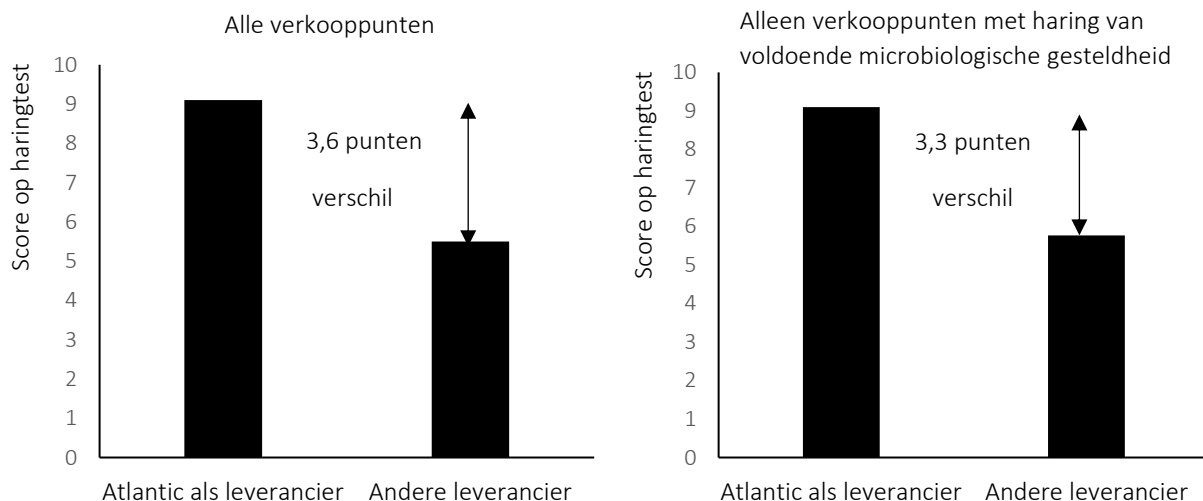
Wanneer een vishandel Atlantic als leverancier heeft, dan hoeft dat niet noodzakelijkerwijs te betekenen dat ook een cursus is gevolgd bij panellid Taal. Naar verluidt gaat dit in de praktijk vaak samen. Daarom richten we ons in deze analyse op de leverancier. Mocht een leveranciersrelatie niet altijd samengaan met het inkopen van een cursus bij Atlantic, dan bestaat er minder reden voor een vertekening in de resultaten.

In het hiernavolgende brengen we eerst het verschil in eindcijfer in beeld, daarna onderzoeken we waar het verschil in eindcijfer vandaan komt (een zogenaamde decompositie-analyse).

Verskil in eindcijfer

Allereerst kijken we hoe het eindcijfer op de haringtest verschilt tussen verkooppunten mét en zónder Atlantic als leverancier. De linker grafiek in Figuur 1 laat zien dat het verschil tussen beide groepen in gemiddelde score op de haringtest 2016 en 2017 3,6 punten bedraagt op een schaal van 0 tot 10. Verkooppunten met Atlantic als leverancier scoren gemiddeld een 9,1. De andere verkooppunten scoren een onvoldoende, gemiddeld net geen 5,5. We zien weinig verschil in gemiddelde score tussen de overige leveranciers.

Figuur 1. *Gemiddelde score van verkooppunten op de haringtest 2016/17, naar leverancier*



Dit verschil in eindcijfer kan versterkt zijn als in de groep die niet Atlantic als leverancier heeft relatief veel rotte appels zitten; zaken die echt haring verkopen die niet verkocht zou moeten worden. Deze rotte appels zouden het gemiddelde sterk omlaag kunnen trekken. Om dit te onderzoeken gebruiken we de microbiologische gesteldheid van de haring zoals vastgesteld door het laboratorium en vermeld door het

panel. Als die onvoldoende is, dan hoort die vis niet verkocht te worden. Als we er vanuit gaan dat de labresultaten correct zijn weergegeven op de website van het AD, dan kunnen we aan de hand daarvan die verkooppunten apart zetten die door de slechte microbiologische gesteldheid van de haring door de mand vallen.

Als we alle 37 van de 292 verkooppunten met een onvoldoende microbiologische gesteldheid van de haring buiten beschouwing laten, dan blijkt dat het beeld nauwelijks verandert. Dit laat de rechtergrafiek in Figuur 1 zien. Verkooppunten met Atlantic als leverancier behalen nog altijd gemiddeld een score van 9,1; andere verkooppunten gemiddeld een 5,8. Het verschil is nu 3,3 punten en blijft statistisch significant.

Conclusie. Verkooppunten met Atlantic als leverancier krijgen gemiddeld genomen een aanzienlijk beter cijfer van het AD testteam dan verkooppunten die een andere leverancier hebben. Het verschil in gemiddeld eindcijfer is 3,6 punten op een schaal van 0 tot 10. Het grote verschil in score is niet te wijten aan een aantal rotte appels in de groep van verkooppunten die niet Atlantic als leverancier heeft.

Het verschil in eindcijfer verklaard

Waar komt grote verschil in het oordeel naar leverancier vandaan? Om deze vraag te beantwoorden reconstrueren we hoe het eindcijfer tot stand is gekomen. Dit doen we aan de hand van een verklaringsmodel. In dit model relateren we de behaalde eindcijfers aan alle factoren die het oordeel mogelijkkerwijs beïnvloeden en schatten de bijdrage van elke factor.

Hoe goed werkt het verklaringsmodel voor het eindcijfer op de haringtest?

Het model voorspelt een verschil in het gemiddelde eindcijfer tussen zaken met en zaken zonder Atlantic als leverancier dat even groot is als het daadwerkelijke verschil. Dit is een belangrijk gegeven, omdat dit onderzoek zich hierop richt. Meer algemeen geldt dat het met het model voorspelde eindcijfer en het daadwerkelijke eindcijfer dicht bij elkaar liggen. De voorspelde ranking van verkooppunten op basis van het eindcijfer van de haringtest en de daadwerkelijke ranking komen zeer sterk overeen (correlatiecoëfficiënt van 0.92).

Factoren die in het verklaringsmodel een rol spelen zijn allereerst de kwaliteitskenmerken die het team noteert dan wel het laboratorium rapporteert: gewicht, temperatuur, vetpercentage, vers van het mes of niet, microbiologische gesteldheid, mate van rijping en kwaliteit van schoonmaken. Daar voegen we de invloed van de vestigingsplaats aan toe. Uit eerder onderzoek weten we dat verkooppunten in Rotterdam en omstreken hoger scoren op de haringtest dan uit de kwaliteit van de geserveerde haring is te verklaren.¹⁴ Ook voegen we toe of een verkooppunt in de top 10 eindigt. Deze zaken worden namelijk twee maal bezocht (zie het tekstkader 'Hoe werkt de haringtest?'), wat van invloed kan zijn op het oordeel. We gaan ervan uit dat de invloed van ieder van deze factoren identiek is voor zaken met en zonder Atlantic als leverancier. Het model en de schattingsresultaten staan in de appendix weergegeven.

In Tabel 1 geven we een overzicht van de invloed van de verschillende factoren op het verschil in oordeel tussen verkooppunten met en zonder Atlantic als leverancier. We kijken hierbij naar de mate waarin beide

¹⁴ Zoals eerder gezegd kan hier de invloed van relatief innige sociale relaties tussen het testteam en ondernemers binnen deze regio achter zitten. Zie voetnoot 1.

groepen verschillen op elke verklaringsfactor – de tweede kolom – en berekenen dan wat de invloed hiervan is op het eindcijfer – de vierde kolom – gegeven de resultaten van het verklaringsmodel.

Een voorbeeld om Tabel 1 te verduidelijken. Het gewicht van de haring bij verkooppunten die niet Atlantic als leverancier hebben ligt 2,17 gram lager vergeleken met verkooppunten die wel Atlantic als leverancier hebben (kolom 2). Elke gram minder kost 0,04 punten op het eindcijfer, zo weten we uit de appendix. Dit levert een verschil in eindcijfer op van 0,09 punten (kolom 4). Op een schaal van 0 tot 10 is dat minimaal.

Tabel 1. *Waarom verkooppunten die niet Atlantic als leverancier hebben lager scoren op de haringtest*

Verklaringsfactor	Verskil tussen verkooppunten met en zonder Atlantic als leverancier	Eenheid	Effect op eindcijfer (0-10)
Gewicht	-2.17	gram	-0.09
Microbiologische gesteldheid	referentiecategorie		
(zeer) goed	0.07	aandeel	-0.01
Voldoende	0.03	aandeel	-0.02
Slecht	0.09	aandeel	-0.01
waarschuwingsfase	0.02	aandeel	-0.04
Afgekeurd			-0.07
Gezamenlijk			
Vetpercentage	referentiecategorie		
beneden 10%	-0.21	aandeel	-0.04
tussen de 10 en 14%	-0.02	aandeel	-0.01
boven 14%			-0.05
Gezamenlijk			
Temperatuur	referentiecategorie		
beneden 7°C	0.29	aandeel	-0.17
tussen 7 en 10°C	0.21	aandeel	-0.36
boven 10°C			-0.52
Gezamenlijk			
Vers van het mes	referentiecategorie		
niet	-0.29	aandeel	-0.51
wel			
Schoonmaken	referentiecategorie		
zeer goed	0.27	aandeel	-0.22
goed	0.25	aandeel	-0.40
matig	0.04	aandeel	-0.12
slecht			-0.73
gezamenlijk			
Rijping	referentiecategorie		
licht	-0.24	aandeel	0.09
gemiddeld	0.34	aandeel	-0.63
sterk	0.06	aandeel	-0.25
bedorven			-0.80
gezamenlijk			
Regio	referentiecategorie		
binnen straal van 30 km rond R'dam	0.60	aandeel	-0.18
meer dan 30 km van R'dam			
Top 10	referentiecategorie		
top 10	0.57	aandeel	-0.54
buiten top 10			

Tabel 1 maakt duidelijk dat we het verschil in eindcijfer tussen verkooppunten met en zonder Atlantic als leverancier niet hoeven te zoeken in het gewicht, de microbiologische gesteldheid en het vetpercentage van de haring. De verschillen zijn klein en de invloed op het eindcijfer marginaal. Dit zijn allemaal objectieve kenmerken van de haring: het vetpercentage en de microbiologische gesteldheid worden door het lab bepaald; de weegschaal geeft het gewicht aan. Ook het feit dat relatief veel zaken die Atlantic als leverancier hebben gevestigd zijn in de regio Rotterdam verklaart weinig van het verschil in oordeel.

Bijna de helft van het verschil zit in een subjective beoordeling van de kwaliteit van schoonmaken van de haring en de mate van rijping van de haring. De plaatsing in de categorieën als 'zeer goed', 'goed', etc. is duidelijk wat anders dan het aflezen van de temperatuur op een thermometer.

De beoordeling van de kwaliteit van schoonmaken zou vertekend kunnen zijn vanwege het feit dat dit een belangrijk onderdeel is van de cursus die één van de leden van het testpanel vanuit Atlantic geeft. Zoals gezegd is vanuit wetenschappelijk onderzoek bekend dat een vertekening in het oordeel waarschijnlijker is naar mate de beoordelaar en beoordeelde elkaar beter kennen. Daarnaast bestaat er natuurlijk de belangenverstrengeling. Hierdoor kan bij zaken met Atlantic als leverancier bijvoorbeeld een verschuiving van het oordeel over de kwaliteit van schoonmaken plaatsvinden van 'goed' naar 'zeer goed'. Zo zien we dat binnen de groep vishandels die goed of zeer goed schoongemaakte haring serveren (ruim 70 procent van alle bezochte zaken) dat vishandels met Atlantic als leverancier bijna uitsluitend in de categorie 'zeer goed' vallen terwijl de verdeling tussen 'goed' en 'zeer goed' bij andere vishandels ongeveer 50/50 is. Voor een direct bewijs dat hier inderdaad sprake is van een vertekening zou het oordeel van het testpanel moeten worden vergeleken met een oordeel van een team die dezelfde haringen beoordeeld zonder te weten waar die vandaan komen.¹⁵

Hetzelfde gaat op voor het oordeel over de mate van rijping. Zo is de verdeling binnen de categorieën 'gemiddeld'/'sterk'/'bedorven' sterk afwijkend tussen beide groepen. Voor zaken met Atlantic als leverancier is de (procentuele) verdeling binnen deze groep 100/0/0; voor de andere zaken 60/35/5. Opvallend is dat 60 procent van de haring die volgens het team bedorven is een zeer goede microbiologische gesteldheid heeft. Dit is zo opvallend omdat het laboratoriumonderzoek onder meer de bacteriën meet die op de haring voorkomen, wat volgens het testteam "een indicatie van de mate van bederf van het product" geeft.¹⁶

Twee andere factoren die ook bijdragen aan het verklaren van het verschil in oordeel zijn de temperatuur van de geserveerde haring (verklaart een halve punt van het verschil in het gemiddelde eindcijfer) en of de haring vers van het mes is (verklaart ook een halve punt). Een soortgelijke invloed heeft de relatief grote vertegenwoordiging van vishandels met Atlantic als leverancier in de top 10.

¹⁵ Voor een aanpak waarbij de keurmeesters niet weten waar de haring vandaan komt, zie de Leidsch Dagblad Haringsmaakttest: <http://sleutelstad.nl/2017/06/24/stevenshof-visspecialist-wint-leidsch-dagblad-haringsmaakttest/> Deze test richt zich puur op smaak. Om een eerlijke aanpak te garanderen hield De Clercq Advocaten en Notarissen toezicht op het correcte verloop van deze test.

¹⁶ Zie <https://www.ad.nl/economie/reacties-haringtest-van-teleurstellend-tot-supertrots~a034b6ea/>

Conclusie. Het grote verschil in gemiddeld eindcijfer tussen verkooppunten met en zonder Atlantic als leverancier is voor bijna de helft te verklaren uit een subjectieve inschatting door het testteam hoe goed de haring is schoongemaakt (zeer goed/goed/matig/slecht) en van de mate van rijping van de haring (licht/gemiddeld/sterk/bedorven). In mindere mate zijn de gemeten temperatuur van de geserveerde haring, of de haring vers van het mes is, en plaatsing in de top 10 van belang voor het verklaren van het verschil in oordeel. Onbelangrijk blijken verschillen in het gewicht, de microbiologische gesteldheid en het vetpercentage van de haring – alle objectief meetbare factoren.

Conclusie

De AD haringtest is zo opgezet dat het testteam weet welke ondernemer de te beoordelen haring verkoopt. Dit kan tot vertekening van het oordeel leiden. Eerder hebben we laten zien dat verkooppunten in dezelfde regio als waar het testteam vandaan komt of werkt een puntenvoordeel genieten dat niet is te herleiden tot de kenmerken van de haring zoals door het team genoteerd of door het laboratorium gerapporteerd (zie voetnoot 1). Met andere woorden: het cijfer op de haringtest hangt niet alleen af van de kwaliteit van de geserveerde haring, maar ook van de vestigingsplaats van de zaak die de haring serveert.

Dit onderzoek richt zich op een belangenconflict van één van de leden van het testpanel. Een panellid is namelijk zakelijk verbonden aan een bepaalde leverancier van haring. Dit weerhoudt het panel er niet van verkooppunten die hun haring bij deze leverancier betrekken te beoordelen.

Het eindcijfer blijkt inderdaad sterk in het voordeel uit te vallen van vishandels met de bewuste leverancier. Op een schaal van 0 tot 10 scoren deze zaken gemiddeld een 9,1, tegen een krappe 5,5 voor alle andere zaken – een verschil van 3,6 punten. Het is dan ook niet verrassend dat verkooppunten met deze leverancier zwaar zijn oververtegenwoordigd in de top 10 van de haringtest.

Vishandels met de bewuste leverancier serveren volgens het testteam gemiddeld genomen dus bijzonder veel betere haring dan de andere bezochte zaken. Een alternatieve verklaring voor het enorme verschil in gemiddelde score is dat het belangenconflict het oordeel van het testteam kleurt. In lijn hiermee is de bevinding dat bijna de helft van het verschil in oordeel zit in een subjectieve beoordeling van de kwaliteit van schoonmaken van de haring (op een schaal van zeer goed, goed, matig tot slecht) en de mate van rijping van de haring (op een schaal van licht, gemiddeld, sterk tot bedorven). Hierbij is het ook opvallend dat 60 procent van de haring die volgens het panel in de categorie ‘bedorven’ valt volgens het laboratorium een zeer goede microbiologische gesteldheid heeft. De laboratoriumtest zou namelijk ook een indicatie geven van de mate van bederf van de haring. Minder belangrijk zijn de objectieve kenmerken temperatuur van de haring en of de haring vers van het mes is, en niet belangrijk zijn de overige objectieve kenmerken gewicht, microbiologische gesteldheid en vetpercentage van de haring.

Deze punten vormen geen direct bewijs van het bevoordelen van vishandels met de bewuste leverancier in de AD haringtest, maar het testteam heeft op basis van dit onderzoek wel alle schijn tegen.

Appendix. Het verklaringsmodel

Het verklaringsmodel ziet er als volgt uit:

eindcijfer = a. gewicht + b. temperatuur + c. vet + d. vers + e. microbiologische gesteldheid + f. rijping + g. schoonmaken + h. regio Rotterdam + i. top10 + j. indicator 2017 + constante + restterm

In dit model staat *gewicht* voor het gewicht van de haring in grammen; *temperatuur* voor de temperatuur in graden celcius van de haring (beneden 7 graden/tussen 7 en 10 graden/boven 10 graden); *vet* voor het door het laboratorium vastgestelde vetpercentage van de haring (beneden 10%/tussen 10 en 14%/boven 14%); *vers* voor of de haring vers bereid is of niet; *microbiologische gesteldheid* voor de door het laboratorium vastgestelde microbiologische gesteldheid van de haring ((zeer) goed/voldoende/slecht/waarschuwingfase/afgekeurd); *rijping* voor de mate van rijping van de haring (licht/gemiddeld/sterk/bedorven); *schoonmaken* voor de kwaliteit van schoonmaken van de haring (zeer goed/goed/matig/slecht); *regio Rotterdam* voor het gevestigd zijn binnen een straal van 30 kilometer van Rotterdam (1/0); *top10* voor verkooppunten in de top 10 van de haringtest; *indicator 2017* voor in 2017 bezochte zaken (de rest is in 2016 bezocht). Het model heeft verder een constante term en tenslotte een restterm; dat wat niet verklaard wordt uit voorgaande variabelen. De termen a, b, c, etc. geven aan hoe deze kenmerken het testcijfer beïnvloeden.

Tabel A1 op de volgende pagina geeft de resultaten weer van de schatting van het verklaringsmodel. Het model verklaart 84 procent van de variantie in het eindcijfer. Veel, maar niet alles. Wat is die overige 16 procent? Allereerst zou de voorspelling van het model op basis van de kenmerken als rijping en kwaliteit van schoonmaken imperfect kunnen zijn. Zo gaat het model ervan uit dat elk kenmerk een afzonderlijke invloed heeft op het eindcijfer, terwijl het ook zou kunnen zijn dat het puntenverlies voor een relatief warme haring die ook nog slecht is schoongemaakt groter is dan het puntenverlies van de hoge temperatuur en het slechte schoonmaken afzonderlijk. Dit is een redelijke veronderstelling, omdat het testpanel zegt dat elke factor inderdaad een afzonderlijke invloed heeft op het eindcijfer. Daarnaast kan het zo zijn dat het eindcijfer beïnvloedt wordt door andere zaken dan de kenmerken van de geteste haring die we in dit onderzoek niet meenemen. Denk bijvoorbeeld aan de invloed van de volgorde van op een testdag bezochte zaken op het oordeel.¹⁷ Daarnaast staat bij de achtergrondgegevens van de haringtest voor elke beoordeelde zaak ook een vrij invulveld met opmerkingen over schoonmaak, geur, smaak en presentatie (bijvoorbeeld: 'frisheid verloren, karamel verkleurd', enz.). Deze opmerkingen komen gedeels terug in de gerapporteerde kenmerken, maar wellicht niet voor de volle 100 procent. Kortom, het model voorspelt het gegeven eindcijfer goed, maar niet perfect.

¹⁷ Zie Victor A. Ginsburgh en Jan C. van Ours, 2003, Expert opinion and compensation: evidence from a musical competition, *American Economic Review*, 93 (1), 289-296.

Tabel A1. Verklaring van het eindcijfer in de haringtest 2016 en 2017

Te verklaren: eindcijfer haringtest (0 tot 10)	
Gewicht (gr)	0,04 (0,01)***
Temperatuur	
beneden 7°C	referentiecategorie
tussen 7 en 10 °C	-0,58 (0,19)***
boven 10 °C	-1,71 (0,22)***
Vetpercentage	
beneden 10%	referentiecategorie
tussen 10 en 14%	0,18 (0,19)
boven 14%	0,66 (0,25)***
Vers van het mes	
niet	referentiecategorie
wel	1,78 (0,20)***
Microbiologische gesteldheid	
(zeer) goed	referentiecategorie
voldoende	-0,14 (0,31)
slecht	-0,50 (0,44)
waarschuwingfase	-0,14 (0,28)
afgekeurd	-2,45 (0,68)***
Rijping	
licht	referentiecategorie
gemiddeld	-0,36 (0,33)
sterk	-1,87 (0,38)***
bedorven	-4,59 (0,49)***
Schoonmaken	
zeer goed	referentiecategorie
goed	-0,82 (0,21)***
Matig	-1,56 (0,22) ***
Slecht	-2,64 (0,43)***
Regio	
binnen straal van 30 km rond R'dam	referentiecategorie
meer dan 30 km van R'dam	-0,30 (0,17)*
Top 10	
top 10	referentiecategorie
buiten top 10	-0,96 (0,32)***
Indicator 2017	
haringtest 2016	referentiecategorie
haringtest 2017	0,21 (0,17)
Constante	3,72 (0,75)***
Aantal waarnemingen	292
Verklaarde variantie (R ²)	0,84

Noot: Tussen haakjes staan de standaardfouten rond de geschatte coëfficiënten. *** statistisch significant op het 1 procent-niveau; ** op het 5 procent-niveau en * op het 10-procent niveau.