

STUDENTENOORDELEN PER OPLEIDING, EEN METHODOLOGISCHE TOELICHTING

Toelichting bij aanpak en resultaten van
de Nationale Studentenenquête 2005

Augustus 2005

DE STUDENTENOORDELEN VAN CHOICE, EEN METHODOLOGISCHE TOELICHTING

Elk jaar organiseert het Centrum Hoger Onderwijs Informatie (CHOICE) een enquête bij honderden opleidingen in het Nederlandse hoger onderwijs. Hoewel de naam Choice pas drie jaar bestaat, heeft de enquête al een historie van tien jaar.

Per opleiding wordt op hetzelfde moment en op dezelfde wijze aan studenten een kwaliteitsoordeel gevraagd over een reeks aspecten van hun opleiding. De vragen betreffen zowel de inhoud en samenhang van het studieprogramma als de vormgeving van het onderwijs, het contact met de docenten, de eventuele stages en de faciliteiten.

Deze Nationale Studenten Enquête (NSE) mondt uit in **studentenoordelen** per opleiding, die enerzijds gepubliceerd worden in de Keuzegids Hoger Onderwijs en andere publicaties voor studiekeizers, en daarnaast op vele niveaus gebruikt worden voor kwaliteitszorg en benchmarking in het hoger onderwijs.

De rapportcijfers zijn strikt genomen geen vergelijkend oordeel - de meeste studenten kennen immers alleen hun eigen opleiding. Toch zijn de studentenoordelen van de NSE in de loop der jaren erkend geraakt als een gezaghebbende indicator voor de 'toestand van het onderwijs' bij opleidingen en instellingen. Door de wijze van steekproeftrekking, de opzet van de vragenlijst en de stapsgewijze verwerking en analyse van de resultaten wordt een optimale betrouwbaarheid en vergelijkbaarheid gerealiseerd.

In dit document wordt in kort bestek¹ inzicht gegeven in de aanpak van de enquête en de betekenis van de resultaten die eruit voortkomen. Daarbij wordt ingegaan op kritische vragen die men bij dergelijk onderzoek dient te stellen, zoals: "Hoe betrouwbaar en representatief zijn de resultaten?". Of: "Wie controleert de kwaliteit van de analyses?"

Met dit rapport geven wij een zo eerlijk mogelijk beeld hoe lang de polsstok van de studentenoordelen eigenlijk is. Met deze transparantie hopen wij bij te dragen aan een passend gebruik van de enquêteresultaten. Tegelijk zijn wij voorbereid op nieuwe discussie en kritische vragen. Wie zulke vragen heeft, nodigen wij van harte uit ze op ons af te vuren.

Wie zoekt naar een meer gedetailleerde kwantitatieve analyse van meerdere jaren studentenoordelen, wordt in dit beknopte rapport nog niet bediend. Wel staat zo'n rapport op stapel. In het studiejaar 2005/2006 verwacht Choice een 'trendrapport' uit te brengen waarin enquêteresultaten van meerdere jaren worden geanalyseerd.

Frank Steenkamp
(penvoerder Choice)

Leiden, 2 augustus 2005

INHOUD

1. Choice: organisatie, verankering	p 3
2. Aanpak van de Nationale Studenten Enquête	p 4
3. Steekproefomvang: betrouwbaarheid, reproduceerbaarheid	p 7
4. Vragenlijst: inhoud, relevantie, samenhang	p 10
5. Resultaten: betekenis, externe validering	p 12
Bijlage: de thema's en indeling van de vragenlijst	p 15

¹ Dit stuk combineert hoofdpunten uit de jaarlijkse formele evaluaties van het project StudieKeuze-Informatie met de resultaten van interne methodenstudies en diverse externe presentaties en discussies. Er wordt niet in details naar deze documenten verwezen. In plaats daarvan zijn de belangrijkste feiten eruit integraal opgenomen.

1. Choice: organisatie, verankering

Centrum. De Nationale StudentenEnquête (NSE) wordt uitgevoerd en gerapporteerd door CHOICE, het Centrum Hoger Onderwijs Informatie voor Consument en Expert in Leiden. Dit centrum is in 2003 formeel van start gegaan als een samenwerkingsverband van organisaties die reeds sinds 1995 gezamenlijk op dit terrein actief waren. Het betreft:

- Research voor Beleid (RvB) te Leiden, onderdeel van de EIM-groep.
- Hoger Onderwijs Persbureau te Leiden (HOP), een zelfstandige stichting

Project. Choice voert de studentenenquêtes uit in het kader van het project Studie-Keuze-Informatie. Dit SKI-project is een overheidsopdracht, die na een Europese aanbestedingsprocedure door het ministerie van Onderwijs (OCW) aan Choice is gegund. Het project loopt van 2003 tot en met 2005, met een optie tot verlenging. Inmiddels heeft staatssecretaris Rutte inderdaad tot verlenging besloten. Dit betekent dat de enquêtes voorsnog tot en met 2008 worden voortgezet.

Organisatie. De rolverdeling binnen Choice is als volgt:

- RvB is verantwoordelijk voor de adequate planning, uitvoering en statistische verwerking van de studentenenquêtes.
- HOP verzorgt de consolidatie van enquêteresultaten in de SKI-database en de rapportage via de Keuzegids en eventuele andere publicaties.

In het team dat het SKI-project uitvoert, wordt ervaring met enquêtemethodes en statistiek gecombineerd met 'veldkennis' van het hoger onderwijs. Medewerkers zijn lid van relevante beroepsverenigingen, waaronder de uit het H.O. voortgekomen Dutch Association of Institutional Researchers (DAIR). Over methodes en deelprojecten uit het SKI-project zijn presentaties verzorgd op DAIR-seminars, op de landelijke Onderwijs Researchdagen (ORD) en op verschillende congressen. Bovendien is er recent in enkele projecten samengewerkt met de Onderwijsinspectie.

Extern toezicht. Naast de interne kwaliteitszorg van RvB en HOP heeft de overheid ook voorzien in extern toezicht bij de uitvoering van het SKI-project. Sinds voorjaar 2003 functioneert daartoe het *Kwaliteitscollege StudieKeuze-Informatie*, met momenteel als voorzitter oud-staatssecretaris mevr T. Netelenbos. Choice legt jaarlijks in een gedetailleerde zelfevaluatie verantwoording af aan dit college. Ook de enquêtemethode en –resultaten zijn daarbij voorwerp van beschouwing.

In dit kader is er speciale aandacht voor methodologische kwesties. Om enkele thema's op het terrein van steekproefgrootte, betrouwbaarheid en samenhang in resultaten uit te diepen is in april 2005 door het ministerie van onderwijs een rondetafelgesprek gehouden met vier hoogleraren. In dit rapport wordt gebruik gemaakt van de aanbevelingen uit deze bijeenkomst, die ook zijn overgenomen door het Kwaliteitscollege.

Relevantie. Naast het Kwaliteitscollege heeft de overheid voor het SKI-project ook een 'Gebruikers-platform' ingesteld. Hierin zijn organisatie van scholieren, studenten en schooldekanen vertegenwoordigd. Dit platform ziet er op toe dat de door Choice verzamelde informatie over opleidingen voldoet aan eisen van actualiteit, relevantie en toegankelijkheid. De ervaringen uit de eerste drie jaar van het SKI-project bracht ook het gebruikersplatform onlangs tot het advies om de huidige studentenenquêtes de komende drie jaar door Choice te laten voortzetten.

2. De aanpak van de Nationale Studentenenquête (NSE)

In de Nationale Studentenenquête (NSE) wordt per opleiding aan studenten gevraagd rapportcijfers te geven voor een reeks kwaliteitsaspecten van hun opleiding en instelling. De 'studentenoordelen' zijn gebaseerd op de gemiddelde scores die door studenten per aspect zijn toegekend. Hier beschrijven wij stapsgewijs de organisatie van de enquête².

Welke opleidingen? N=1200. Cyclus en bundeling

Hoewel de NSE grootschalig is opgezet (ruim 20.000 ondervraagden per jaar), is er niet voor gekozen alle relevante opleidingen elk jaar te onderzoeken. Analyse van de resultaten heeft immers geleerd dat deze een aanzienlijke stabiliteit kennen (zie paragraaf 3).

Deze stabiliteit maakt het mogelijk ieder jaar over alle voltijdse bacheloropleidingen³ geldige studentenoordelen te publiceren, die gebaseerd zijn op een enquêtecyclus van maximaal drie jaar. De spelregels zijn als volgt:

- *Per discipline.* Verwante opleidingen worden, behoudens inhaalacties⁴, in hetzelfde jaar onderzocht. Vergelijking van oordelen begint altijd bij de groep verwante opleidingen.
- *Frequentie.* Opleidingen worden tenminste eens in de drie jaar onderzocht; in belangrijke disciplines met relatief grote instroom is de cyclus tweejaarlijks.

Op deze wijze komen elk jaar ruim 400 opleidingen of combinaties van opleidingen aan bod. Die combinaties worden gevormd als een instelling meerdere verwante opleidingen aanbiedt, die elk voor zich te kleinschalig zouden zijn voor een enquête. Een voorbeeld zijn de wo-opleidingen Frans, Spaans en Italiaans, die in de enquête worden gebundeld in een cluster 'Romaanse talen'. Per instelling wordt dus bij elk van deze talen hetzelfde studentenoordeel gepubliceerd, met een voetnoot die verwijst naar de bundeling.

Als resultaat van de geschetste cyclus kunnen elk jaar studentenoordelen over meer dan 1200 opleidingen worden gepubliceerd.

De steekproef: aselect, N=50

Voor het verkrijgen van betrouwbare resultaten is het van belang om een representatieve steekproef uit de studentenpopulatie te enquêteren. In de loop der jaren is hiervoor de volgende aanpak ontwikkeld:

- *Bestand.* Basis voor de steekproef, zijn de actuele NAW-gegevens van ingeschreven studenten per opleiding, dat verkregen wordt van de IB-Groep.
- *Dwarsdoorsnee.* De enquête wordt gericht op een dwarsdoorsnede van de gehele studentenpopulatie - dwars op studiejaren, geslachten en vooropleidingstypen. Bij grote opleidingen wordt per opleiding 'at random' een steekproef van 300 studenten getrokken. Bij kleinere opleidingen start de procedure met alle studenten.
- *Vaste enquêteperiode.* De enquête loopt voor alle opleidingen van begin maart tot eind april. Reden voor deze keus is o.a. dat eerstejaars studenten dan al een meer dan oppervlakkige indruk van hun opleiding hebben.
- *Telefonische benadering.* Om onbedoelde uitval of selectie van ondervraagden te voorkomen, wordt de enquête niet per post of op locatie gehouden maar telefonisch. Bij deze benaderingswijze is het aantal weigeringen bescheiden (<15%) en de representativiteit het hoogst. Om de respons te maximaliseren, worden respondenten die de telefoon niet opnemen nog tweemaal opnieuw gebeld.

² Deze beschrijving betreft vooral de basisversie van de Nationale Studentenenquête, gericht op bekostigde voltijdse bachelor-opleidingen. Recent zijn versies ontwikkeld voor reguliere deeltijdopleidingen, Ma- en niet-bekostigde Ba-opleidingen. Deze enquêtes kennen dezelfde principes en structuur. Er zijn geringe verschillen in steekproef en/of vragenlijst. In de handleiding bij de SKI-database en op de website van Choice worden actuele details over deze varianten opgenomen.

³ Alleen te kleine en te jonge opleidingen worden niet onderzocht, omdat hier geen betrouwbare resultaten gegarandeerd kunnen worden. Choice hanteert als vuistregels: minstens 15 eerstejaars, en minstens 3 jaargangen studenten.

⁴ Soms kan een opleiding in jaar X niet onderzocht worden, vanwege 'te jong' of technische (steekproef-)problemen. Zonodig vindt dan in jaar X+1 een 'herkansing' plaats, waarna de resultaten worden ingevoegd bij de groep verwante opleidingen.

- *Netto steekproef = 50.* Bij grote opleidingen gaat de enquête door totdat er 50 studenten zijn bereikt. Bij kleinere opleidingen ligt dit aantal lager, met een minimum van 20% van de studentenpopulatie. Dit blijkt voldoende voor de nagestreefde precisie (zie par. 3)

De vragenlijst

De Nationale Studentenenquête bevat diverse controlevragen en vragen naar de achtergrond van de student (klopt de studie? Welke leeftijd, vooropleiding?). De kern wordt echter gevormd door de vragen waarop de studentenoordelen per opleiding worden gebaseerd. Dit centrale deel van de vragenlijst kent de volgende opzet:

- *Rapportcijfers op 30 deelaspecten.* Aan studenten wordt gevraagd een oordeel in rapportcijfers (een 10-puntsschaal) te geven, over een reeks aspecten van de opleiding. Er is naar gestreefd om alleen oordelen te vragen over zaken waarover studenten redelijkerwijs een oordeel kunnen hebben. Verder is de vraagstelling zakelijk. De stijl van een populariteitspoll wordt zo veel mogelijk vermeden. Dus geen vragen zoals: "Ben je tevreden over je docenten?", maar een uitnodiging om te oordelen over de vakkennis, didactische kwaliteit en bereikbaarheid van docenten.
- *Ordering in 10 vaste thema's.* De enquêtevragen zijn geordend in 10 thema's, die alle bekende issues van de kwaliteitszorg bestrijken (van onderwijsinhoud en docenten tot organisatie, studeerbaarheid en faciliteiten). De issues zijn wel gedefinieerd vanuit de invalshoek van het onderwijsproces zelf. Bekende thema's uit visitatie en accreditatie zoals de formele organisatie en papieren kwaliteitszorg komen niet aan de orde.

Vraagformuleringen worden met enige regelmaat bijgesteld, maar de 10-deling van de kwaliteitsissues houdt Choice zo veel mogelijk stabiel - dit omwille van de vergelijkbaarheid over meerdere jaren. Juist per 2005 is overigens ook de thema-indeling vernieuwd. Op deze herindeling is al vanaf 2003 in de vragenlijst geanticipeerd, zodat direct weer over drie jaar historie beschikt kan worden.

De precieze vragenlijst en indeling ervan is opgenomen als bijlage bij dit rapport. Op de kwaliteitsbewaking van de vragenlijst gaan we in bij paragraaf 4.

Verwerking en presentatie van de oordelen

- *Getrapte rapportcijfers.* Per opleiding wordt voor elke vraag de gemiddelde score van alle respondenten berekend. Deze rapportcijfers worden groepsgewijs gemiddeld tot scores voor 10 kwaliteitsissues. Het (ongewogen) gemiddelde van deze 'clusterscores' is de totaalscore van de opleiding. In de praktijk variëren de rapportcijfers per vraag van individuele respondenten tussen 1 en 10 en de gemiddelde vraagscores per opleiding tussen 3 en 9. De clusterscores van opleidingen variëren van 4 tot 8,5. De totaalscores van opleidingen lopen uiteen van 5,9 tot 7,9.
- *Marges.* Omdat niet de hele populatie ondervraagd is, zijn de resultaten slechts een benadering van "het" studentenoordeel. Er is statistische onzekerheid. Om deze grijpbaar te maken, wordt van alle cijfers per opleiding 95%-betrouwbaarheidsmarge berekend en opgenomen in de SKI-database. De marge bij de totaalscore wordt gepubliceerd in de Keuzegids en in de 'schermen' bij de database. De marge geeft aan binnen welke range het cijfer zou uitkomen (met 95% zekerheid), als alle studenten ondervraagd werden. In de praktijk is de marge bij afzonderlijke vragen ongeveer 0,5 punt, bij de clusterscores ongeveer 0,3 punt en bij de totaalscores rond 0,15 punt. Als we dus een totaalscore van 6,85 publiceren, weten we in dit geval dat de 'werkelijke' score tussen 6,7 en 7,0 punt ligt. [NB: in alle gevallen waar de marge op de totaalscore 0,25 punt of meer bedraagt, geeft Choice dit apart aan.]

- *Context van verwante studies.* Mogen we het oordeel van de pabo-studenten over het niveau van de stof vergelijken met dat van de universitaire chemici? Choice moedigt dit niet aan. Tussen disciplines bestaan verschillen in type student en in onderwijscultuur. Bovendien zijn de pabo en de chemie-opleiding geen directe concurrent van elkaar⁵. Om die reden staat vergelijking tussen verwante opleidingen binnen de Nationale Studentenenquête centraal. Choice zelf presenteert studentenoordelen over opleidingen in de Keuzegids altijd primair per discipline. Ook aan andere gebruikers van deze oordelen wordt dit aanbevolen. Voorzover er toch studentenoordelen per instelling worden gepresenteerd, bouwt Choice deze op uit de *relatieve scores* van die instelling in de diverse disciplines.

Significanties. Een vraag bij gepubliceerde oordelen is altijd of de gevonden verschillen statistisch voldoende betrouwbaar zijn. Choice berekent daarom bij de 10 clusterscores en de totaalscore van elke opleiding, of deze "significant" (voor 95% zeker) afwijken van het gemiddelde van de verwante opleidingen. Deze significanties werden tot 2004 in de ranglijsten van de Keuzegids weergegeven met pijltjes omhoog en omlaag bij de rapportcijfers.

Volgens de door onze opdrachtgever verzamelde methodologen (zie p. 3, "extern toezicht") kon deze presentatie misbegrepen worden door studiekeizers. Als één opleiding met een totaalscore van 7,11 gemarkeerd werd als significant positief en een tweede opleiding met een score van 7,10 kreeg deze markering niet, kon de indruk ontstaan van een harde klassegrens tussen beide opleidingen. In werkelijkheid voldeed de meting bij de ene opleiding (net) wel aan het 95%-criterium en de andere (net) niet. Van een hard verschil tussen beide opleidingen was uiteraard geen sprake.

Choice zal nu in de ranglijstjes per studie in 2005 geen significanties meer weergeven. Wel blijft dit stuk statistiek van nut als indicator voor benchmarking: voor opleidingen die hun eigen positie ten opzichte van de 'concurrentie' willen beoordelen, blijven de significantiescores daarom opvraagbaar.

⁵ Zie ook de nadere beschouwing over vergelijkbaarheid van oordelen in paragraaf 5.

3. Representativiteit, betrouwbaarheid, reproduceerbaarheid

Bij het beoordelen van de enquêteresultaten is een aantal thema's relevant. De meest gestelde vraag is hoe betrouwbaar een studentenoordeel op basis van 50 studenten kan zijn. Die vraag komt hier aan de orde, maar er zijn ook andere vragen van belang.

De representativiteit van de steekproef

Bij een survey als de NSE gaan we uit van een studentenpopulatie waaruit we een 'random' steekproef hebben getrokken. In de praktijk telt elke opleiding jonge en oude, minder en meer gemotiveerde studenten en studenten met positieve en negatieve ervaringen. Er bestaat een risico dat de enquête onbedoeld tot scheve vertegenwoordiging van zulke groepen bevat.

Voorzover mogelijk, worden zulke effecten door de onderzoeks aanpak uitgesloten. De telefonische benadering is mede hierom gekozen (bij een enquête op locatie speelt een onbedoelde selectie op feitelijke aanwezigheid, motivatie, enz). Op twee manieren worden verder eventuele selectie-effecten gemonitord. De reden van uitval van respondenten wordt bijgehouden. En de verdeling van respondenten over studie jaren, geslachten en vooropleiding wordt gerapporteerd. Deze laatste kan vergeleken worden met bekende kenmerken van de gehele populatie. Uit deze gegevens blijkt het volgende:

- Uitval tijdens de enquête komt vooral door 'technische' redenen. Geen telefoonnummer beschikbaar, niet bereikbaar, gegevens onjuist. Het percentage weigeringen is gering. Op 26.000 daadwerkelijke telefonische contacten betrof dit in 2005 minder dan 15%.
- De enquête biedt, voor zover meetbaar, in het algemeen een evenwichtige afspiegeling van de populatie wat betreft leeftijd, geslacht, vooropleiding en studievordering.
- Van buitenlandse studenten is duidelijk dat deze in de enquête ondervertegenwoordigd zijn. Choice heeft een Engelstalige vragenlijst, maar het probleem doet zich voor op een ander niveau. Van deze groep zijn veel minder telefoonnummers beschikbaar.

De conclusie is dat de steekproef van de NSE op de tot nu toe gecontroleerde dimensies representatief geacht mag worden voor de populatie van *Nederlandse* (dwz: reeds eerder in Nederland verblijvende) studenten bij de betreffende opleiding.

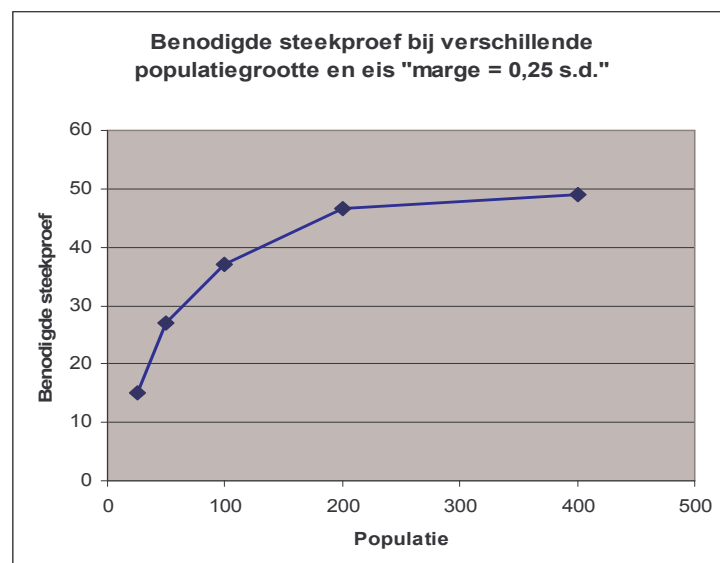
Om meer zekerheid over de representativiteit van de resultaten te verkrijgen, zal Choice de komende jaren ook controlevragen stellen over motivatie en ambitie van studenten.

Betrouwbaarheid bij deze steekproefomvang

Dit is een vast onderwerp van de statistiek bij enquêtes. Als men de spreiding in resultaten (de standaarddeviatie of s.d.) kent, valt bij elke steekproefgrootte de betrouwbaarheidsmarge te berekenen⁶.

Bij de nationale studentenenquête is de s.d. van de totaalscore bij de meeste opleidingen 0,6 punt. Volgens de wetten van de statistiek is dan bij een oneindig grote populatie een steekproef van 50 studenten genoeg om een marge van 0,15 punt te realiseren. ($0,25 \times s.d.$)

Bij opleidingen met minder dan 200 studenten kan dezelfde precisie bereikt worden met een kleinere steekproef (zie grafiek). De steekproef voor de masterenquête is in 2005 volgens deze regels opgebouwd, om bij alle opleidingen dezelfde nauwkeurigheid te realiseren.



⁶ Een voorwaarde is wel dat de scores 'normaal verdeeld' zijn. Dit is bij de NSE in het algemeen het geval.

En hoe werkt dit nu in praktijk uit? Uit de volgende cijfers blijkt dat de gerealiseerde marge (of: marge m), gemiddeld over alle opleidingen*, in de voltijdenquête precies overeenkomt met de streefwaarde.

<u>Enquête-aflevering</u>	<u>m clusterscores</u>	<u>m totaalscore</u>
Voltijd 2003	0,21 tot 0,27	0,15
Voltijd 2004	0,19 tot 0,28	0,15
Voltijd 2005	0,19 tot 0,28	0,15
NBI 2004/5	0,20 tot 0,29	0,15
Deeltijd 2004	0,24 tot 0,33	0,19
Deeltijd 2005	0,23 tot 0,36	0,19
Masters 2005	0,31 tot 0,44	0,23*

*) Dit zijn gerealiseerde cijfers, voorafgaand aan opschoning. Bij enkele opleidingen werden te weinig studenten bereikt en/of was de marge op de totaalscore groter dan 0,3 punt. Deze oordelen worden niet gepubliceerd.

Behoudens een enkele uitzondering vertonen de enquêteresultaten dus de beoogde precisie. In gevallen waarin de oordelen een onverwacht grote marge vertonen, zal Choice de resultaten niet - of met een duidelijke disclaimer - publiceren.

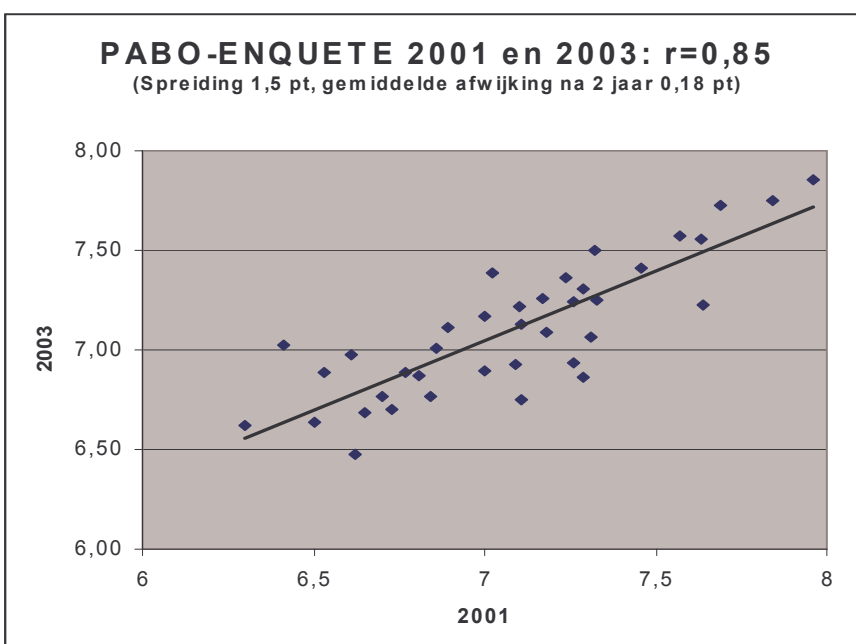
Stabiliteit over de jaren

Een andere vraag betreft de 'houdbaarheid' van enquêteresultaten. Choice neemt in zijn database en daarop gebaseerde publicaties bij een aantal opleidingen nog oordelen op van twee jaar geleden. Is de actualiteit van deze cijfers getoetst?

Inderdaad heeft Choice de stabiliteit van de studentenoordelen in het recente verleden getoetst. Omwille van de ijkings van de resultaten is bij drie grote universitaire studies (rechten, economie en geneeskunde) in de periode van 1995 tot 2003 bijna jaarlijks een enquête gehouden, met een dubbele steekproefomvang. Op deze wijze werd de precisie van elke meting vergroot en werd een beeld verkregen van de stabiliteit over de jaren.

Uit dit meerjarige onderzoek blijkt dat - mits de hoogste en laagste eindoordelen meer dan 0,5 punt uiteen liggen - de ranglijst per studie een aanzienlijke stabiliteit vertoont. Er zijn wel verschuivingen zichtbaar, maar deze bedragen in 2 jaar tijd gemiddeld niet meer dan 0,12 punt. Over een periode van 3 jaar is de gemiddelde verschuiving niet groter.

Een grote hbo-opleiding die elke twee jaar aan een enquête wordt onderworpen is de pabo. Ook hier bleek bij vergelijking van de resultaten uit 2003 en 2001 een sterke overeenkomst



tussen beide ranglijsten. Strikt genomen betekent dit dat de resultaten uit 2001 ook na twee jaar nog een redelijk actueel beeld gaven; desalniettemin verving Choice ze op dat moment door nieuwe studentenoordelen.

Deze stabiliteit lijkt in tegenstrijd met het beeld dat opleidingen na kritiek snel tot actie overgaan om knelpunten op te lossen. Ook als dit beeld juist is, zijn er echter

verschillende mechanismen die een snelle verschuiving in studentenoordelen verhinderen:

- *Doorlooptijd* van veranderingen - Als Choice in maart 2001 enquêteert en de resultaten in oktober 2001 publiceert, kunnen vernieuwingen niet eerder doorgevoerd worden dan per september 2002. De enquête in maart 2003 komt dan nog vroeg - als men de traagheid van organisaties en het na-ijlen van student-tevredenheid in aanmerking neemt.
- *De meting*. Direct ingevoerde veranderingen in opleidingen leiden zelden in één klap tot grote verschuiving in studentenoordelen. Vernieuwing start vaak in de propedeuse, maar in de enquête van Choice vormen de eerstejaars gemiddeld slechts 30 tot 35 procent van de ondervraagde populatie. Het voordeel van deze werkwijze is dat ook een beeld gegeven wordt van de onderwijskwaliteit in hogere jaren, maar de andere kant van de medaille is de 'gedempte' meting van vernieuwingen.

De meest abrupte stijgingen in het studentenoordeel zijn de afgelopen jaren door Choice gemeten bij opleidingen en instellingen die verhuisden naar nieuwbouw en tegelijk zwaar investeerden in nieuwe faciliteiten. De positieve ervaring met de faciliteiten gaat in zulke gevallen nogal eens samen met gunstiger oordelen over andere zaken, zoals de docenten en de communicatie van de opleiding. In veel gevallen blijkt dit positieve effect van verhuizing overigens binnen enkele jaren weg te ebben.

4. De vragenlijst: relevantie en samenhang

De vragenlijst van de Nationale Studentenenquête is het product van een lang ontwikkelingstraject. De thema's uit het nationale 'discours' over onderwijskwaliteit zijn erin neergeslagen, net als ideeën uit andere enquêtes en issues die scholieren belangrijk blijken te vinden volgens de instroommonitor en vergelijkbare onderzoeken.

Bijna elk jaar vinden er ook aanpassingen plaats in de vragenlijst. Vaak subtiele, na kritiek of vragen van gebruikers. De hoofdstructuur van de enquête wordt echter, omwille van de vergelijkbaarheid over de jaren, zo veel mogelijk constant gehouden.

Een risico van een zo lange geschiedenis (met eerste pretests en diepte-interviews die dateerden uit de vorige eeuw) is toch dat het onderzoek 'out of touch' zou raken met de actualiteit. Daarom is in het kader van het SKI-project nieuw onderzoek gedaan om de relevantie van gestelde vragen en de inhoudelijke samenhang in de vragenclusters te toetsen. Dit is gebeurd voor alle deelenquêtes binnen de NSE. Hier rapporteren wij over de breedst opgezette evaluatie, die de enquête bij voltijd-bacheloropleidingen betrof.

De relevantie van de issues voor betrokkenen bij studiekeuze

In 2004 is uitvoerig onderzoek gedaan naar de actuele ideeën over relevante studiekeuze-informatie bij scholieren, dekanen, ouders en studenten. Het betrof zowel groepsgesprekken, als telefonische enquêtes en de input in een interactieve studiekeuzetest.

Alleen de belangrijkste resultaten worden hier weergegeven.

Naast elkaar worden twee sets oordelen getoond over het 'belang' van de issues uit de NSE:

- a) *NRC-/Choice-Studiekeuzetest* najaar 2004. Totaal 13.500 personen, in grote meerderheid scholieren in de studiekeuzeleeftijd. Respondenten konden aan issues een 'gewicht' toekennen om hun eigen ranking van opleidingen te sturen. Niet alle hoofdissues van de NSE waren expliciet in de test opgenomen.
- b) *Monitor-module binnen de NSE*. Hierin geven 5000 van de huidige studenten, als volleurde 'ervaringsdeskundigen' aan welke kwaliteitsissues volgens hen van belang zijn.

Belang van issues uit NSE volgens twee doelgroepen				
Onderwerp	(a) scholieren		(b) huidige studenten	
	havo	vwo	hbo	wo
Inhoud	8,0	8,4	7,9	8,3
Docenten	7,5	7,7	7,4	7,5
Voorber. Arbmarkt	6,8	6,7	7,4	7,5
Keuzeruimte	-	-	7,2	7,3
Samenhang progr.	-	-	7,1	7,3
Faciliteiten	7,6	7,8	7,2	7,2
Werkvormen	-	-	7,0	7,2
Communicatie	6,5	6,2	7,0	7,2
Prakt. Organisatie	6,8	6,8	6,9	7,1
Gebouwen	-	-	6,7	6,8

We zien in het algemeen gunstige scores: men vindt de gemeten issues relevant. Verder springt de "inhoud" van het onderwijs er als belangrijkste uit, terwijl enkele praktische punten wat minder zwaar wegen. Toch zijn de verschillen niet groot. Zelfs de kwaliteit van gebouwen wordt (volgens de studenten van 2004) met een "7 min" nog vrij belangrijk gevonden.

Aan dit subtiele verschil in gewicht komt Choice al tegemoet, door ervoor te zorgen dat de meerderheid van de 10 issues in de actuele opzet van de enquête het onderwijs zelf betreft. Slechts 4 issues gaan over organisatie en praktische randvoorwaarden. Er is geen basis om bij de standaardberekening van totaalscores per opleiding een verschil in weging tussen de issues aan te brengen. Dit zou slechts de transparantie van de enquête-opzet verlagen.

De samenhang binnen de 10 vragenclusters van de enquêtes

De vragenlijst van de NSE bestaat uit tien clusters: de scores op de onderliggende vragen worden steeds gemiddeld tot één clusterscore. De hantering van zulke gebundelde scores of 'schalen' is legitiem, mits er tussen de scores van de betreffende vragen voldoende samenhang bestaat. Dit is de afgelopen jaren in de NSE-resultaten getoetst. Deze toetsing heeft mede bijgedragen aan de recente herindeling van de vragenclusters. In 2003 en 2004 is daarom de samenhang van de deels vragen in de nieuw gevormde clusters berekend. Hier worden toetsingsresultaten uit de voltijdenquête gepresenteerd:

Als maat voor de samenhang wordt *Cronbach's alpha* gehanteerd. Deze grootte kan variëren tussen 0 en 1; een score van 0,5 wordt acceptabel geacht – ons streven is 0,6 of meer. De resultaten worden zowel getoond voor 2003 als 2004.

Consistentie vragenlijst NSE - voltijd bacheloropleidingen

	2003		2004	
	HBO	WO	HBO	WO
1. Inhoud	.58	.52	.64	.59
2. Keuzeruimte	.48	.52	.43	.47
3. Samenhang	.50	.56	.49	.49
4. Werkvormen	.55	.50	.57	.61
5. Voorbereiding loopbaan	.53	.51	.55	.55
6. Kwaliteit docenten	.69	.61	.69	.62
7. Communicatie	.67	.65	.65	.60
8. Studeerbaarheid	.63	.53	.67	.64
9. Gebouwen	.71	.66	.74	.69
10. Faciliteiten	.56	.54	.57	.61

De samenhang blijkt gemiddeld ruim voldoende, met scores rond de 0,6. Vooral bij de enquête in het WO is - na bijschaven van formuleringen - in 2004 sprake van een lichte verbetering: het aantal alpha-scores van 0,6 of hoger (vet gemarkeerd) nam duidelijk toe. De clusters keuzeruimte en samenhang blijven het zwakste. Beide vertonen nog een lichte achteruitgang tot scores iets onder de 0,5 (cursief gemarkeerd).

De deeltijdenquête laat in het algemeen iets hogere scores voor samenhang zien, met een gemiddelde ruim boven 0,6. Alleen de clusters 'voorbereiding loopbaan' (WO) en 'samenhang programma' (HBO) scoren onder de 0,5.

Een even gunstig verhaal geldt voor de masterenquête. Alle vragenclusters voldoen hier aan de gestelde minimumeis. Acht van de tien scoren zelfs hoger dan 0,6.

Al met al geeft de samenhang van de vragenclusters geen reden tot zorg. In 2005 is daarom geen wijziging aangebracht in de indeling van vragen. Wel is opnieuw geschaafd aan enkele formuleringen. De effecten hiervan worden in najaar 2005 in detail bekend.

5. De resultaten: betekenis, externe validering

Bij enquêtes met vergelijkende, evaluerende of voorspellende pretenties vallen altijd lastige vragen te stellen. Statistisch kan er aan vele eisen voldaan worden, maar hoe weten we dat er echt gemeten wordt wat de onderzoekers pretenderen?

In het geval van de Nationale Studentenuquête vragen we oordelen aan groepen studenten per opleiding, en kennen daarna aan die oordelen een ruimere betekenis toe:

- *vergelijkbaarheid*: we nemen aan dat de oordelen eerlijk vergeleken kunnen worden. Maar hebben de studenten op verschillende plaatsen wel vergelijkbare normen en verwachtingen?
- *kwaliteitsindicator*: we meten tevredenheid, en suggereren dat er een relatie bestaat met de feitelijke kwaliteit van opleidingen en instellingen. Maar is dat voldoende aannemelijk?

Het zijn geen vragen waar snel een definitief antwoord op te geven valt. Maar de vele analyses en discussies die wij aan dit onderwerp hebben gewijd, geven Choice wel het vertrouwen dat de wijze van presentatie van studentenoordelen legitiem is - met alle daarbij behorende nuancering. We gaan hier kort op beide genoemde kwesties in.

De vergelijkbaarheid van de oordelen

Omdat wij geen grote aantallen studenten op verschillende plaatsen kunnen laten studeren, moeten we bij deze vraag steunen op indirecte bewijsvoering:

- *tussen disciplines* zijn de studentenoordelen, zowel op empirische als fundamentele gronden, beperkt vergelijkbaar. Het onderwijsproces op een journalistenopleiding, een pabo of een opleiding natuurkunde kent zeer verschillende dimensies. Ook is er weinig reden om aan te nemen dat havo-meisjes naar de pabo gaan met dezelfde normen en verwachtingen als de slimme jongetjes die voor natuurkunde kiezen. De ervaring leert ook dat de journalistiekstudenten veel meer kritiek hebben dan hun collega's op de pabo. Choice erkent dat cultuurverschillen tussen disciplines een grote rol spelen, en presenteert ze bij voorkeur slechts als wat ze zijn: verschillen in tevredenheid.
- *tussen instellingen*, in dezelfde discipline of dwars op de disciplines (na correctie voor disciplineverschil!), pretendeert Choice een grotere vergelijkbaarheid. Dit is mede gebaseerd op hoe het hoger onderwijs in Nederland georganiseerd is (redelijk uniform voortgezet onderwijs, landelijk uniforme toelatingseisen en collegegelden) en op feiten over studiekeuzegedrag en studentenstromen. Voor studenten in een regio zijn de instellingen werkelijk onderling uitwisselbare concurrenten. In deze setting trekken verwante opleidingen in hoge mate vergelijkbare groepen studenten. Dit wordt gestaafd door populatiegegevens, en door resultaten die we in de volgende paragraaf bespreken.
- *Uitzondering (1): klein confessioneel*. De ervaring leert dat er een uitzondering bestaat. Bij *zeer* kleine instellingen die door hun identiteit een "eigen markt" van studiekeuzers kennen en een sterke loyaliteit van de eigen studenten, zijn de studentenoordelen niet altijd vergelijkbaar met die van de collega's elders. Er zijn gevallen van kleine religieuze of antroposofische opleidingen waar de studenten zeer positief oordeelden - terwijl op basis van een visitatierapport grote vraagtekens bij het studieklimaat gesteld moesten worden. Dit betekent niet dat bij zulke opleidingen nooit sprake is van kritiek, wel dat bij de presentatie van oordelen extra behoedzaamheid vereist is.
- *Mogelijke uitzondering (2): dure particuliere opleidingen*. Sinds kort verzamelt Choice ook studentenoordelen bij niet-bekostigde instellingen (NBI). Veel van de instellingen nemen, door hun prijsstelling, hun onderwijsmix en/of specialisatie op specifieke doelgroepen, op

de onderwijsmarkt een andere positie in dan reguliere opleidingen. Dit betekent dat ook hier in meer of mindere mate sprake is van een afwijkende markt en een afwijkende studentenpopulatie. Gemeten verschillen in kwaliteitsoordeel kunnen dus mede een afspiegeling zijn van verschillen in ambitie en verwachtingen van studenten. Choice zal dit aspect nader onderzoeken, maar publiceert de studentenoordelen over sommige NBI-opleidingen in 2005 wel samen met de oordelen over reguliere voltijdopleidingen. Dit gebeurt met een duidelijke voetnoot, die de lezer attendeert op de aparte status van de opleiding.

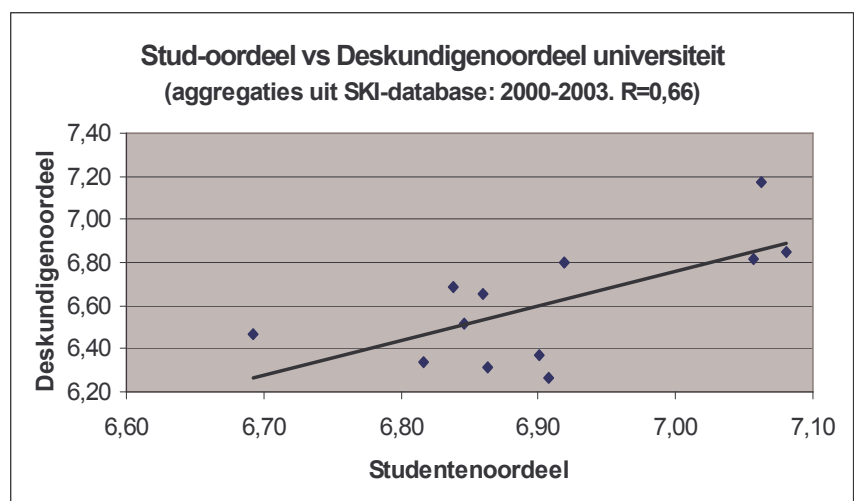
- *geen Randstad-effect.* Vaak wordt gesuggereerd dat studenten in de Randstad in het algemeen kritischer zijn dan studenten in het oosten van het land. Natuurlijk zijn er cultuurverschillen, maar leidt dit tot andere eisen of normen met betrekking tot het geboden onderwijs? Het is een lastige kwestie. Om een beeld van regionale verschillen te krijgen, heeft Choice enkele jaren geleden aan alle studenten een oordeel gevraagd over een dienst die landelijk opereert – de Informatie Beheer Groep (IBG). Studenten in de ‘provincie’ bleken precies even kritisch te oordelen als studenten in de grote steden. Het verwachte Randstad-effect werd niet bevestigd.

Externe validering: relatie met feiten uit andere bronnen

Een van de manieren om te beoordelen of de studenten enquêtes de realiteit meten die ze pretenderen te meten, is onderzoek naar de samenhang met metingen door andere partijen of op basis van anderssoortige gegevens. Hier presenteren wij er enkele, die steun geven aan de hiervoor uiteengezette visie:

- *Verband met andere studenten enquêtes.* Dit verband lijkt misschien vanzelfsprekend, maar dat is het toch niet per definitie. Elke enquête kent een andere vragenlijst, de aard van steekproef kan afwijken en het moment van het onderzoek speelt een rol. Toch is tot nu toe bij iedere vergelijking met andere grootscheepse enquêtes dat er een duidelijk verband is in resultaten. In 2003 is dit bijvoorbeeld geanalyseerd bij de enquête van Elsevier/NIPO in de sector HEO commercieel/communicatie (38 opleidingen). Ondanks de genoemde verschillen in aanpak was bleek er een sterk statistisch verband, met een correlatie van 0,66 tussen de Choice- en Elsevier-studentenoordelen. Eerder waren minutieuze analyses gedaan met interne enquêtes van de Hogeschool van Amsterdam en de Hogeschool Leiden. Hier werden tussen de ranglijsten van opleidingen correlaties van respectievelijk 0,85 en 0,95 berekend. Dit betekent dat de Choice-enquête een werkelijkheid meet die samenhangt met wat andere vergelijkbare enquêtes meten. Dit verband is sterker dan de invloed van verschil in onderzoeks aanpak en van meet-onzekerheid (zie betrouwbaarheidsmarges, p 7)

- *Verband met visitatiescores.* Oordelen van visitatiecommissies komen op geheel andere wijze tot stand dan studentenoordelen. Daarom is het interessant om te bezien of er tussen beide bronnen een verband bestaat. In 2003 is dit bepaald voor hele instellingen. Zowel uit studenten- als

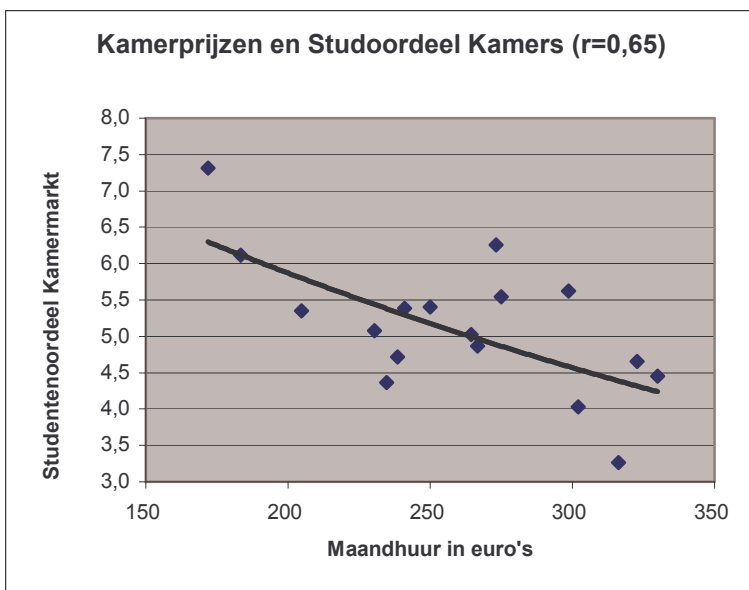


deskundigenoordelen⁷ per opleiding zijn gebundelde instellings-scores berekend. Het verband is voor 13 universiteiten in grafiek gezet.

Er is inderdaad een bescheiden samenhang.

Daarmee maakt de vergelijking aannemelijk dat studenten- en deskundigenoordelen “overlappende werkelijkheden” benaderen. Anders gezegd: Maastricht en Twente (rechtsboven en rechts) profileren zich sterker met onderwijskwaliteit dan de UvA of de TU Delft (geheel links en 2^e van links). Beide metingen maken dit verschil tastbaar – een wederzijdse validering.

- *Verband met harde cijfers: de kamermarkt.* Het is ook interessant om de studentenoordelen te relateren met harde feiten. Het probleem is dat er over onderwijskwaliteit weinig harde feiten te vinden zijn. Alleen de studietempocijfers bieden enig houvast. Sprekender is een ander voorbeeld, waarbij gebruik is gemaakt van de rapportcijfers die studenten in de marge van de NSE ook mogen geven voor de kamermarkt in hun studiestad. Gevraagd wordt een oordeel over de verkrijgbaarheid van een ‘betaalbare kamer’. Deze rapportcijfers zijn door ons in 2003 naast landelijke cijfers gelegd over kamerprijzen voor kamerzoekenden (bron: vraagprijzen www.kamernet.nl). Het resultaat laat een duidelijke samenhang zien. Ook dit mag opgevat worden als validering van de studentenoordelen. De *verschillen* in rapportcijfers die studenten in diverse steden geven, tonen een duidelijk verband met objectief meetbare verschillen.



⁷ De scores uit visitatierapporten zijn door Choice de afgelopen jaren bewerkt en samengevat in de Keuzegids Hoger Onderwijs. In verband met verschillen in aanpak van visitatiecommissies (en dus nog geringere vergelijkbaarheid over de disciplines dan bij studentenoordelen) zijn deze scores sinds 2004 niet meer opgenomen in de SKI-database. Choice zal de komende tijd de resultaten van het nieuwe accreditatiestelsel proberen te benutten voor studiekeuze-informatie.

BIJLAGE

THEMA'S EN INDELING VRAGENLIJST NATIONALE STUDENTENENQUETE 2005

(de 'nieuwe' clustering die met terugwerkende kracht per 2003 is ingevoerd) (*formulering vragen niet exact!*)

1. Inhoud programma:

- De interessantheid van de geboden studiestof
- Het niveau van de studiestof
- Het niveau van de tentamens

2. Keuzeruimte:

- De ruimte voor bijvakken en keuzevakken
- De mogelijkheid om eigen afstudeerrichting of specialisatie te kiezen
- De mogelijkheid om in het buitenland te studeren

3. Samenhang programma

- De samenhang tussen de studie jaren
- De samenhang tussen de afzonderlijke vakken

4. Werkvormen

- Het stimuleren van een kritische houding
- Het gebruik van 'activerend' onderwijs
- De oefening van praktische vaardigheden

5. Voorbereiding loopbaan

- De mate waarin het studieprogramma een helder beeld geeft van de arbeidsmarkt
- De praktische voorbereiding op de arbeidsmarkt
- De kennismaking met wetenschappelijk onderzoek (WO)/ toepassen van onderzoek (HBO)

6. Kwaliteit docenten

- De inhoudelijke deskundigheid van de docenten
- Het vermogen van de docenten om de stof helder over te dragen en je leerproces te stimuleren
- De beschikbaarheid van de docenten

7. Communicatie opleiding

- De tijdige bekendmaking van toetsresultaten en andere beoordelingen
- De communicatie over je rechten en plichten als student
- De communicatie over praktische zaken, zoals roosters e.d.

8. Studeerbaarheid

- De begeleiding van de studievoortgang
- De haalbaarheid om de propedeuse af te ronden in de tijd die er voor staat
- De haalbaarheid om de opleiding af te ronden in de tijd die er voor staat
- De mogelijkheden om examens/tentamens te herkansen

9. Gebouwen

- De geschiktheid van de collegezalen en lesruimtes
- De hoeveelheid werkplekken voor zelfstudie
- De geschiktheid van de practicum- en oefenruimtes

10. Faciliteiten

- De bibliotheek- en mediafaciliteiten
- De computer- en netwerkfaciliteiten
- De beschikbaarheid van het onderwijsmateriaal

Voor elk vragencluster wordt het ongewogen gemiddelde van de vraagscores berekend. Dit geldt als de clusterscore. De 10 clusterscores worden gemiddeld tot één totaalscore per opleiding.

NB: Voor deeltijdse Ba-opleidingen en voor masteropleidingen gelden aangepaste vragenlijsten. De studentenoordelen over deze opleidingen worden apart gepresenteerd. Ook voor niet-bekostigde opleidingen (NBI) geldt een licht aangepaste vragenlijst. De studentenoordelen over deze opleidingen worden waar mogelijk gepresenteerd in vergelijking met hun bekostigde tegenhangers. Details over de aangepaste vragenlijsten worden door Choice apart gepubliceerd.