



Datasets en bijbehorende opgaven

Traktatie van het 50-jarige Cito: opgaven voor SE statistiek met ICT
NWD 1 februari 2019

Ivo Claus (ivo.claus@cito.nl)

Irene van Stiphout (irene.vanstiphout@cito.nl)

Ger Limpens (ger.limpens@cito.nl)



Statistiek volgens cTWO

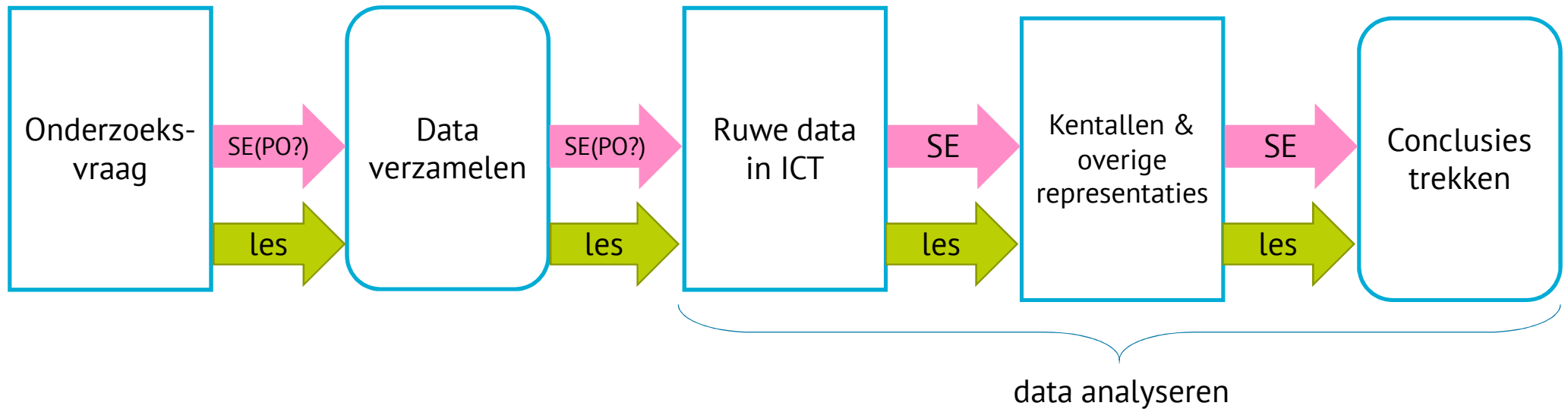
Eindrapport cTWO (vernieuwingscommissie) dec 2012:

Het onderwerp statistiek (domein E) wordt op een meer realistische en probleemgeoriënteerde manier benaderd dan voorheen. Uitgangspunt daarbij is de empirische cyclus van data verzamelen, data analyseren en conclusies trekken. ICT wordt gebruikt om grote datasets te analyseren.

Uit de syllabi havo wiskunde A en vwo wiskunde A in pilotperiode ('11 – '15):

Domein	In CE	Moet in SE	Mag in SE
E Statistiek		x	

Empirische cyclus



Van pilot naar definitief examenprogramma

Eindrapport cTWO havo wiskunde A: [...] Domein E, Statistiek en kansrekening, was in de pilot uitsluitend onderdeel van het schoolexamen (SE). Een reden hiervoor was onder meer dat het geïntroduceerde ICT-element [...] niet past bij de wijze waarop het CE wordt afgenomen. Ook toetsing van de empirische onderzoekscyclus lijkt minder goed mogelijk binnen de CE-setting. De pilotdocenten staan echter kritisch tegenover de keuze om statistiek buiten het CE te houden: **de vrees bestaat dat statistiek daardoor in de dagelijkse schoolpraktijk het kind van de rekening wordt**, omdat het CE op school vaak prioriteit krijgt boven het SE [...] Men pleit voor opname in het CE of het stellen van voorwaarden aan SE-toetsing. In een kleine veldraadpleging [...] onderschrijven ook niet-pilotdocenten dit probleem. Om deze redenen stelt cTWO in haar uiteindelijke programma een andere verdeling tussen SE en CE voor dan in de pilotprogramma's.

Intermezzo: vwo wiskunde A en C

- *Eindrapport cTWO vwo wiskunde A:* [...] Pilotdocenten Wiskunde A vwo maken in meerderheid geen bezwaar tegen het niet centraal toetsen van het domein Statistiek en kansrekening. Bovendien is dit domein niet eenvoudig inwisselbaar voor een ander onderwerp dat in het kader van de 60/40-regeling goed tot zijn recht zou komen indien het niet in het CE getoetst zou worden.
- *Eindrapport cTWO vwo wiskunde C:* [...] Docenten gaven [...] aan, geen behoefte te hebben aan toetsing van statistiek binnen het centraal examen Wiskunde C; zij ervaren het CE-programma in de huidige vorm als 'breed genoeg'.

Syllabus nu (vanaf 2017)

Examen HAVO
wiskunde A

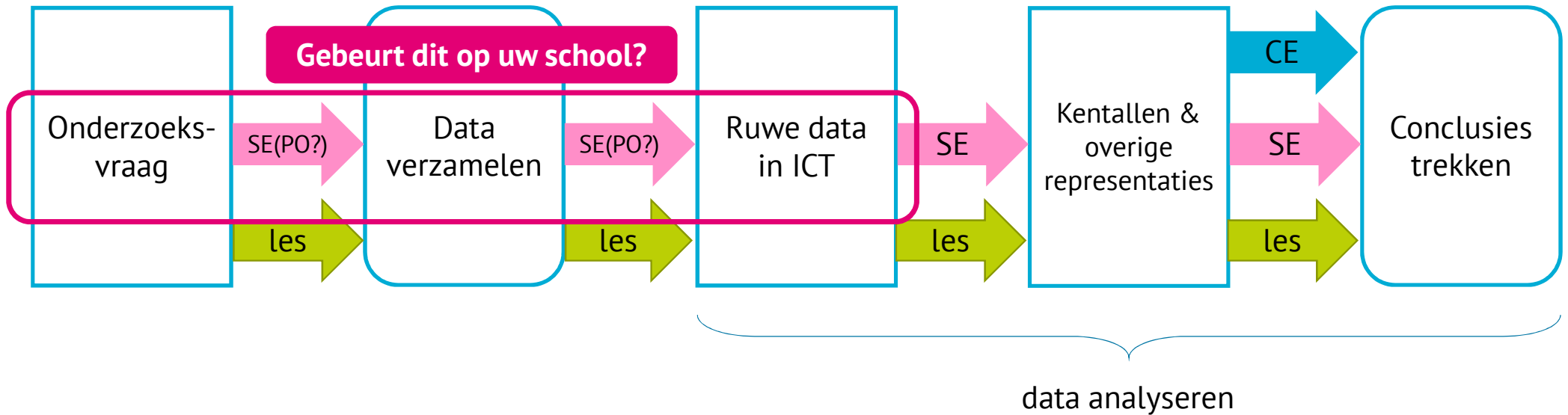
Subdomein E5: Statistiek met ICT

De kandidaat beheerst statistisch ICT-gebruik in relatie met de subdomeinen E1, E2, E3 en E4 om grote datasets te interpreteren en te analyseren [...]

Domeinindeling

Domein	Subdomein	In CE	Moet in SE	Mag in SE
E Statistiek	E1: Presentaties van data interpreteren	x		x
	E2: Data verwerken	x		x
	E3: Data en verdelingen	x		x
	E4: Statistische uitspraken doen	x		x
	E5: Statistiek met ICT			x

Empirische cyclus



Het lijkt eenvoudig

- Een ruwe dataset vinden
- Bewerken (leerlingvriendelijk maken)
- Vragen bedenken

- Lastig: waar vind ik die?
- Kost veel tijd
- Kost veel tijd (met name vragen van het type 'conclusies trekken')

of... (bij geen tijd)

- Een SE uit een proefwerkbundel halen (een niet ongebruikelijk gebruik)
- Oude CE-opgaven gebruiken (een niet ongebruikelijk gebruik)

- Die zijn er niet (toch?)
- Die zijn er niet ('geen' ICT bij CE)

Cito 50 jaar: een traktatie



- Twee datasets
 - COOL (Cohort Onderzoek OnderwijsLoopbanen, 7000 leerlingen havo 5 in 2013)
 - Staandehoudingen (1700 staandehoudingen in Florida 1^e half jaar 2015)
- Twee sets vragen
 - COOL: selectie vragen noodzakelijk
 - Staandehoudingen: selectie vragen niet voor de hand liggend
- Met (in ieder geval) Excel en VuStat uit te voeren
- Correctievoorschrift op basis van VuStat bijgeleverd



Constructie en beschikbaarstelling

- Geconstrueerd in 2018 door (oud-)constructeurs, ervaren in constructie statistiekopgaven CE
- Onder leiding van toetsdeskundigen Cito
- In WiskundE-brief (9 dec '18 na vooraankondiging in sep) berichtgeving: aanmelding via e-mail
- Antwoord (vanaf 10 dec '18) via e-mail met zipbestand met alle bestanden
- Nasturen enquête (vanaf 9 jan '19)
- Uitlekgevaar?



Ontvangst

- Veel interesse voor het materiaal: 280 aanmeldingen (tot 15 januari '19)
- Reacties op enquête: 62 (tot 30 januari '19) [komt straks!]



Inventarisatie onder aanwezigen

Niet ontvangen	
Wel ontvangen, van plan in te zetten	
Wel ontvangen, niet van plan in te zetten	
Wel ontvangen, al ingezet	

VuStat

- VuStat: (gratis) statistische software
- Nog niet erg bekend bij docenten (lijkt het)
- Downloadbare (windows-)versie (werkt met vus-bestanden)
- Online app (werkt met json-bestanden: vus-bestand in windows-versie exporteren naar ascii met TAB als scheidingsteken)
- <https://www.vustat.eu/apps/stat/index.html>



COOL: dataset

- 6976 records/leerlingen
- 21 velden

Data analyse | Bestand Dataset COOL.vus

Bestand Data Beeld Grafiek Tabel Kentallen Resampling Opties ?

Labels

Nr	CIJFER_CSE_NED	CIJFER_CSE_WIA	CIJFER_CSE_WIB	CIJFER_CSE_ENG	CIJFER_SE-GEM	UREN_SCHOOL
1	3.0	**	**	2.0	6.0	35
2	3.2	6.7	**	4.7	**	*
3	3.4	**	**	2.6	**	*
4	3.8	**	**	3.7	6.5	33

UREN_TV	UREN_HUISWERK	UREN_BAANTJE	UREN_VRIENDEN	UREN_SPORT	UREN_INTERNET	UREN_UTGAAN	UREN_COMPSEL	UREN_OVERIG	ZELFSTANDIGHEID
3	3	12	10	4	3	10	1	*	4.066667
*	*	*	*	*	*	*	*	*	**
*	*	*	*	*	*	*	*	*	**
3	4	12	20	3	6	10	0	*	3.588235

COOL: vragen

Leeftijd en geslacht

- 1) Maak een kruistabel van de variabelen LEEFTIJD en GESLACHT
(1p)
- 2) Bereken welk percentage van de leerlingen in de steekproef 18 jaar of ouder is.
(2p)
- 3) Bereken welk percentage van de 17-jarigen een meisje is.
(2p)

Computervaardigheden

Computervaardigheden
procedureel

Computervaardigheden
(procedureel)

COOL: vragen

Syllabus:

Subdomein E2 Data verwerken

3. beargumenteren welke centrummaat (gemiddelde, mediaan of modus) en welke spreidingsmaat (standaardafwijking, interkwartielafstand, spreidingsbreedte) geschikt zijn om een dataset te karakteriseren;

Sommige leerlingen geven met een zeer groot getal aan dat ze heel veel uitgaan, zoals 98 en 90. Dit zijn waarschijnlijk fouten, al dan niet opzettelijk.

13) Vul in de tabel in de tweede kolom bij elke van die drie spreidingsmaten in of die zeer grote getallen bij de variabele UITGAAN invloed op die spreidingsmaat hebben. Omcirkel telkens 'ja' of 'nee'.
(3p)

Concep-
tueel

Naam van de spreidingsmaat	De zeer grote getallen hebben invloed op deze spreidingsmaat	Waarde
	ja / nee	
	ja / nee	
	ja / nee	

UREN UITGAAN
0
5
6
0
5
98
*
*
*
2
10

COOL: vragen

Syllabus:

Subdomein E2 Data verwerken

5. uit gegeven data andere data afleiden [...]

Computer-
vaardigheden

Beeldschermtijd

Sommige ouders vinden dat hun kinderen te veel 'achter een beeldscherm zitten'. In deze opgave vertalen we 'achter een beeldscherm zitten' met 'tv kijken' of 'internet gebruiken' of 'computerspelletjes spelen'.

19) Maak een nieuwe variabele UREN_BEELDSCHERM die de totale tijd per week weergeeft die een leerling besteedt aan tv kijken, internetten en computerspelletjes spelen. Bereken daarmee welk percentage van de leerlingen in de steekproef gemiddeld minimaal 4 uur per dag achter een beeldscherm zit. (4p)

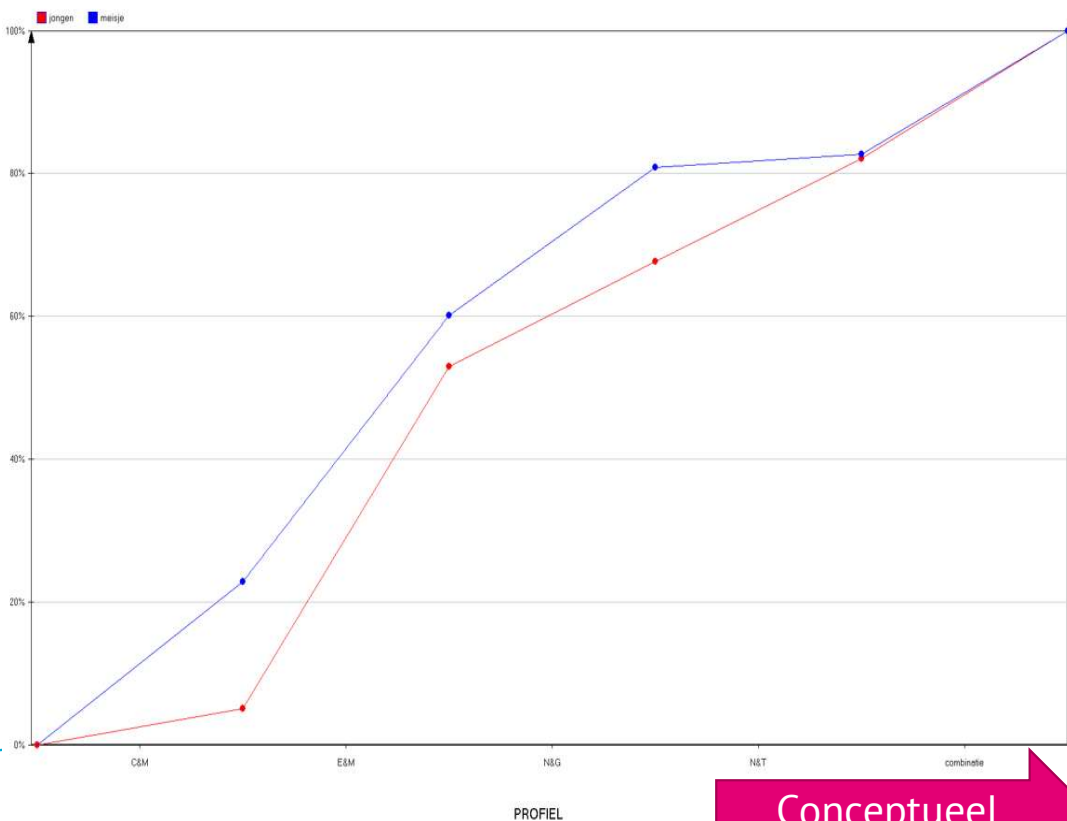
UREN_TV	UREN_INTERNET	UREN_COMPSEL
3	3	1
*	*	*
*	*	*
3	6	0
7	60	5
*	*	*
4	4	3
5	5	*
5	0	45
6	6	0
2	5	0
15	1	21
20	20	20
4	10	0
6	20	0
*	*	*

COOL: vragen

Syllabus:

Subdomein E4 Statistische uitspraken doen

7. conclusies uit statistisch onderzoek met behulp van kwalitatieve en kwantitatieve argumenten kritisch beoordelen, al dan niet in het kader van de empirische cyclus (onderzoeksvraag, data verzamelen, data analyseren, conclusies trekken).



Karin wil onderzoeken hoe groot het verschil in profiel is tussen jongens en meisjes. Om dat te onderzoeken, maakt zij gebruik van $\max V_{cp}$. Eerst maakt zij de volgende relatieve cumulatieve frequentiepolygoon:

In de grafiek kun je aflezen dat $\max V_{cp}$ kleiner is dan 20%. Daarom is het verschil in profiel tussen jongens en meisjes gering, zo is de conclusie van Karin.

8) Leg uit waarom deze aanpak niet correct is. (2p)

COOL: vragen

Subdomein E4 Statistische uitspraken doen

- het verschil tussen groepen kwantificeren
- 1. groepen vergelijken op een gegeven kenmerk;
- 2. aan de hand van een gegeven berekeningswijze het verschil tussen twee groepen kwantificeren;

Computer-vaardigheden,
conceptueel/procedureel



Juan beweert:

“Er is een groot verschil tussen het CSE-cijfer wiskunde A van de leerlingen in de steekproef met een C&M-profiel en het CSE-cijfer wiskunde A van leerlingen in de steekproef met een E&M-profiel.”

17) Onderzoek of Juan gelijk heeft.
(3p)

Staandehoudingen: dataset

- 1740 records/leerlingen
- 22 velden

ID staandehouding	datum	maand	tijd	dagdeel	districtsnaam	ID district	geslacht bestuurder	leeftijd bestuurder
FL-2015-000113	2015-01-01	januari	06:50	ochtend	Palm Beach County	12099	M	64
FL-2015-000299	2015-01-01	januari	09:17	ochtend	Hendry County	12051	V	42
FL-2015-000627	2015-01-01	januari	12:15	middag	Okeechobee County	12093	M	39
FL-2015-000734	2015-01-01	januari	13:18	middag	Leon County	12073	M	62
FL-2015-001064	2015-01-01	januari	16:25	middag	Duval County	12031	V	*
FL-2015-001864	2015-01-02	januari	09:15	ochtend	Hamilton County	12047	M	31
FL-2015-001993	2015-01-02	januari	10:27	ochtend	Alachua County	12001	M	32

doorzoeking uitgevoerd?	type doorzoeking	resultaat staandehouding	gearresteerd?	ID agent
nee	n.v.t.	dagvaarding	nee	3208
nee	n.v.t.	waarschuwing	nee	2542
nee	n.v.t.	dagvaarding	nee	2494
nee	n.v.t.	waarschuwing	nee	1204
nee	n.v.t.	dagvaarding	nee	3372
nee	n.v.t.	dagvaarding	nee	3137
nee	n.v.t.	dagvaarding	nee	2880



Staandehoudingen: vragen

Syllabus:

Subdomein E1 Presentaties van data interpretern en beoordelen

5. beoordelen of een variabele kwalitatief of kwantitatief, discreet of continu, ordinaal of nominaal is;

Definitie

3) Het bestand bevat heel wat variabelen. Vul bij de onderstaande variabelen eerst één van de woorden kwalitatief/kwantitatief in en vul daarna een ander geschikt statistisch woord in.

a) De variabele 'leeftijd bestuurder' is een ... variabele. Deze variabele is een variabele op een ... schaal.

(2p)

b) De variabele 'rang agent' is een ... variabele. Deze variabele is een variabele op een ... schaal

(2p)

Staandehoudingen: vragen

Syllabus:

Subdomein E2 Data en verdelingen

1. verdelingen kwalitatief beschrijven en/of kwalitatief vergelijken, waarbij gebruik gemaakt wordt van klokvormige, meertoppige, uniforme en scheve verdelingen, centrum en spreiding, staarten en uitschieters;

5) Onderzoek of de leeftijden van de bestuurders normaal verdeeld zijn. Geef een geschikte representatie en licht je antwoord toe.
(3p)

← Computervaardigheden
procedureel

Staandehoudingen: vragen

Syllabus:

Subdomein E4 Statistische uitspraken doen

1. groepen vergelijken op een gegeven kenmerk;
2. aan de hand van een gegeven berekeningswijze het verschil tussen twee groepen kwantificeren;

Sommige mensen in de Verenigde Staten denken dat White people bij een staandehouding vaker een waarschuwing krijgen dan Black people.

- 8) Onderzoek met behulp van phi of er sprake is van een gering, middelmatig of groot verschil. Vul daartoe eerst onderstaande tabel in.

	Black	White	totaal
waarschuwing			
geen waarschuwing			
totaal			

(5p)



Computervaardigheden
procedureel

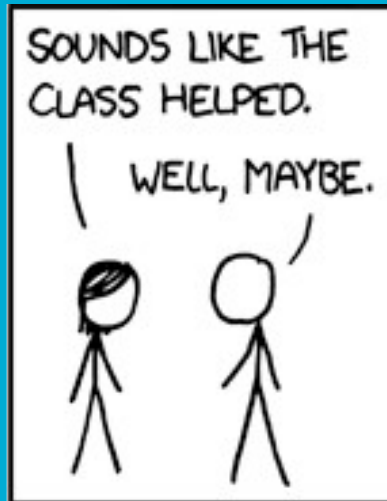
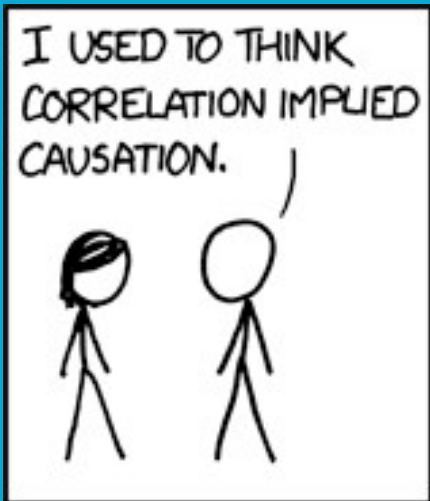
Vragen/discussie

- Wat moet er getoetst worden?
 - Definities
 - Computervaardigheden
 - Procedureel
 - Conceptueel
- Kant en klaar (*Staanhoudingen*) of zelf samenstellen (*COOL*)?
- Meer vragen met groter onderzoekend karakter gewenst?
- Voorziet een SE als dit in een behoefte? Behoeft voor vwo?
- Als Cito (of een andere aanbieder) in de toekomst een soortgelijk product tegen betaling van €300 (per school) zou aanbieden, zou u de school dan adviseren het product aan te schaffen?

Resultaten enquête

- We gaan 'live' naar de resultaten op internet.





Bedankt voor uw aandacht

cito.nl

T (026) 352 11 11

klantenservice@cito.nl

se-statistiek@cito.nl

Ook snel en eenvoudig contact via Facebook en Twitter

