

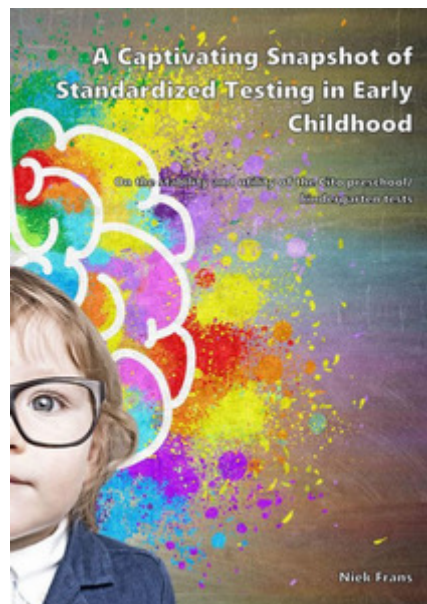
Cito-kleutertests? Een E! 😊

Ewald Vervaeke, lid kerngroep WSK

Op 19 september 2019 is de dertigjarige orthopedagoog Niek Frans gepromoveerd op een proefschrift over de kleutertests van het Cito: *Een boeiende momentopname van gestandaardiseerde testen in de vroege kinderjaren: over de stabiliteit en bruikbaarheid van de Cito-kleutertests*.¹



Niek Frans (geboren in 1989).



Voorkant van Frans' proefschrift.

Wrijving tussen observeren en tests

Frans stelt vast dat er sinds het opgaan in 1985 van de kleuterschool in de basisschool wrijving is ontstaan tussen de gerichtheid van kleuterleerkrachten op speels leren en volgen van de ontwikkeling met observaties aan de ene kant en de gerichtheid in het primaire onderwijs op programmatisch instrueren en toetsen met statistisch genormeerde meerkeuzetests aan de andere kant. Het Cito (Centraal Instituut Toets Ontwikkeling) speelt daarin een hoofdrol.

In 2013 begon de politiek de kant van de kleuterleerkrachten te kiezen, met de [motie die voorheen verplichte kleutertests facultatief maakte](#). Dit leidde tot het [regerakkoord van 2017](#) (p.11) op grond waarvan kleutertests voortaan verboden zouden zijn.

Kleutertests zijn volgens Frans goeddeels ingevoerd om te kijken of kleuters in hun verdere schoolloopbaan wellicht slechte lezers en rekenaars zouden zijn en omdat vroegtijdig ingrijpen bij 'risicokinderen' beter zou zijn dan verhelpen op latere leeftijd.



Cito-kleutertest 'Taal' voor groep 2.



Cito-kleutertest 'Rekenen' voor groep 2.

¹ N. Frans, *A captivating snapshot of standardized testing in early childhood*, Groningen, Rijksuniversiteit Groningen, 2019.

Het testen van kleuters kent echter veel problemen. Om er enkele te noemen:

- bij kleuters zegt één testresultaat niet veel;
- kleuters zijn slecht testbaar omdat ze de bedoeling vaak niet begrijpen en bijvoorbeeld het antwoord aankruisen of omcirkelen, dat hun persoonlijke voorkeur heeft;
- de prestaties van kleuters kunnen erg schommelen, ook vanwege vermoeidheid, wisselende concentratie of gewoon een baaldag;
- kleuters hebben niet allemaal hetzelfde ontwikkelingstempo;
- door de vormgeving en werkwijze van de tests komen altijd veel kinderen in de risicogroep, namelijk de 20% laagstscorders in het I-V-systeem of de 25% laagstscorders in het A-E-systeem.

Cito's 'signaleringen'

Merkwaardig bij dit alles vindt Frans het volgende. Aan de ene kant noemt het Cito zijn tests 'signaleringen van problemen', wat inhoudt dat het Cito ze een voorspellende waarde toedicht zonder ze overigens met zoveel woorden voorspellend te noemen. Aan de andere kant heeft het Cito nooit onderzoek gedaan naar die signalerende/voorspellende waarde. Het zou voor de hand liggen om bij kleuters die nu een D of E halen, na te gaan wat er over 3 of 4 jaar van hun lezen en rekenen is geworden, maar dat doet het Cito dus niet.

5.3.3 Voorspelling per leerling

Let op! 'Voorspelling' is alleen van toepassing op het Leerlingrapport.

Rapport:	Leerlingrapport	Niveaucodering:	Ii/m V	<input type="checkbox"/> Oude normen	Schooljaar:	2014-2015
Weergave:	Grafiek en Tabel	<input checked="" type="checkbox"/> Toon legenda	<input type="checkbox"/> Toon referentieniveaus			
Sortering:	Achtemaam	<input checked="" type="checkbox"/> Voorspelling				

Aan de hand van de bekende toetsresultaten van een leerling en van de onderzoeksgegevens over de ontwikkeling van kinderen maakt het programma een voorspelling voor de toetsresultaten die de leerling het komende afnamemoment(en) zal behalen. Een stippellijn en twee * geven de voorspelde ontwikkelingslijn in de grafiek aan.

Leerlingrapport



Fragment uit de Cito-publicatie [Computerprogramma LOVS; gebruikershandleiding; versie 4.13 \(december 2019, p.78\)](#)², waarin het Cito spreekt over 'voorspelling' in het kader van zijn leerlingvolgsysteem.

Frans heeft daar wel onderzoek naar gedaan. Hij heeft van 431 kinderen naar alle taal- en rekentestcores in de groepen 1-4 gekeken en van 1402 kinderen die scores in de groepen 2-5 gevolgd.

In de groep van 431 kinderen heeft Frans vooral naar de 25% laagstscorders gekeken – vanwege de risico's die men hen doorgaans toedicht: dat waren 143 kinderen op taal en 101 kinderen op rekenen. Hij ontdekte dat slechts 12% consistent D of E blijft scoren op taal en slechts 17% op rekenen. Omgekeerd ging 47% van de 143 kinderen die de eerste keer op taal nog een A- of B-score hadden behaald, in groep 3 naar een D- of E-score. Dat bleek ook het geval te zijn bij 35% van de 101 kinderen die de eerste keer op rekenen nog A of B scoorden. De A- en B-scores bleken dus lang niet altijd een garantie op succes in groep 3 te zijn.

Bij de groep van de 1402 kinderen ontdekte Frans onder meer dat 10,7% van de kinderen voor taal en 12,1% voor rekenen van groep 2 naar groep 5 een groei laat zien, die aanzienlijk afwijkt van die van de overige kinderen. Dat betekent dus dat elk van die kinderen moeilijk te diagnosticeren is aan de hand van één testafname.

Volgens Frans kan men pas na vijf afnames een enigszins zekere uitspraak doen over de vraag of een afwijking stelselmatig is of een toevallige schommeling. Over Cito-kleutertests concludeert hij dan ook dat ze 'niet sensitief genoeg zijn om structurele afwijkingen in de groei te onderscheiden van willekeurige fluctuaties in de scores' (p.129).

Wat denken kleuterleerkrachten er zelf van?

Frans heeft ook aan mensen uit de praktijk gevraagd hoe ze de rol van kleutertests ervaren. Hij deed dat door aan 97 proefpersonen (61 kleuterleerkrachten, 29 IB'ers en 7 anderen) een vragenlijst voor te leggen en vervolgens met 6 van hen een vraaggesprek te houden. Die 6 personen heeft hij geselecteerd op grond van hun antwoorden in de vragenlijst.

Een belangrijke uitkomst uit die vragenlijst is dat de proefpersonen Cito-kleutertests niet als instrumenten zien om de resultaten van het onderwijs mee te beoordelen of te verbeteren, maar ervaren die óf als een afstraffing als de scores laag

² Van internet geplukt op 26 maart 2020.

zijn óf als een bevestiging van de eigen indrukken als de scores gemiddeld of hoog zijn. Velen van hen vinden het ook niet eerlijk om kinderen vragen te stellen over iets dat hen nog niet is aangeboden of gaan alvast op een Cito-kleutertest trainen omdat ze denken dat hun kleuters dat kennelijk zouden moeten kunnen of weten. Daarmee wordt de normering van het Cito feitelijk onderuitgehaald, want door te trainen voor een test ontstaat een vertekend beeld aangezien er er ook scholen zijn die niet trainen.

Het boek bevat 18 passages uit de vraaggesprekken. Hier volgen drie voorbeelden³:

- Tegen tests: ‘Die hokjes van groepsplannen: dat telt niet voor onderwijsinspectie want die kijken niet naar kleuterklassen [...], dus dan denk ik “Ja voor wie doen we dat?”. Voor dat stukje onzekerheid bij onszelf? Voor de ouders die graag een rapportje zien? Terwijl je van KIJK nog een heel mooi rapport kan uitdraaien [...] Want ja, kinderen doe je daar ook geen plezier mee’.
- Tests als bevestiging: ‘Die toetsen zijn heel fijn om te checken of het ook klopt wat ze geleerd hebben, [...] van “Nou, ik verwacht dat dat kind eigenlijk heel goed mee kan en die doet in de les ook altijd goed mee”, geen opvallendheden dus; die zal wel hoog scoren en als je dat dan ook ziet is dat voor de leerkracht wel een bevestiging. Zo wordt het hier op school ook wel gezien’.
- Negatieve ervaringen met kleutertest: ‘Ik ben blij dat we de toetsen niet gedaan hebben, want het gaf ook veel frustratie voor de kleintjes. Die binnenkwamen. En die moesten daar aan een tafeltje gaan zitten. Met een blad voor hun neus [...] dat moeten ze nou in groep 2 wel, maar dan zijn ze al wel weer een stukje verder. In hun ontwikkeling. Ze zijn [...] er meer klaar voor dan in groep 1’.

Alle lof!

Ik vind het een goed en belangrijk proefschrift. Van mij krijgt Niek Frans alle lof voor het feit dat hij heeft gedaan wat het Cito al die tijd heeft nagelaten: het signalerende/voorspellende karakter van zijn kleutertests onderzoeken.

Het onderwijsveld heeft vanaf het begin zijn vraagtekens gehad bij meerkeuzetests en bij het feit dat resultaten altijd slechts afwijkingen van populatiegemiddeldes zijn. Op zijn best stroken ze met je eigen indruk (zoals in de tweede passage hierboven), maar in alle gevallen heb je er niets aan voor de praktijk want er komt niet uit wat je zou kunnen doen om de score hoger te krijgen of hoog te houden.

In mijn bespreking beperk ik me tot drie punten: tests als non-toetsen, de ervaringen van de *Werk- en Steungroep Kleuteronderwijs* (WSK) met het Cito en de uitweg die Frans’ proefschrift blijkt te hebben.

1. Tests als non-toetsen

Hierboven schrijf ik consequent over ‘tests’, terwijl het Cito zelf vrijwel altijd het woord ‘toets’ gebruikt⁴ en ook Frans in zijn samenvatting vaker ‘toets’ zegt dan ‘test’.⁵ Ik doe dat om de volgende reden.⁶

Binnen het toetsgebeuren moeten we volgens mij een onderscheid maken tussen toetsen die op iets *inhoudelijks* zijn gebaseerd en toetsen die *methodologisch* van aard zijn.

Een voorbeeld van een toets met een inhoudelijke basis is de leesproef.⁷

Als Tess bijvoorbeeld TESS, MAMA, PAPA en TOM (broer) heeft geschreven, kan men met haar letters woorden als MOP, PET, POT, POST en STAM maken.

Als ze PET als ‘P, e, t’ (louter hakken) of als ‘P, e, t; step’ (hakken-en-gissen) leest, is ze zeker niet leesrijp.

Als ze PET als ‘P, e, t; pet’ (hakken-en-plakken) of als ‘Pet’ (onmiddellijk lezen) leest, is ze in elk geval plakvaardig en moet je nog verder toetsen of ze ook leesrijp is. De kern van de houdbaar gebleken theorie hierachter is dat ‘P, e, t’ en ‘P, e, t; step’ op *onomkeerbaar* psychologisch functioneren duidt en ‘P, e, t; pet’ of ‘Pet’ op *omkeerbaar* psychologisch functioneren: om de klank /e/ aan de klank /p/ te kunnen plakken moet je in je hoofd met /e/ naar /p/ *teruggaan* en dan weer *vooruit* naar T. De leesproef is dus ingebed in een aangetoonde psychologische theorie over de ontwikkeling van het lezen.

Een voorbeeld van een toets die methodologisch van aard is, is een Cito-kleutertest. In *Kleuter in beeld*, het observatiesysteem dat het Cito al eind 2018 aan het voorbereiden was, staan bijvoorbeeld deze drie vragen:

³ Met dank aan Niek Frans die me de oorspronkelijke woorden deed toekomen.

⁴ In de Cito-publicatie *Computerprogramma LOVS; gebruikershandleiding; versie 4.13* staat het woord ‘test’ vier keer en het woord ‘toets’ 869 keer – afgeleide en samengestelde woorden als ‘testen’, ‘toetsafname’ en dergelijke meegerekend.

⁵ www.rug.nl/research/portal/files/95431764/Samenvatting.pdf; van internet geplukt op 26 maart 2020.

⁶ Zie verder E. Vervaeke, *Basisonderwijs zonder basis*, Rotterdam, Gelling, 2016, p.203-236.

⁷ www.wsk-kleuteronderwijs.nl/wp-content/uploads/2013/11/2013-11-19-bijlage-met-schrijfproef-en-leesproef-in-pdf-formaat.pdf.

Voorbereidend lezen		Aantal goed		
Leest korte, nieuwe woorden met mkm-structuur		<input type="text"/>		
Voorbereidend schrijven				
Schrijft/stempelt eigen naam	<i>Deels goed = beetje</i>	Nee	Beetje	Ja
Voorbereidend schrijven				
Kan eigen naam naschrijven/ nastempelen naar aanleiding van voorbeeld	<i>Deels goed = beetje</i>	Nee	Beetje	Ja

Fragment uit *Kleuter in beeld* van het Cito, onderdeel 'Oriëntatie op geschreven taal'.

Uiteraard mogen die vragen gesteld worden, maar ten eerste: het Cito maakt niet duidelijk wat het lezen van MKM-woorden, schrijven/stempelen van de eigen naam en naschrijven/nastempelen ervan met 'oriëntatie op geschreven taal' te maken heeft en al helemaal niet wat 'oriëntatie op geschreven taal' met leesrijpheid te maken heeft, want óf het Cito suggereert dat *Kleuter in beeld* gebruikt kan worden om na te gaan of een kind aan lezen toe is óf het Cito suggereert dat niet maar onthoudt zich ervan om schoolbesturen, scholen en leerkrachten af te raden om *Kleuter in beeld* te gebruiken om leesrijpheid te bepalen.

Ten tweede: het Cito beoordeelt de antwoorden niet in het kader van een houdbaar gebleken theorie, in dit geval een theorie over de psychologische ontwikkeling van het lezen, maar geeft er punten aan. Bij de drie vragen in het fragment hierboven zullen de punten zoiets zijn als: aantal goed gelezen MKM-woorden; 0/1/2 of 1/2/3; en nog eens 0/1/2 of 1/2/3. Vervolgens gaat het Cito met alle scores – dus ook die drie – rekenen: gemiddelde van de populatie; afwijking van dat gemiddelde per kind; getalsmatige verbanden met andere tests; enzovoort.

Die berekeningen zullen best kloppen, maar dat is de kwestie niet. Bij die puntenverdelingen en berekeningen dienen vragen gesteld te worden als:

- Waarom is het *aantal* goed gelezen woorden van belang en niet de *wijze van lezen* (louter hakken, hakken-en-gissen, hakken-en-plakken en onmiddellijk lezen van de leesproef)?
- Waarom is in het geval van de puntenverdeling 1/2/3 'beetje' twee keer zoveel waard als 'nee', 'ja' drie keer zoveel waard als 'nee' en 'ja' anderhalf keer zoveel waard als 'beetje'?
- Waarom is in het geval van de puntenverdeling 1/2/3 'beetje' bij 'schrijft/stempelt eigen naam een beetje' even veel waard als 'beetje' bij 'schrijft/stempelt eigen naam een beetje na', namelijk allebei 2 punten?

Samenvattend kunnen we zeggen dat er ten minste twee soorten psychologische onderwijsstoetsen⁸ zijn:

- op inhoudelijke inzichten gebaseerde rijpheidstoetsen die houdbaar zijn;
- op methodologische overwegingen gebaseerde toetsen die niet houdbaar zijn.

Om het dit onderscheid ook woordelijk tot uiting te laten komen, gebruik ik voor houdbare toetsen de term 'proef' ('rijpheidsproof') en voor onhoudbare toetsen de term 'test' ('Cito-kleutertest').⁹ Dat sluit ook aan bij het feit dat de methodologie van het Cito dezelfde is als die van de testpsychologie.

Gaandeweg heb ik in mijn verhaal over proeven en tests aangegeven dat tests non-toetsen zijn. Dat valt ook te concluderen uit Frans' proefschrift:

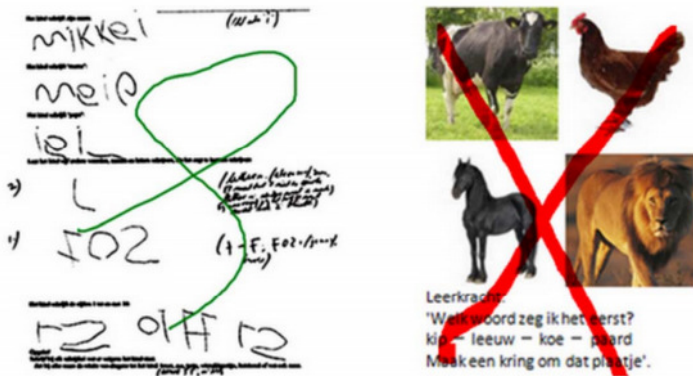
- van het signalerende en voorspellende karakter van Cito-kleutertests blijkt bitter weinig;
- leerkrachten met leerlingen uit moeilijke bevolkingsgroepen voelen zich bestraft en die met leerlingen uit andere bevolkingsgroepen zien hun indrukken hooguit bevestigd (en, wat Frans niet vermeldt, negeren tegenstrijdigheden tussen eigen indrukken en testuitslagen) maar zijn niet echt enthousiast;
- onderwijstests roepen het intraineren op bij kinderen die ergens nog niet aan toe zijn.

Vooraf dat niet-voorspellende acht ik schrikbarend. Grote groepen kinderen die van A/B naar D/E gaan en omgekeerd: iets dergelijks heb ik in mijn onderzoek heb ik iets dergelijks nog nooit meegemaakt. Binnen mijn (ontwikkelingspsychologische en allerm minst statistische) perspectief is het ondenkbaar dat grote groepen kinderen die in groep 2 hakkend-en-plakkend of onmiddellijk lezen in groep 3 weer louter-hakkend of hakkend-en-gissend blijken te lezen.

Cito-kleutertests zijn heel slechte toetsen! De WSK heeft dat duidelijk afgebeeld in een brochure:

⁸ Er is nog een derde soort onderwijsstoets, namelijk een toets om te bepalen in welke mate een kind dat ergens rijp voor is, de hem aangeboden leerstof zich eigen heeft gemaakt. Dat zijn vorderingstoetsen of vorderingsproeven.

⁹ Terwijl ik dit schrijf, zitten we volop in de coronacrisis. Daarin wordt ook vaak van 'tests' gesproken. Daar is op zich niets mee mis, maar het zal duidelijk zijn dat die 'tests' feitelijk proeven zijn in de zin van 'toets die op houdbare inzichten is gebaseerd'.



Proeven: ja; tests: nee.¹⁰

Ook ik heb onderzoek gedaan naar Cito-kleutertests en mijn oordeel erover wordt door Frans' onderzoek bevestigd. Ik heb de leesrijpheidstoets (dit is leesproef plus schrijfproef) en de Cito-test *Taal voor kleuters* met elkaar vergeleken.¹¹ Deze test blijkt geen enkel verband te hebben met leesrijpheid. Dit komt onder meer omdat het Cito lezen en schrijven als taalvaardigheden opvat terwijl alleen spreken en luisteren taalvaardigheden zijn - lezen en schrijven zijn (de)coderingsvaardigheden.

Tests zijn non-toetsen zijn onhoudbaar en moeten onmiddellijk van het toneel verdwijnen.

Ik geef het Cito voor zijn kleutertests een E! 😊

2. Ervaring van de WSK met het Cito

Naar aanleiding van het [normeringsonderzoek van het Cito voor de kleutertoets/kleuterobservatie-instrument *Kleuter in beeld – Taal*](#) heeft de WSK het Cito deze vragen toegestuurd:

- 'Wilt u ons uw definities van de begrippen "toets", "observatie-instrument" en "normering" laten weten?'.¹²
- 'We zien uw wetenschappelijk onderbouwde reactie op deze vier punten graag tegemoet'.¹²
- 'We zouden graag van u de verzekering krijgen dat er in het voortraject op geen enkele manier psychometrische normering is verricht; dat de door u bedoelde validatie louter is gebaseerd op de psychologische ontwikkeling van het kind'.¹³

Het Cito wilde deze vragen niet schriftelijk maar wel mondeling beantwoorden.¹⁴ Mede op aandringen van het Ministerie van OCW is de WSK op 11 juni 2019 het gesprek aangegaan, maar daarin heeft het Cito (ook volgens henzelf) geen antwoord op onze vragen gegeven. Enzovoort.¹⁵

Het eind van het liedje is dat het Cito – zonder zelfs maar een begin van een antwoord op een van de drie vragen gemaakt te hebben – op 7 november 2019 doodleuk schrijft: 'Staan er volgens u nog vragen open waarop Cito geen of geen volledige reactie heeft gegeven, dan verneem ik dat graag om deze alsnog te realiseren'.¹⁶

De WSK heeft de regering en de Tweede Kamer laten weten wat zijn vermoeden is voor Cito's handelwijze: '*het Cito zou moeten erkennen dat hun nieuwe kleutertoetsen wel degelijk inferentieel-statistisch genormeerd zijn en geen volledige ontwikkelingspsychologische inhoud hebben en dat die nieuwe kleutertoets wel degelijk louter getalsmatig en allerminst objectief is, maar het Cito draait hier om andere dan wetenschappelijke en maatschappelijke redenen omheen*'.¹⁷

Sedert de motie van 2013 is het Cito erin geslaagd om de zogeheten Expertgroep aan het Ministerie van OCW te koppelen, waardoor alle toetsen in een leerlingvolgsysteem tests moeten zijn.¹⁸ In reactie hierop heeft de Tweede Kamer de Expertgroep uit kleutertoetsen gehaald door het aannemen van een amendement.¹⁹

3. Uitweg: half gestructureerde onderzoeksvragen

Dagblad *Trouw* plaatste op 6 januari 2020 een [vraaggesprek met Frans](#). Daarin laat hij zich ook enigszins positief uit over tests: 'Je kunt de toets gebruiken als grove inschatting van niveau, zolang je maar begrijpt dat deze inschatting nogal kan wisselen per toets'. Dat grove kan ik niet bevestigen in mijn onderzoek naar de Cito-taalttest: van de veertien kinderen is er maar één leesrijp, maar die heeft geen A op de Cito-test *Taal voor kleuters*, maar een C, dus onder gemiddeld. Er is nog een kind met een C en dat is volstrekt niet leesrijp: het kan zijn naam niet schrijven en denkt dat van mijn naam 'd' de eerste letter is en 'E' de laatste. Dit doet me denken aan een thermometer die zowel bij -30°C als bij +30°C 0°C aangeeft...

¹⁰ WSK-brochure [Minimale competenties voor de beginnende kleuterleerkracht](#), 18 april 2015, p.6.

¹¹ Zie noot 6 en de [samenvatting](#) van een artikel uit 2014 dat geantdateerd in 2013 is verschenen: 'Leesrijpheid: leesrijpheidstoets versus Citotoets "Taal voor kleuters"', *Struktur en genese*, 2013, vol.26, p.25-50.

¹² [Brief van de WSK aan het Cito, 9 november 2018](#).

¹³ [Brief van de WSK aan het Cito, 28 november 2018](#).

¹⁴ [Brief van het Cito aan de WSK, 14 november](#); [brief van het Cito aan de WSK, 10 december 2018](#).

¹⁵ De verdere e-correspondentie is te volgen op www.wsk-kleuteronderwijs.nl/downloads/; zoek op 'cito'.

¹⁶ [Brief van het Cito aan de WSK, 7 november 2019](#).

¹⁷ [Brief van de WSK aan de regering en de Tweede Kamer, 12 december 2019](#).

¹⁸ [Brief van de WSK aan de regering en de Tweede Kamer, 12 december 2019](#), zoek op 'expertgroep'.

¹⁹ Het [amendement](#) is op 4 februari 2020 door de Tweede Kamer aangenomen.

Nogmaals: tests zijn non-toetsen en kunnen en zullen uit de onderwijspraktijk verdwijnen. Maar hoe kunnen kinderen dan wel zinvol getoetst worden? Frans geeft volgens mij in zijn vraaggesprekken met zes proefpersonen de uitweg uit het testdrijfzand: half gestructureerde vragen. Daarin bepaalt een onderzoeker zijn doel en stemt hij daar de vraag of vragen op af, die hij zeker wil stellen, maar voor het overige past hij andere vragen aan aan de reacties van elke proefpersoon afzonderlijk. Als men dit goed doet, kan men bij kinderen de resultaten altijd in het kader van een theorie duiden, zoals 'onomkeerbaar' of 'omkeerbaar' in de leesproef. Immers, aan Durk die DURK, MAMA, PAPA en GINA (zus) heeft geschreven, laat men niet PET lezen zoals bij Tess, maar wel GUM, dat men dan weer niet aan Tess laat lezen.

Wat zou het Cito van Frans' proefschrift vinden? *Trouw* aan het slot: 'Cito was niet bereikbaar voor een reactie'... Het Cito laat zich graag op 'zeker weten' voorstaan (zie onderstaand logo), maar weet het Cito het zelf wel zo zeker?

