

# TOETSEN EN TOETSVRAGEN ANALYSEREN

E-merge  
Digitaal toetsen  
2 november 2016

# INTRODUCTIE



Sander Schenk

Hogeschool Rotterdam sinds 1999

Instituut voor Financieel Management

Docent – manager – beleidsadviseur

Rode draad: digitaal toetsen

Toets & Leer

*N.B. Stel gerust vragen tussendoor*

# JARGON – WAT HOORT WAARBIJ?

## Lijst 1

Item

Cesuur

Sleutel

Toetsscore

Afleider

Betrouwbaarheid

Validiteit

## Lijst 2

Test – hertest

Fout antwoord

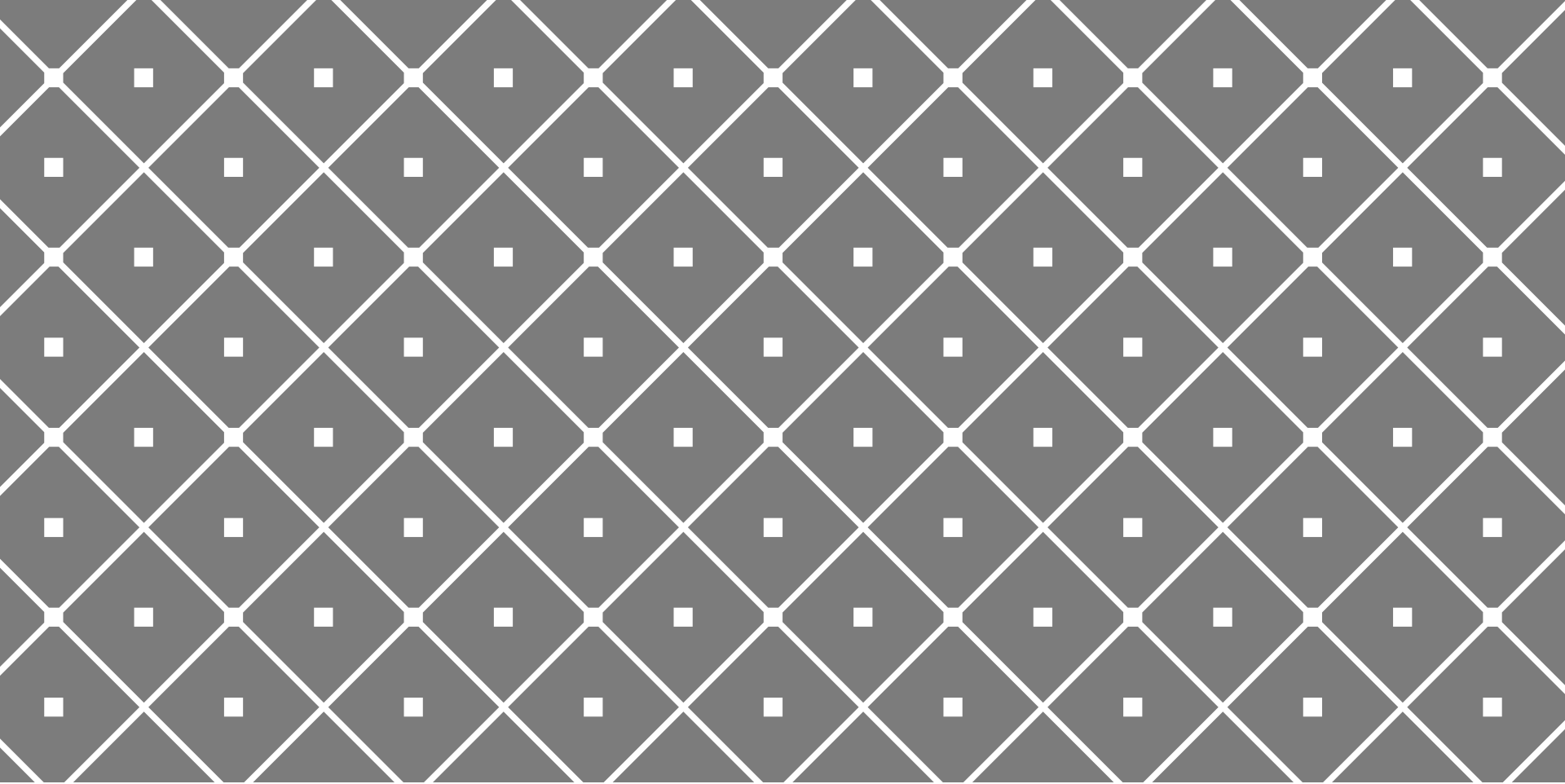
Punten

Vraag

Zak-slaaggrens

Toetsmatrijs

Juiste antwoord



# WAAROM TOETSANALYSE?

Vraag 1 van 6

# DOELEN TOETS- EN ITEMANALYSE 1

Kwaliteit van toets controleren

Niveau borgen

Verkeerde beslissingen voorkomen

Volgende toetsen verbeteren

Aansluiting op gegeven onderwijs beoordelen

# DOELEN TOETS- EN ITEMANALYSE 2

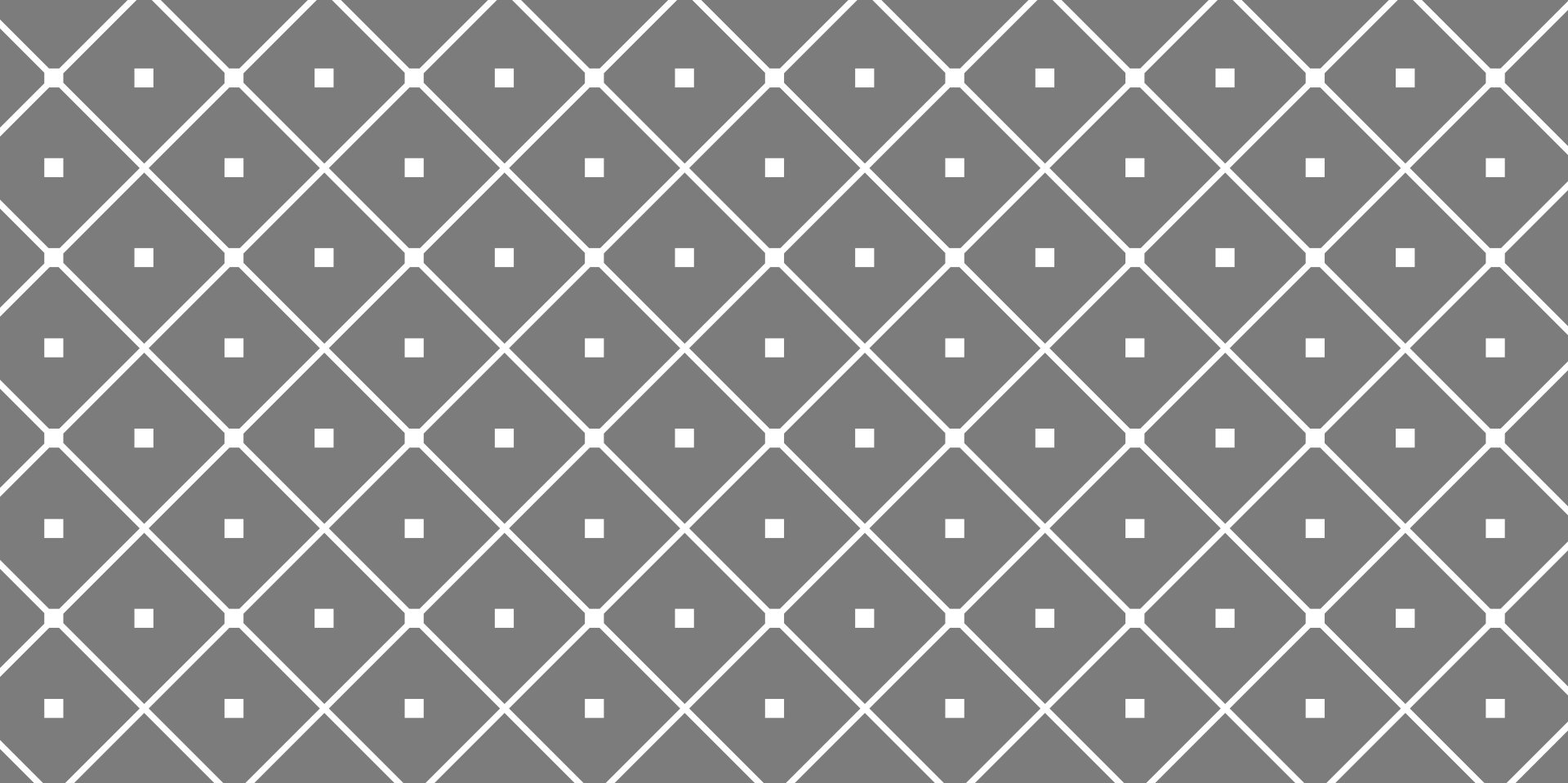
Sleutelfouten herstellen

Corrigeren voor moeilijkheid toets

Moeilijke en makkelijke vragen identificeren

Weinig onderscheidende vragen herkennen (kaf en koren)

Kwaliteit van afleiders beoordelen



# WELKE ANALYSETOOL?

Vraag 2 van 6

# HULPMIDDELEN BIJ TOETSANALYSE

## **Analysemodule van eigen digitaal toetssysteem**

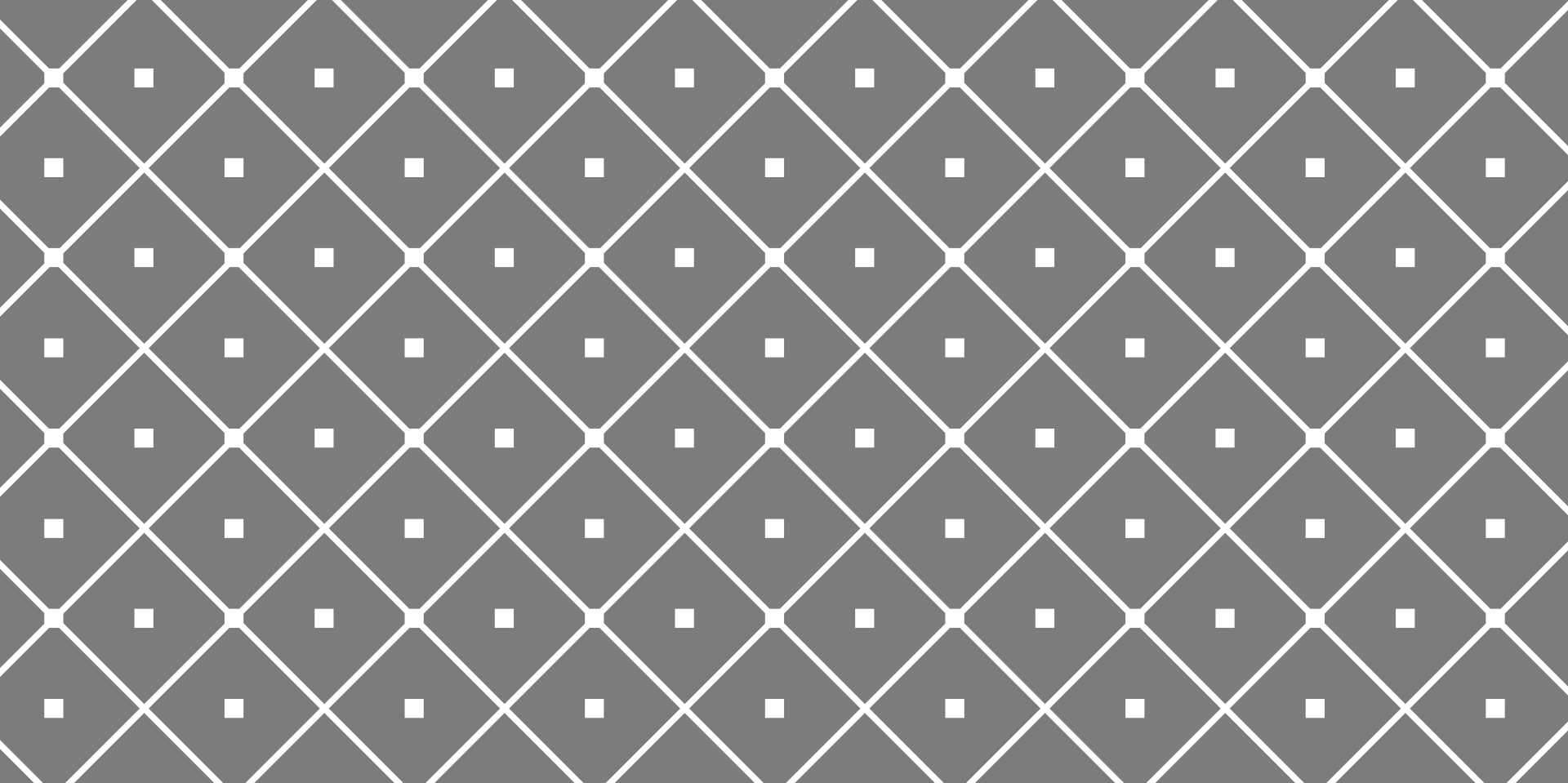
Mogelijkheden verschillen, maar meestal prima bruikbaar. Vergt investering om pakketeigen jargon en opties te leren kennen.

**ConTest (Teleform)** Uitgebreide rapportages. Vooral voor digitale analyse van op papier afgenomen meerkeuzetoetsen.

**qDNA** Algemene en gratis hulptool (zie [qdnatool.org](http://qdnatool.org)) die helpt bij het maken en interpreteren van een toets- en itemanalyse

**Excel / SPSS / R** Voor de doe-het-zelver, vergt veel psychometrische kennis en vaardigheid





# WIE IS VERANTWOORDELIJK?

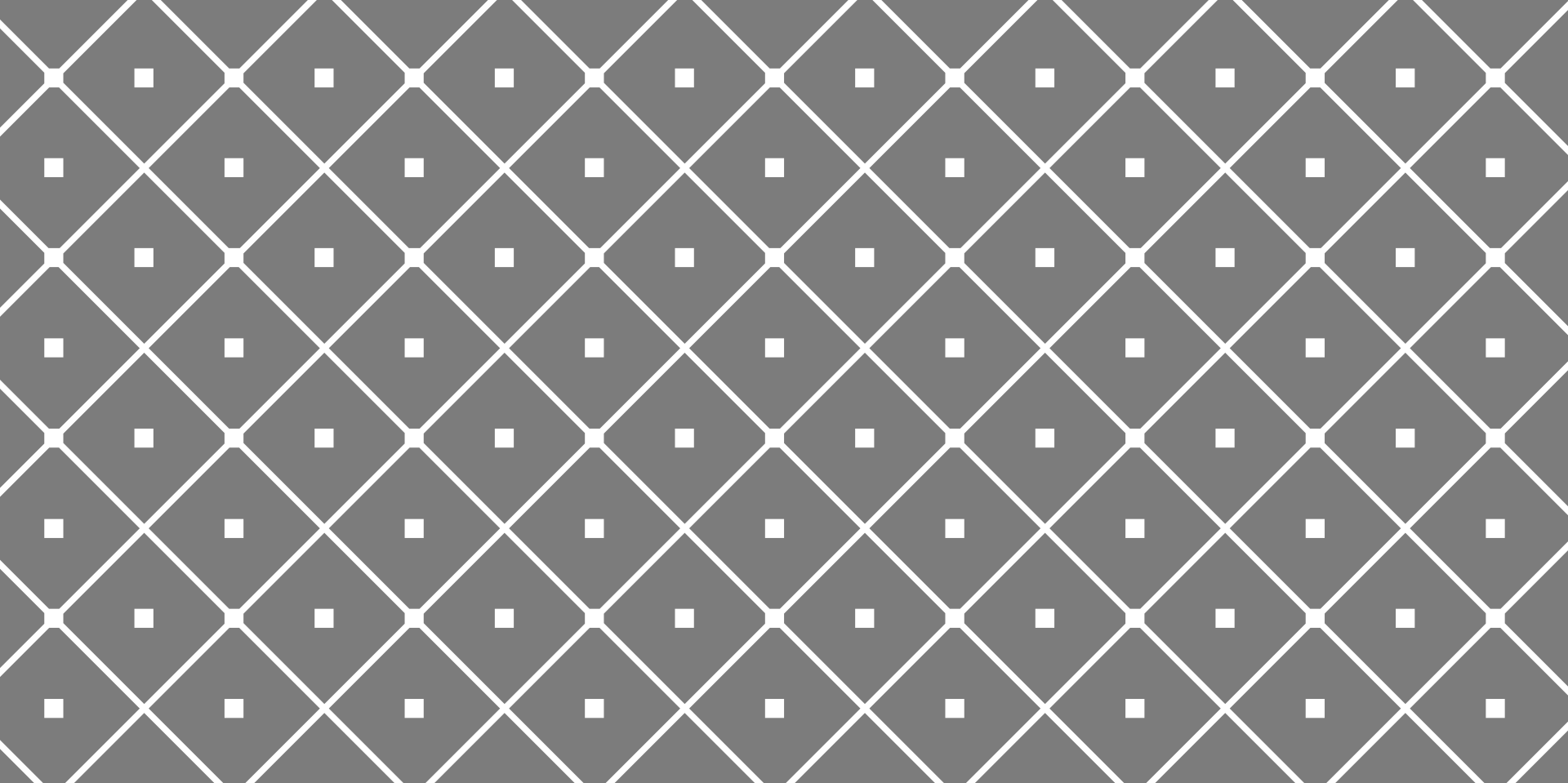
Vraag 3 van 6

# WAT IEDERE DOCENT MOET KUNNEN

De **kwaliteit** van een toets en toetsvragen (items) **beoordelen** op basis van een toets- en itemanalyse.

**Acties definiëren** die op basis van een toets- en itemanalyse genomen moeten worden.

Naar behoefte **om ondersteuning vragen** bij een toetsdeskundige/psychometrist.



# WANNEER ANALYSEER JE?

Vraag 4 van 6

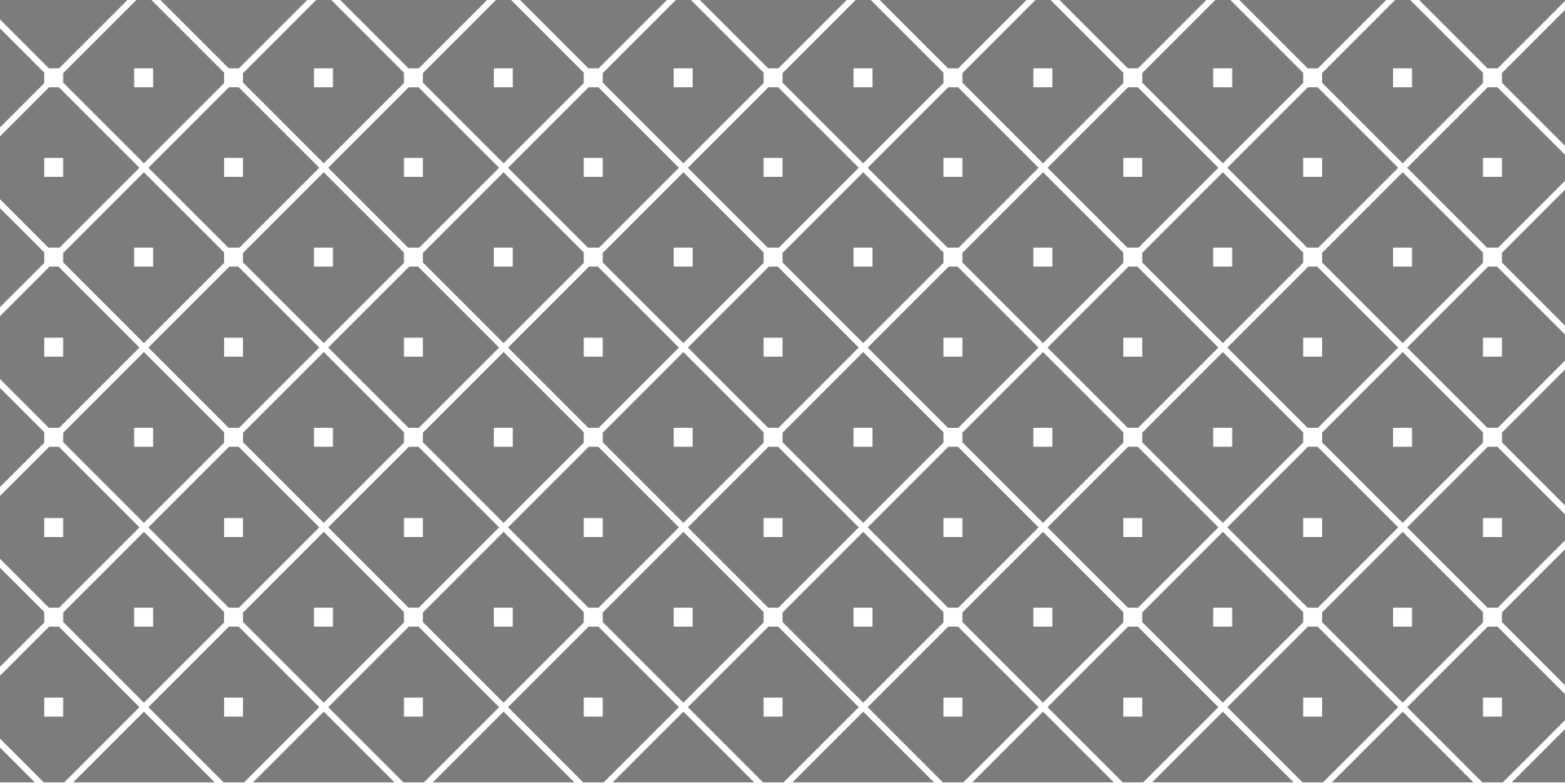
# WANNEER TOETS- EN ITEMANALYSE

Zo snel mogelijk na toetsafname

Vóór invoeren cijfers in studentinformatiesysteem

Vóór inzage studenten

Na eventuele aanpassingen opnieuw TIA uitvoeren



# WAT ANALYSEER JE?

Vraag 5 van 6

# WAAR OP LETTEN ?

Aantal deelnemers toets

Eerste kans of herkansing?

Frequentieverdeling van scores

Statistieken op toetsniveau

Statistieken op vraagniveau

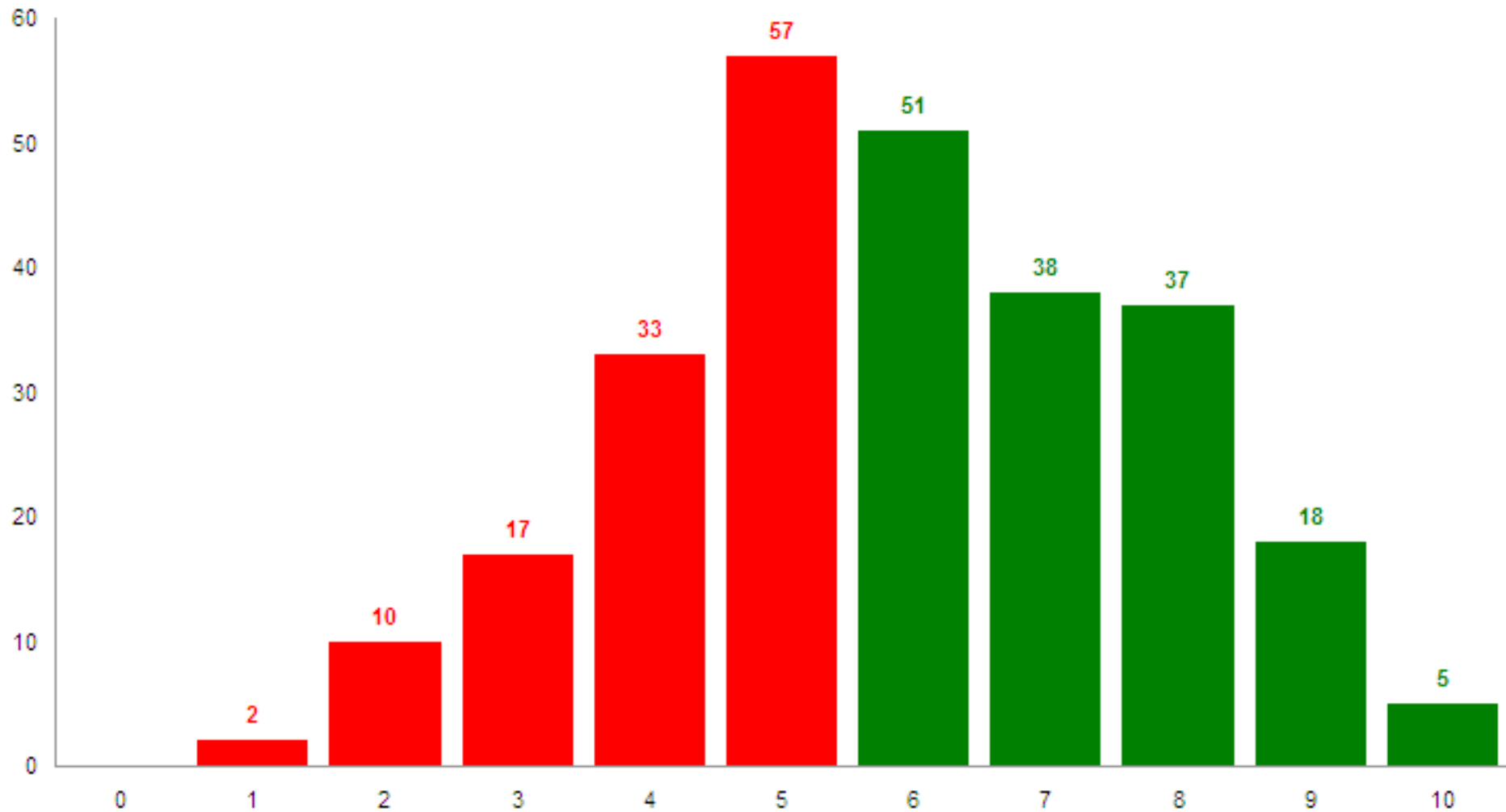
Proportie goed ( $p$ )

Onderscheidend vermogen ( $rit$ )

Waarde van afleiders ( $f$  en  $z$ )

Toetsopgaven erbij!

# FREQUENTIEVERDELING



# GEMIDDELDE MOEILIKHEID TOETS

De gemiddelde moeilijkheid van een toets komt uit op 0,69.

Was dit een makkelijke of een moeilijke toets?



# BETROUWBAARHEID

Een summatieve toets heeft een Cronbach's alfa van 0,7

Is dit genoeg om aan de toets conclusies te verbinden?



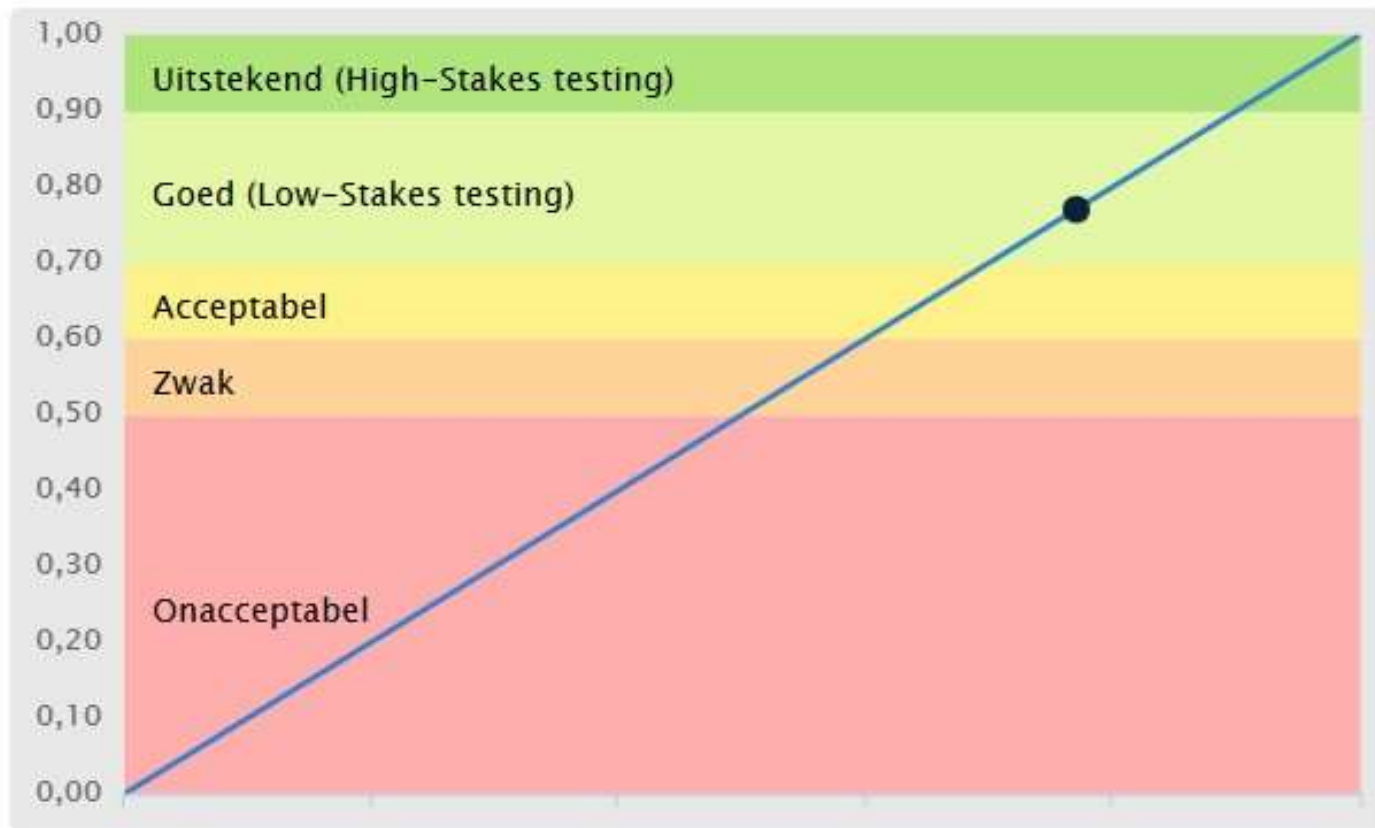
# CRONBACH'S ALFA

## **Cronbach's alfa**

Cronbach's alfa kan worden gebruikt als schatting voor de betrouwbaarheid van een toets. RemindoToets berekent deze bij toetsmatrijzen die ten minste één keer zijn afgenomen met dezelfde (sub)set van vragen. Deze analyse is voornamelijk van toepassing op toetsmatrijzen met een vaste set vragen.

# CRONBACH'S ALFA

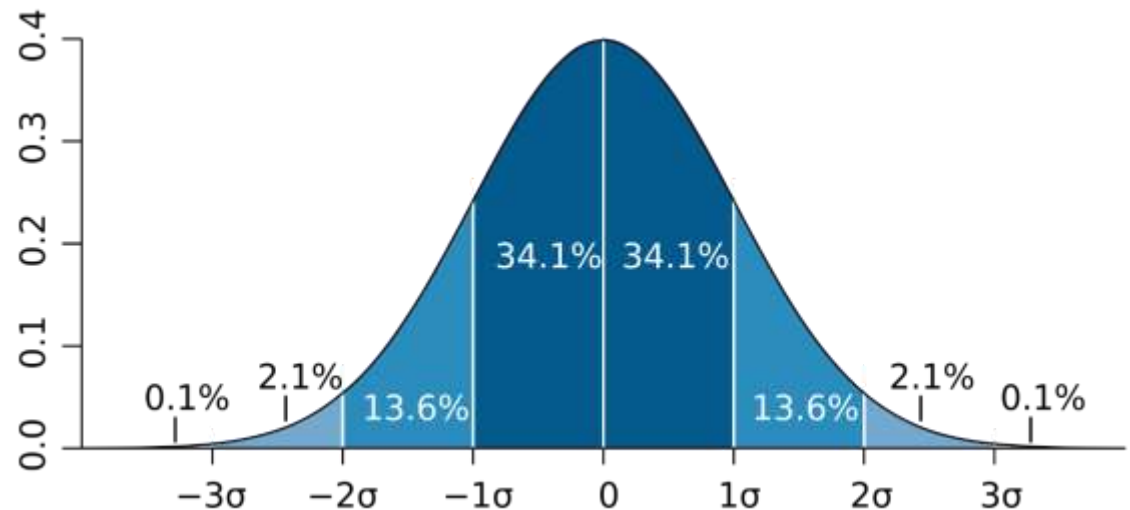
✓ Cronbach's alfa gebaseerd op 30 vragen en 176 toetsafnames is 0,77.



# STANDAARDMEETFOUT

Bij een toets waarbij de cesuur op 26 punten ligt, is de standaardmeetfout 7,14.

Kun je met 95% zekerheid zeggen dat een student die 12 punten scoort terecht is gezakt?



# MISCLASSIFICATIES

Tabel 3: Percentage foute zak/slaag beslissingen, gegeven de toetsbetrouwbaarheid en het zakpercentage  
(Uit: Dousma e.a., 1997)

Zak %	Betrouwbaarheid											
	0,0	0,50	0,55	0,60	0,65	0,70	0,75	0,80	0,85	0,90	0,95	1.00
5	10	8	7	7	7	6	6	5	4	4	3	0
10	18	14	13	12	11	11	10	9	8	6	4	0
15	26	18	18	17	16	14	13	12	10	8	6	0
20	32	23	21	20	19	17	16	14	12	10	7	0
25	38	26	25	23	22	20	18	16	14	11	8	0
30	42	29	27	25	24	22	20	18	15	12	9	0
35	46	31	29	27	25	23	21	19	16	13	9	0
40	48	32	30	29	27	24	22	20	17	14	10	0
45	50	33	31	29	27	25	23	20	18	14	10	0
50	50	33	31	30	28	25	23	20	18	14	10	0

# P-WAARDE

Een toetsvraag heeft een p-waarde van 0,75.

Kun je stellen dat dit een makkelijke vraag was?

# TE HOGE P-WAARDE?


## Vraag









Wat is **geen** wettelijke kenmerk van een overeenkomst volgens art. 6:213 BW?

- (A) een meerzijdige rechtshandeling
- (B) een of meer partijen
- (C) betaling
- (D) aangaan van een verbintenis

1 pt. ✓

## Gegeven antwoorden

 Pas antwoordmodel aan

Antwoord	Score	Aantal antwoorden
 C	1	 163
 B	0	 6
 A	0	 4
 D	0	 3

# TE LAGE P-WAARDE?

94	0,00		0,00	0,00	 
95	0,83				
95	0,72				
95	0,75		0,17	0,43	 

De P-waarde van deze vraag is lager dan 0,4, wat betekent dat minder dan 40% van de kandidaten in staat was deze vraag correct te beantwoorden. Wellicht was deze vraag te moeilijk?

























# LAGE P-WAARDE ONDERZOEKEN

Wat kan Ali maximaal lenen?

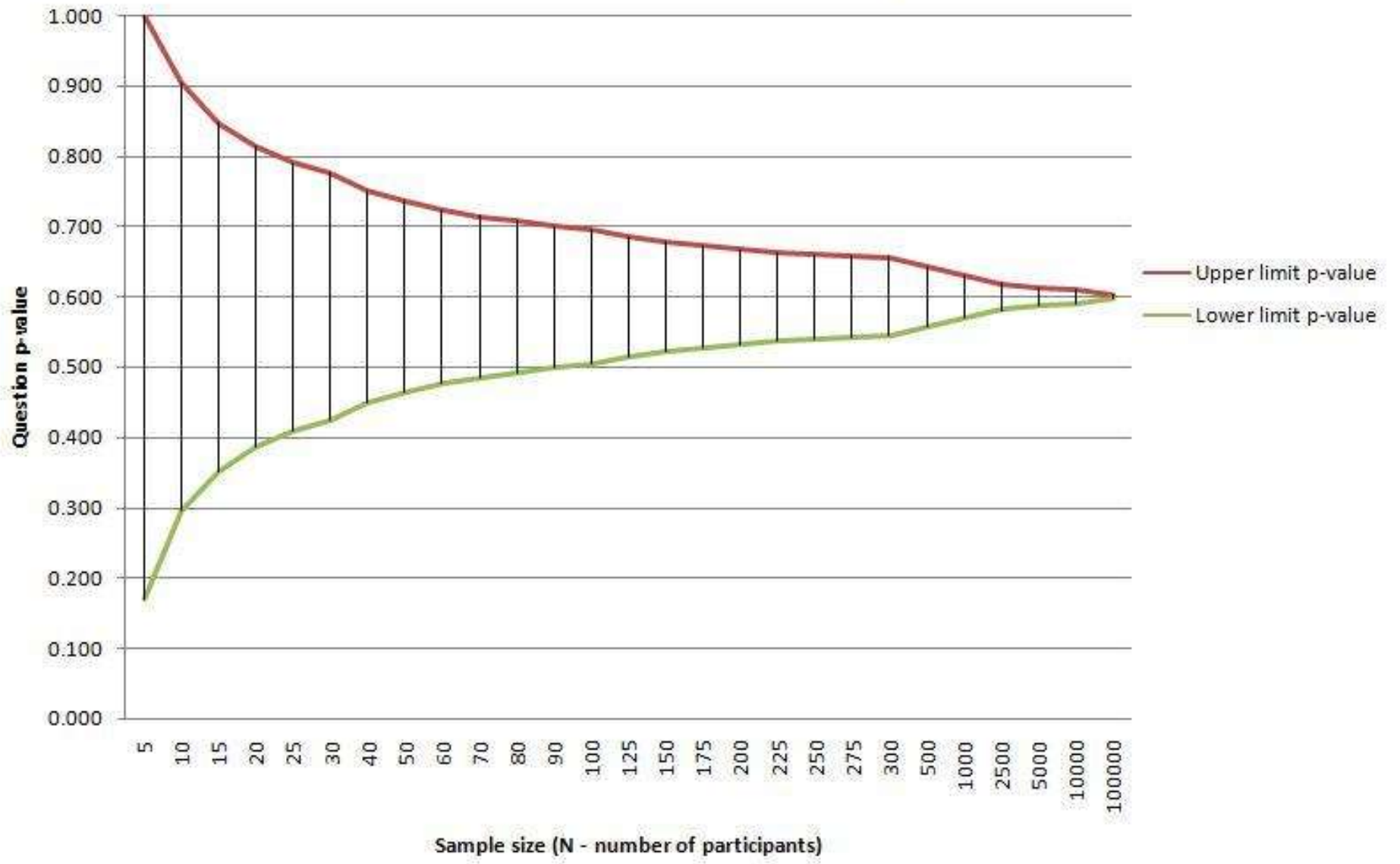
€  11.225 1 pt.

Gegeven antwoorden

 Pas antwoordmodel aan

Antwoord	Score	Aantal antwoorden
 11.212,5	0	 25
 1.837,5	0	 16
 35.837,5	0	 6
 913,25	0	 6
 0	0	 4
 41.837,5	0	 3
 45.662,5	0	 3
 29.962,5	0	 2
 45.212,5	0	 1
 11.187,5	0	 1
 4.485	0	 1

# How sample size influences the stability of p-values (observed p-value of 0.6 used as example)



# A-WAARDEN

Een afleider bij een meerkeuzevraag met vier alternatieven heeft een A-waarde van 0,02.

Wat zegt dit over de kwaliteit van de afleider?

# KWALITEIT VAN AFLEIDERS 1

		Volledig overzicht	Details van het resultaat	Analyseer vragen			
		Resultaten	P	$R_{ir}$	STD		
en klassieke benadering 52	versie 2	151	0,47	0,40	0,50	 	
en klassieke benadering 51	versie 2	152	0,66	0,40	0,47	 	
en klassieke benadering 49	versie 2	150	0,79	0,32	0,41	 	
en klassieke benadering 50	versie 2	151	0,75	0,15	0,43	 	

P-waarde en  $R_{ir}$ -waarde van geel gearceerde vraag ziet er prima uit.

Niks meer aan doen dus?

# KWALITEIT VAN AFLEIDERS 2

Rule 4	
Code	IFM.MAN.01.04.02_XN0003 XD-Kritische kanttekeningen klassieke benadering 51
Question ID	1001
Version number	2
Interaction number	
Aantal resultaten	152
STDEV-waarde	0,476
P-waarde	0,658
RIT-waarde	0,487
RIR-waarde	0,393
A-waarde (A)	0,171
A-waarde (B)	0
A-waarde (C)	0,658
A-waarde (D)	0,171

Dezelfde vraag.

Antwoord B blijkt een A-waarde van 0 te hebben.

Die afleider heeft dus niet 'afgeleid'

# KWALITEIT VAN AFLEIDERS 3

Antwoordanalyse voor vraag IFM.MAN.01.04.02\_XN0003 XD-Kritische kantf...



## Vraag

Volgens de nieuwe of de moderne benadering van strategisch management is het belangrijkste van strategisch management dat?

- (A) Strategisch management synoniem is aan strategische planning
- (B) Strategisch management alleen voor de cash cows van een organisatie belangrijk is
- (C) Strategisch management synoniem aan strategisch denken
- (D) Een organisatie op zoek moet gaan naar het evenwicht tussen de middelen van de organisatie, de sterktes, de zwaktes, de kansen en de bedreigingen

1 pt. ✓

## Gegeven antwoorden

Pas antwoordmodel aan

Antwoord	Score	Aantal antwoorden
✓ C	1	100
✗ D	0	26
✗ A	0	26

Afsluiten

# RIT- EN RIR-WAARDEN

Je komt in een toetsanalyse een vraag tegen met een Rit-waarde van  $-0,4$ .

Moet je deze vraag verwijderen uit de toets?

# NEGATIEVE RIT/RIR

81	0,21		-0,08		0,41	 
81	0,85	De R <sub>ir</sub> -waarde van deze vraag is negatief, wat betekent dat er een negatieve correlatie is tussen het goed beantwoorden van de vraag en het resultaat op de toets.				
81	1,00					

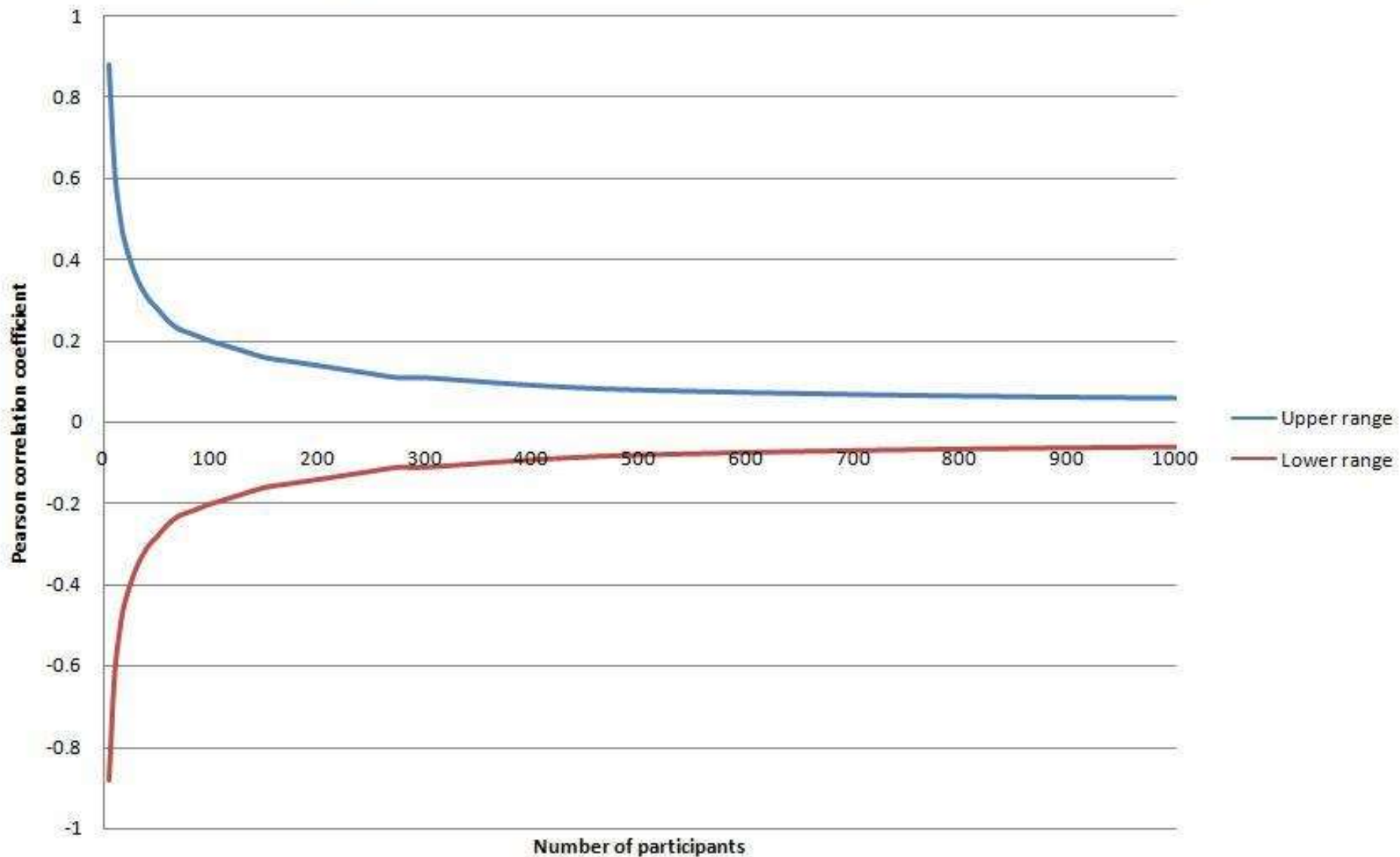
Een negatieve Rit/Rir-waarde moet altijd onderzocht worden.

Mogelijke verklaringen?

- Goede studenten op verkeerde been gezet (te ver doorgedacht)
- Sleutelfout?
- Vraag te weinig gesteld in toets



## How sample size influences the stability of Pearson correlation coefficients (observed item total correlation of 0 used as example)




# NEGATIEVE RIT/RIR

## Vraag


Stelling I: misdrijven hebben in tegenstelling tot overtredingen een sanctie in de vorm van een gevangenisstraf







Stelling II: doen plegen is een vorm van medeplichtigheid en/of OVAR

- (A) alleen stelling I is juist
- (B) alleen stelling II is juist
- (C) beide stellingen zijn onjuist

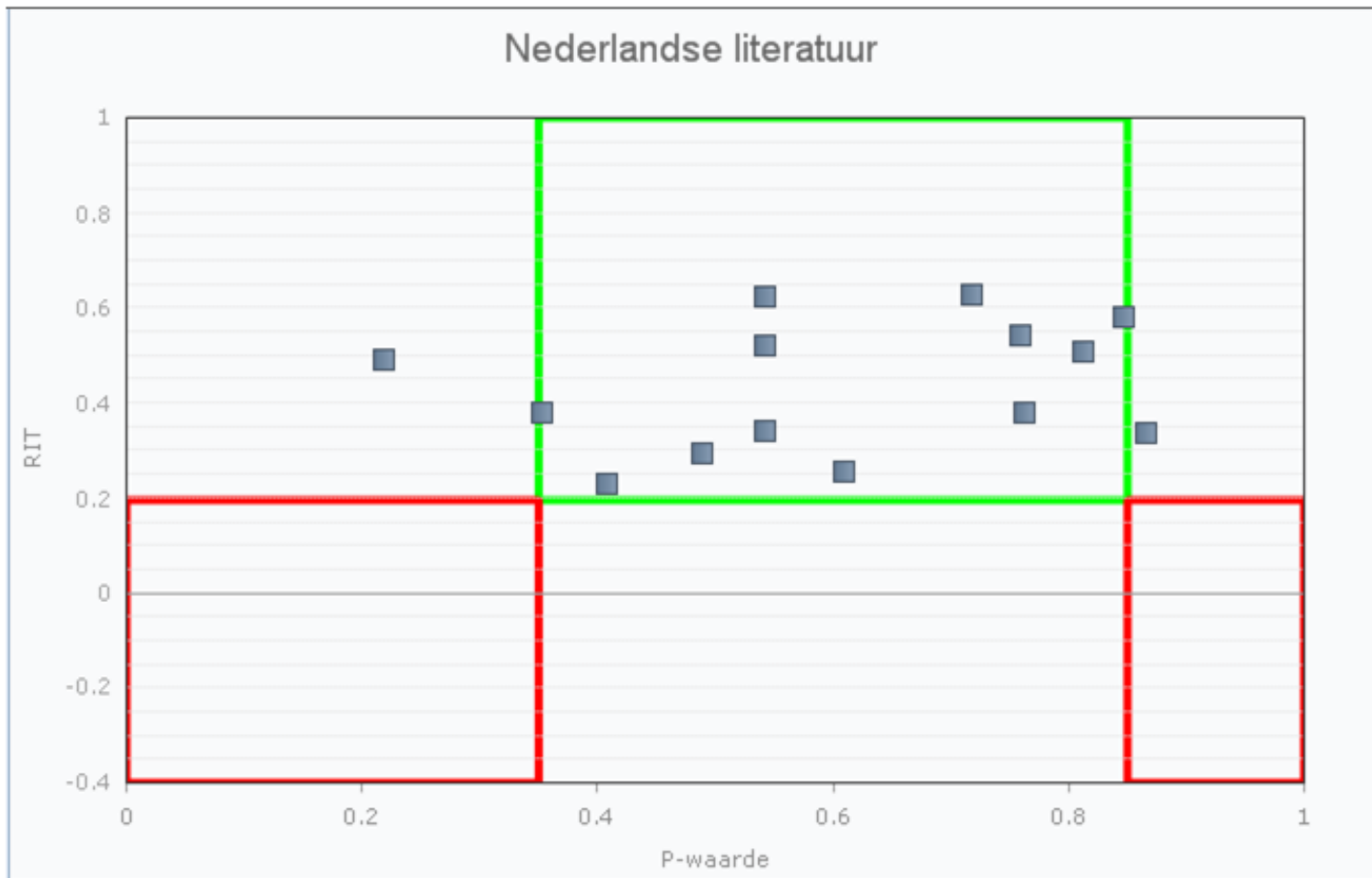
1 pt. 

## Gegeven antwoorden

 Pas antwoordmodel aan

Antwoord	Score	Aantal antwoorden
 C	0	 48
 A	1	 17
 B	0	 16

# DE IDEALE VRAGEN?



*P*-waarde

Moeilijkheid toetsvraag



Coëfficiënt *alfa*

Betrouwbaarheid toets



*A*-waarde

Kwaliteit afleiders



Meetfout

Betrouwbaarheid toets

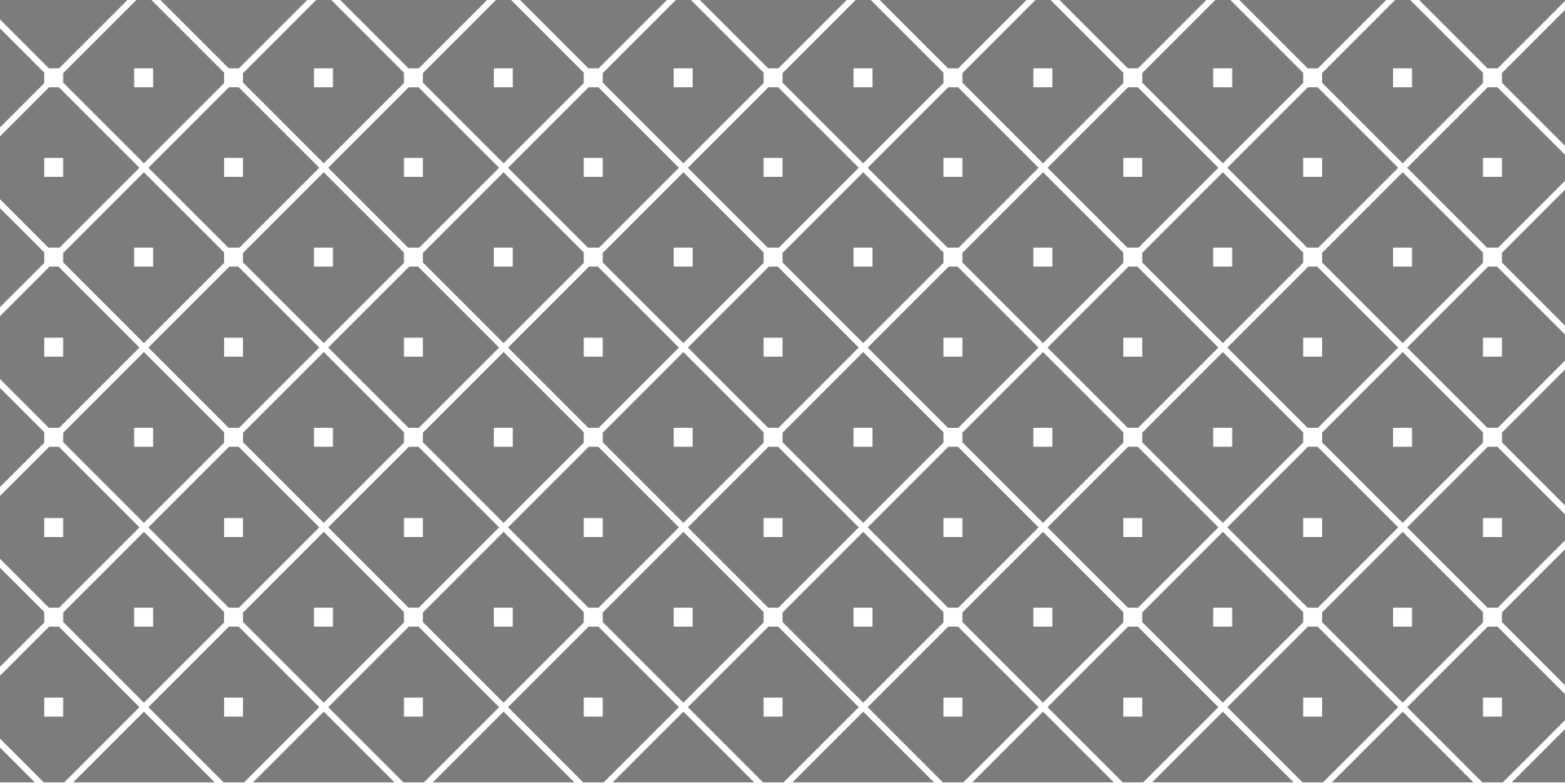


*R<sub>it</sub>*-waarde

Onderscheidingsvermogen vraag



Altiid checken!



**WELKE ACTIE ONDERNEEM JE?**

Vraag 6 van 6

# TOETS GEANALYSEERD, WAT NU?

## Direct

Vragen laten vervallen?

Vragen bij iedereen goed rekenen?

Extra antwoord goed rekenen?

Cesuur aanpassen?

## Later

Vragen verbeteren

Meer vragen in toets opnemen?

Goede vragen als basis voor volgende toets gebruiken

# EXTRA ANTWOORD GOED REKENEN

[actief] Antwoordmodel aangepast voor het toetsmoment "FSMCKR2114"     
door Sander Schenk op 1 nov. 2016 15:29:12 (motivatie: Antwoorden €  
11.212,50 en € 11.225,00 worden beide goed gerekend. Het verschil is een gevolg van  
het al dan niet afronden van tussenliggende bedragen in de berekeningen.)

# VRAGEN LATEN VERVALLEN?

<i>Beslismatrix</i> <i>UMCG</i>	$p > 0.85$	$0.85 < p <$ raadkans	$p <$ raadkans
$R_{it} > 0$	✓	✓	✗
$R_{it} = 0$	✓	discussie	✗
$R_{it} < 0$	✓	✗	✗



# CESUUR AANPASSEN?

Cohen-Schotanus:

Een toets waarbij niet wordt gecorrigeerd voor de moeilijkheid **deugt niet!**

# AANPASSING CESUUR

**Resultaat**

Cijfer	5,5 (cijferopbouw)
Aangepast	↑
Score	24 (Maximale score: 40)
Resultaat	Voldoende
Logboek	Resultaatlogboek Afnamelogboek



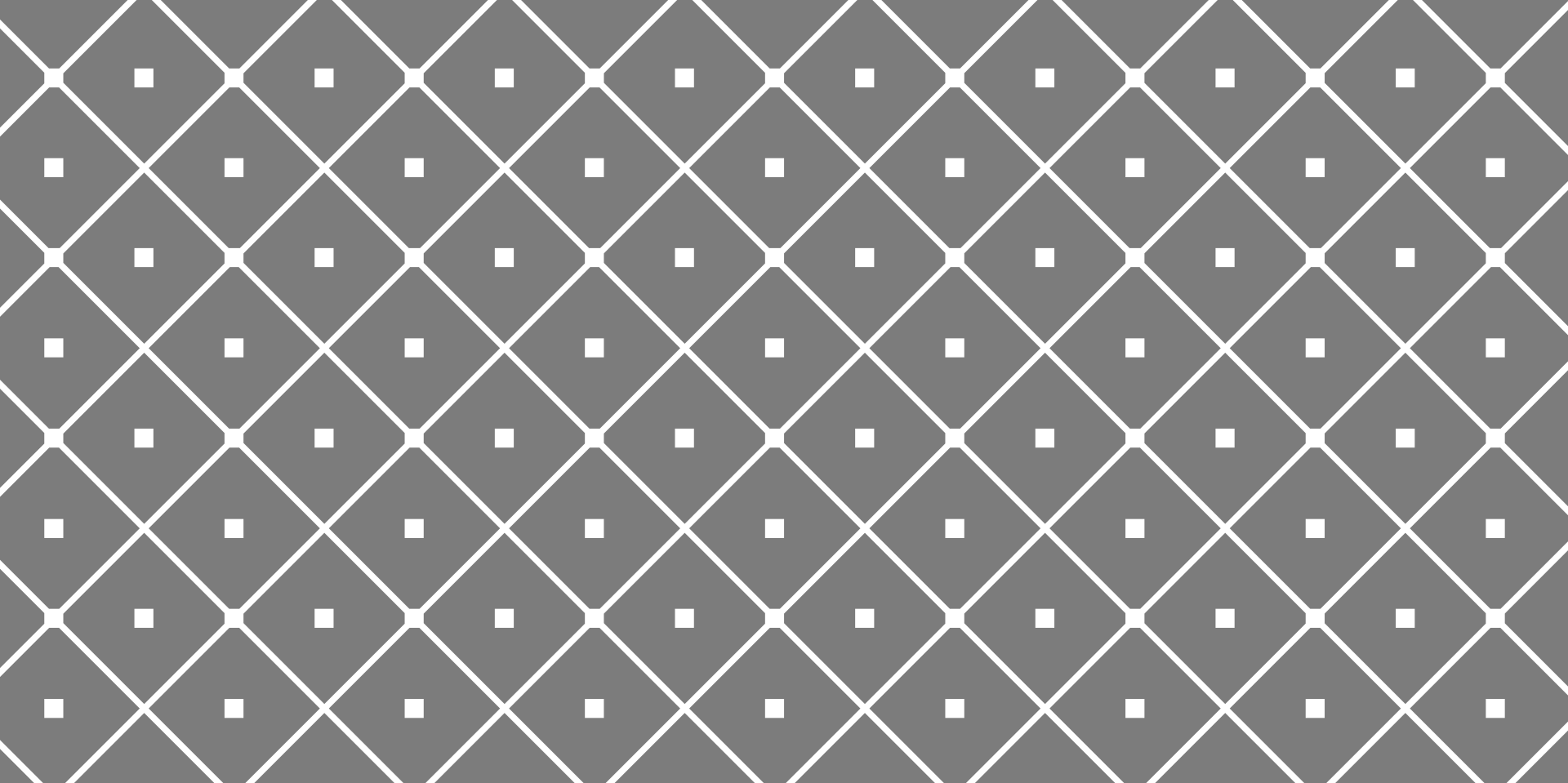
[Meer details...](#)

Transparantie: student kan zien dat resultaat is aangepast (in dit geval omhoog).

In 'Resultaatlogboek' kunnen details ingezien worden.

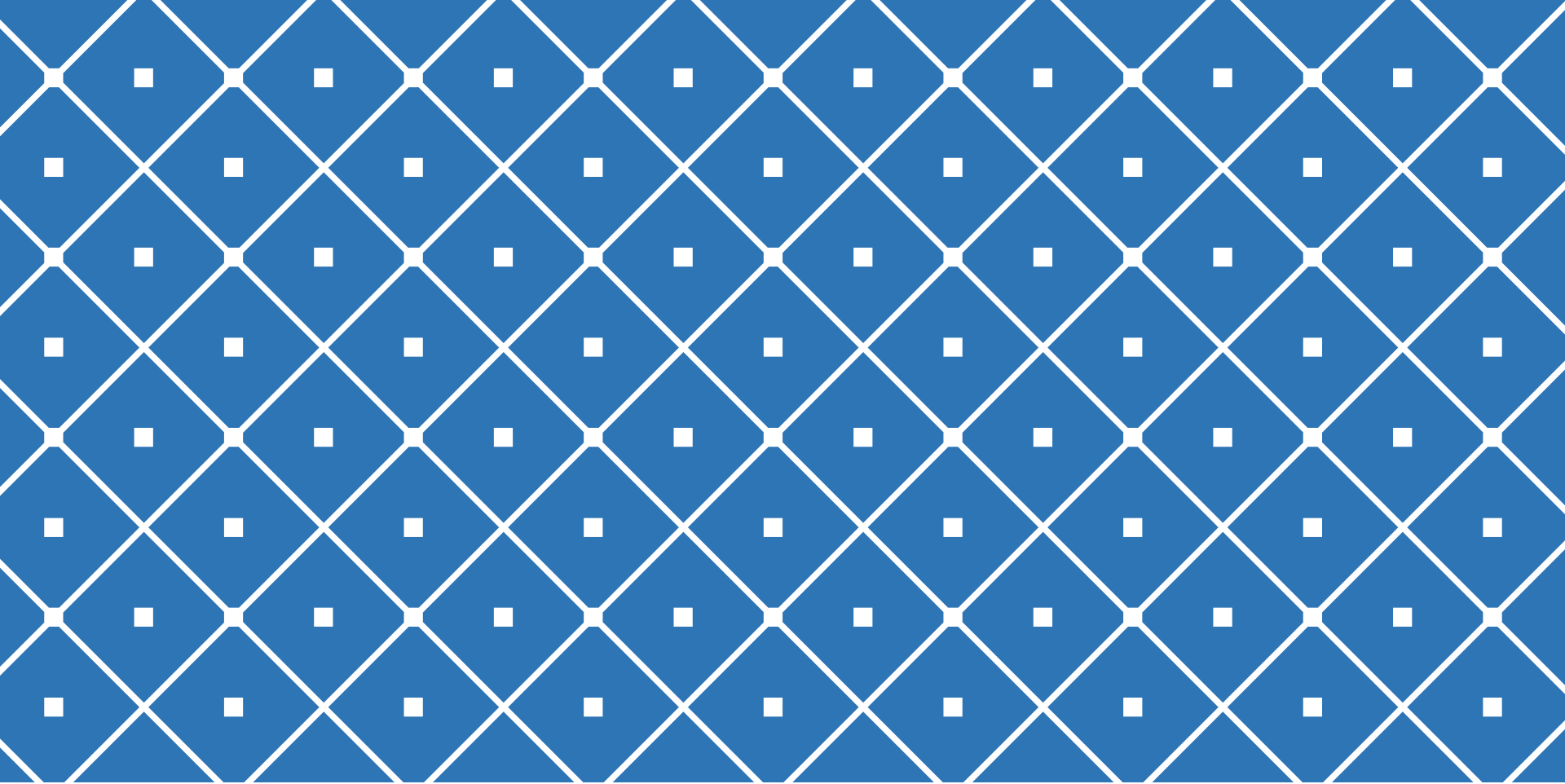
# AANPASSING CESUUR

Actie	Gegevens
De cesuur is aangepast	<p><b>Nieuwe beoordeling</b> 5,5 (geslaagd: Ja)</p> <p><b>Oude beoordeling</b> 5,2 (geslaagd: Nee)</p> <p><b>Nieuw scoregebied</b> Voldoende Je hebt 65% of meer van de punten behaald. Dit is voldoende,</p> <p><b>Oud scoregebied</b> Onvoldoende Je hebt minder dan 65% behaald. Dit is helaas niet voldoende,</p> <p><b>Motivatie:</b> Op grond van een analyse van de toetsresultaten is besloten de zak-slaaggrens (cesuur) bij te stellen van 65% naar 60% voor het cijfer 5,5. Let op, deze aanpassing geldt alléén voor deze toetsafname. Bij volgende afnames van deze toets kan de cesuur weer anders zijn.</p>
De status van het resultaat is aangepast	<b>Nieuwe status:</b> Afgesloten.



# CONTEST-RAPPORTAGE

Bestuderen  
Interpreteren  
Bespreken  
Acties definiëren



# TOETSEN EN TOETSVRAGEN ANALYSEREN

E-merge  
Digitaal toetsen  
2 november 2016