



# Correlatie en regressie met GeoGebra Suite

*Aanpak in de 2<sup>de</sup> graad doorstroomfinaliteit*

*Els Coussement*

*Els.Coussement@arteveldehs.be*

Vlaamse Vereniging voor Wiskunde Leraars

Digisprong in de wiskunde

25 maart 2023

*Spreidingsdiagrammen en trendlijnen zijn nieuw in de tweede graad doorstroomfinaliteit. Leerlingen interpreteren puntenwolken en zoeken naar een verband tussen twee grootheden. Cruciaal hierbij is dat leerlingen gebruik kunnen maken van ICT. In dit artikel lees je hoe we GeoGebra Suite kunnen inzetten bij dit leerstofonderdeel.*

In Uitwiskeling 38/1 (Deprez & Moons, 2022) tonen Johan Deprez en Filip Moons hoe dit deeltje bivariate statistiek in de tweede graad aan bod kan komen. De leerlingen krijgen de belangrijkste kernideeën van correlatie en regressie mee zonder dat de leerkracht ingaat op de abstracte wiskundige opbouw. Zij gaan in spreidingsdiagrammen op zoek naar enkele eenvoudige verbanden: lineair, recht evenredig, omgekeerd evenredig en zuiver kwadratisch. Heel belangrijk is dat we onze leerlingen leren dat we pas een besluit trekken na het bekijken van het spreidingsdiagram, het interpreteren van de correlatiecoëfficiënt als er een vermoeden is van een lineair verband én het bestuderen van de trendlijn in de puntenwolk.

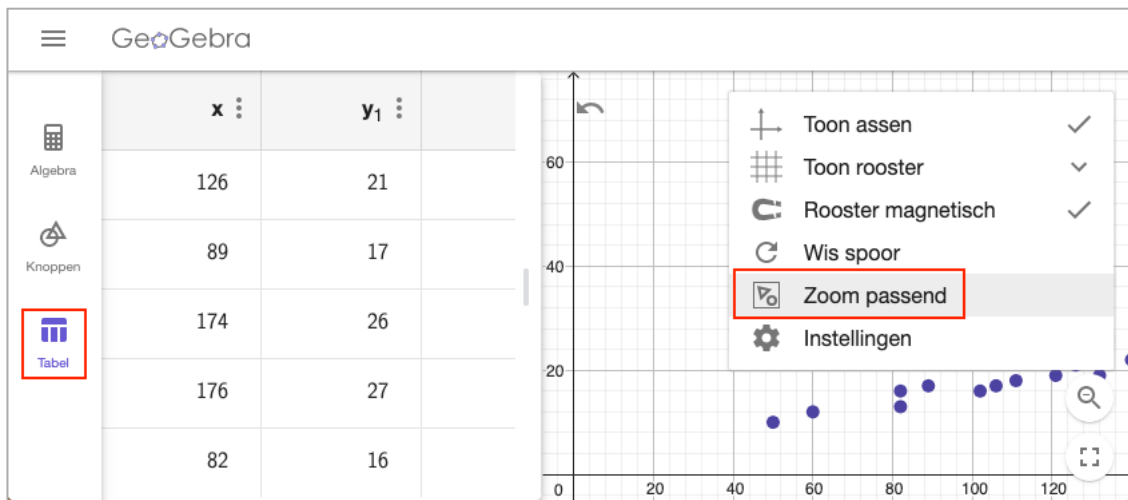
Bij elk soort verband waarvan sprake is in de eindtermen vind je hieronder een voorbeeld dat uitgewerkt is met de grafische rekenmachine van GeoGebra. Je kan hiervoor de aparte app ‘Grafische rekenmachine’ of de app ‘GeoGebra Suite’ gebruiken. Meer voorbeelden en oefeningen zijn te vinden in volgend GeoGebraboek: <https://www.geogebra.org/m/nmxcp3t8> (Coussement, 2022).

### Lineaire regressie

Amerikaanse onderzoekers hebben ontdekt dat er een verband bestaat tussen het getjirp van krekels en de temperatuur. In de tabel hieronder vind je data opgemeten in de Verenigde Staten in 2007 voor de Amerikaanse sneeuwboomkrekel. De grootheid  $A$  is het aantal keer dat de krekels tjirpen per minuut en de grootheid  $T$  is de temperatuur in Celsius. (Callaert, Bekaert, Goethals, Provoost, & Vancaudeberg, 2023)

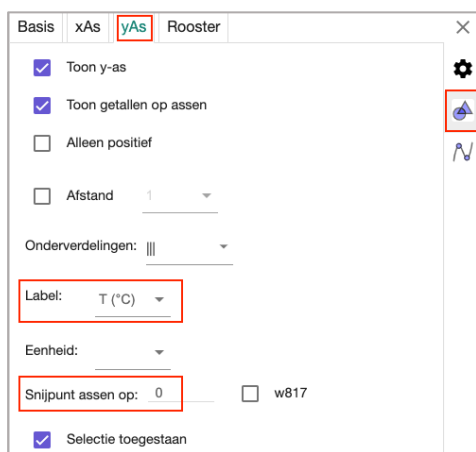
$A$ (aantal/min)	$T$ ( $^{\circ}C$ )	$A$ (aantal/min)	$T$ ( $^{\circ}C$ )	$A$ (aantal/min)	$T$ ( $^{\circ}C$ )
126	21	102	16	82	13
89	17	140	22	132	19
174	26	60	12	111	18
176	27	148	23	121	19
82	16	106	17	50	10

We onderzoeken nu of we op basis van de tjirpfrequentie van de sneeuwboomkrekels de temperatuur kunnen schatten. We maken hiervoor een spreidingsdiagram in de grafische rekenmachine door het tabelvenster te openen. Hiervoor klikken we links op het icoon *Tabel* (zie figuur 1). Daar voeren we alle gegevens in. De waarden voor de grootheid  $A$  komen in de kolom  $x$  en de waarden voor de grootheid  $T$  komen in de kolom rechts ervan. Bovenaan de tweede kolom verschijnt automatisch  $y_1$ . Eenmaal de data ingevoerd zijn, vinden we in het tekenvenster de puntenwolk terug. Om alle punten in het zichtbare deel van het tekenvenster te krijgen, kan de optie *Zoom passend* bij het tandwiel (rechtsboven in het tekenvenster) handig zijn.

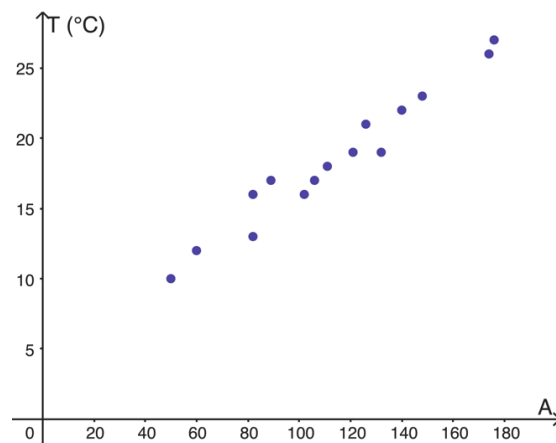


**Figuur 1** – Het maken van een spreidingsdiagram

Toch streeft een leerkracht vaak een betere lay-out na voor spreidingsdiagrammen die in cursusteksten of toetsen komen. Door rechtsboven onder *Zoom passend* op het tandwiel te klikken, komen we bij de instellingen terecht. Daar kunnen we bijvoorbeeld het label bij een as aanpassen (zie figuur 2) en ervoor zorgen dat de assen elkaar niet snijden in de oorsprong. Verder kunnen we ook wijzigingen aanbrengen aan het rooster, het zichtbare tekenvenster bepalen, de kleur van de assen kiezen, de lettergrootte aanpassen, ...



**Figuur 2** – Lay-out van het spreidingsdiagram aanpassen via de instellingen



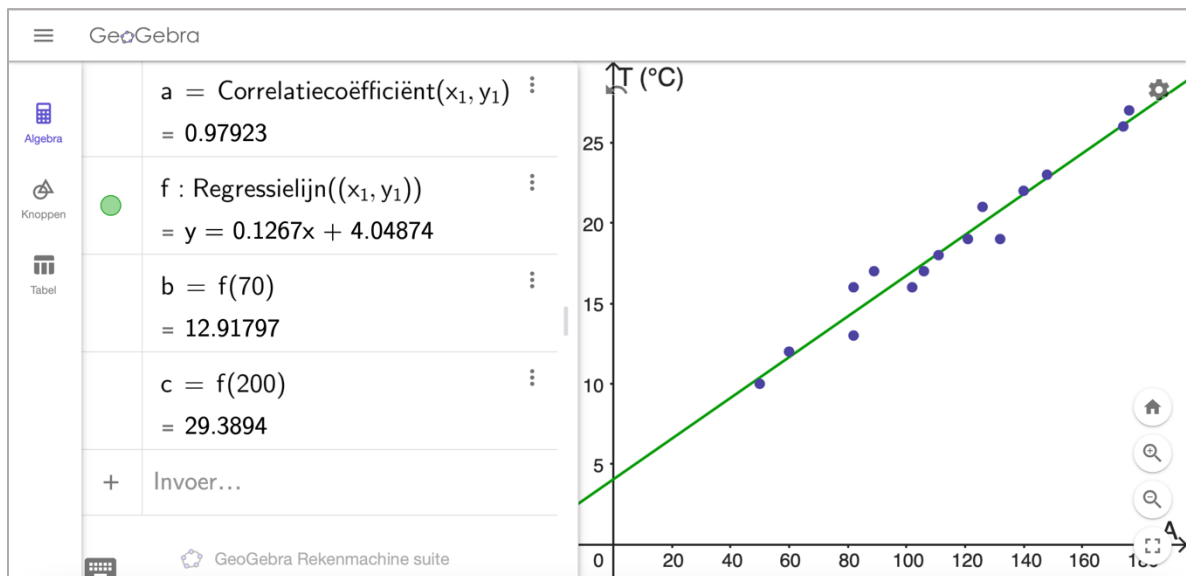
**Figuur 3** – Tjirpen en temperatuur: spreidingsdiagram

Het spreidingsdiagram (zie figuur 3) doet vermoeden dat er een sterk positief lineair verband is tussen de grootheden  $A$  en  $T$ . We berekenen bijgevolg de correlatiecoëfficiënt van Pearson. Hiervoor gaan we naar het algebrafenster en gebruiken het commando  $Correlatiecoëfficiënt(x_1, y_1)$ . Met  $x_1$  verwijzen we naar de data in de kolom  $x$ , met  $y_1$  verwijzen we naar de data in de kolom  $y_1$ . We vinden dat  $r = 0,97923$ , wat ons vermoeden bevestigt. Het is dus verantwoord om op zoek te gaan naar de regressierechte, de rechte die het best dit lineair verband beschrijft. Hiervoor maken we gebruik van het commando  $Regressielijn((x_1, y_1))$ . Let op! De twee haakjes zijn hierbij noodzakelijk! We vinden het volgende verband:

$$T = 0,1267 \cdot A + 4,0487$$

De regressierechte verschijnt bij de puntenwolk (zie figuur 4) en lijkt goed aan te sluiten bij de data. We kunnen dit verband bijgevolg gebruiken voor inter- en extrapolatie. Als de krekels 70 keer tjirpen per minuut, schatten we een temperatuur van  $13^{\circ}C$ . Tellen we in één minuut 200 keer het getjirp van krekels, dan zal de temperatuur vermoedelijk ongeveer  $29^{\circ}C$  zijn. Met de laatste uitspraak moeten we

wel iets voorzichtiger omspringen omdat een extrapolatie altijd minder betrouwbaar is dan een interpolatie.



**Figuur 4** – Het berekenen van de correlatiecoëfficiënt en de regressierechte

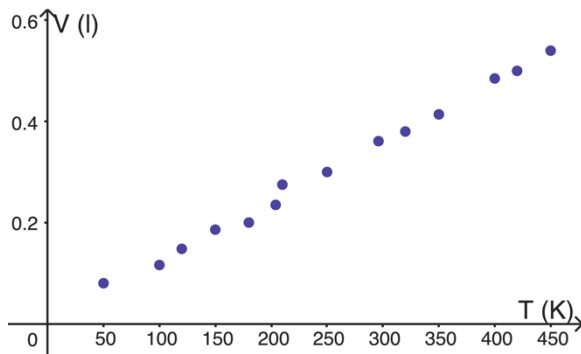
Merk op dat we de regressierechte ook kunnen vinden door in de tabel te klikken op de drie horizontale puntjes naast  $y_1$ . We kiezen vervolgens voor *Regressie* en *Linear*. Door op *Plot* te klikken krijgen we dezelfde regressierechte, al ziet het commando in het algebrafenster er wel wat complexer uit.

### Recht evenredig verband

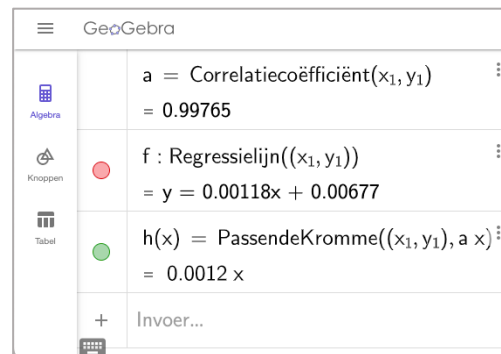
Een leerkracht fysica laat zijn leerlingen experimenten uitvoeren om hiermee de wet van Gay-Lussac te illustreren. Bij een gegeven hoeveelheid gas wordt de druk constant gehouden. Hieronder vind je voor elke meting de temperatuur  $T$  in Kelvin en het volume  $V$  van het gas in liter. Deze dataset is een variant op een dataset die gebruikt wordt in een werktekst van de UHasselt. (Callaert, Bekaert, Goethals, Provoost, & Vancaudeberg, 2023)

$T$ (K)	$V$ (l)	$T$ (K)	$V$ (l)	$T$ (K)	$V$ (l)
50	0,080	204	0,235	350	0,414
100	0,116	210	0,275	400	0,485
120	0,148	250	0,300	420	0,500
150	0,186	296	0,361	450	0,540
180	0,200	320	0,380		

Opnieuw starten we met het maken van een spreidingsdiagram (zie figuur 5). Deze puntenwolk doet vermoeden dat er een sterk positief lineair verband is tussen  $T$  en  $V$ . De correlatiecoëfficiënt  $r = 0,99765$  bevestigt dit vermoeden.



**Figuur 5** – Gay-Lussac: spreidingsdiagram



**Figuur 6** – Gay-Lussac: correlatiecoëfficiënt en mogelijke trendlijnen

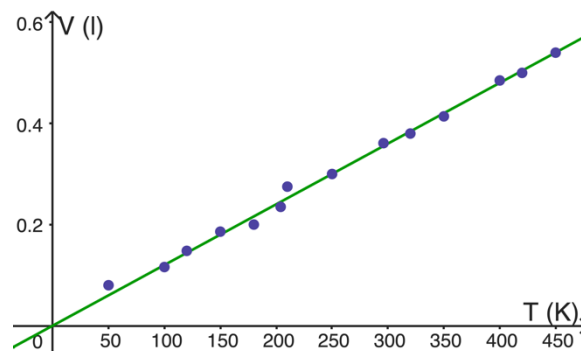
We vinden volgende regressierechte (zie figuur 6):

$$V = 1,18 \cdot 10^{-3} \cdot T + 6,77 \cdot 10^{-3}$$

Toch is dat niet wat wij voor ogen hebben. De wet van Gay-Lussac zegt dat er een recht evenredig verband is tussen  $V$  en  $T$  als de druk constant is. We zoeken dus een rechte door de oorsprong. Hiervoor tikken we het commando  $PassendeKromme((x_1, y_1), a x)$  in in het algebravenster. We vinden hiermee het volgende verband (zie figuur 6):

$$V = 1,20 \cdot 10^{-3} \cdot T$$

Figuur 7 toont dat deze rechte vrij goed aansluit bij de meetresultaten.



**Figuur 7** – Gay-Lussac: spreidingsdiagram en trendlijn

### Omgekeerd evenredig verband

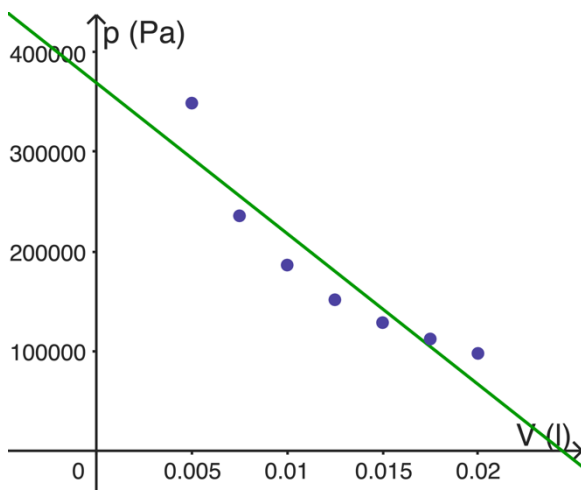
Analoog kan deze leerkracht ook proeven laten uitvoeren waaruit de wet van Boyle kan gehaald worden. Bij een gegeven hoeveelheid gas wordt nu telkens de temperatuur constant gehouden. In de volgende tabel vind je voor elke meting het volume  $V$  in liter en de gemiddelde druk  $p$  in Pascal. (De wet van Boyle (CoachLab Experiment), 2004)

$V$ (l)	$p$ (Pa)
0,0200	97 790
0,0175	112 260
0,0150	128 660
0,0125	151 580
0,0100	186 390
0,0075	235 790
0,0050	348 810

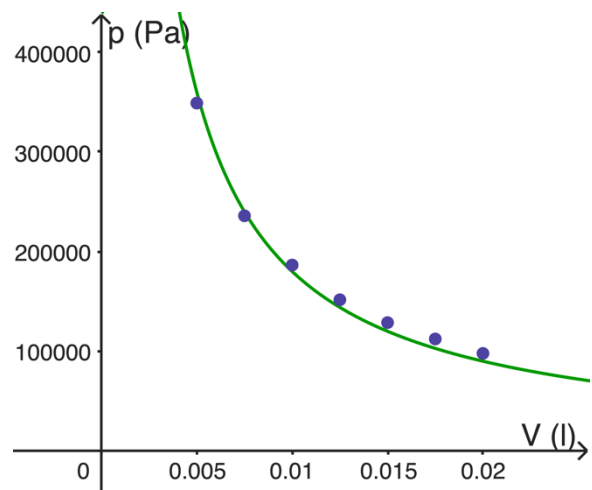
Met het spreidingsdiagram in figuur 8 en de correlatiecoëfficiënt  $r = -0,9277$  zouden we het idee kunnen hebben dat er een sterk negatief lineair verband bestaat tussen de twee grootheden. De regressierechte, getekend in figuur 8, lijkt op het eerste gezicht goed aan te sluiten bij de datapunten. Als we nader toekijken, zien we echter dat de afwijkingen tussen de datapunten en de trendlijn hier niet zo toevallig lijken te zijn. Links en rechts liggen de datapunten boven de regressierechte en in het midden liggen ze eronder. Dat is een signaal dat erop wijst dat een lineair verband misschien niet het beste model is. We zoeken hier beter naar een gebogen trendlijn. In dit geval suggereert de puntenwolk een omgekeerd evenredig verband tussen de twee grootheden. Met het commando  $\text{PassendeKromme}\left((x_1, y_1), \frac{a}{x}\right)$  vinden we de functie  $f$  met  $f(x) = \frac{a}{x}$  ( $a \in \mathbb{R}$ ) waarvan de grafiek het best de datapunten benadert. GeoGebra geeft de trendlijn met vergelijking  $y = \frac{1799,4099}{x}$ . Volgend model kunnen we dus gebruiken om het verband tussen  $p$  en  $V$  te beschrijven:

$$p = \frac{1799}{V}$$

In figuur 9 zien we dat deze gebogen trendlijn veel beter aansluit bij de gegeven datapunten dan de regressierechte. Voor inter- en extrapolatie kunnen we bijgevolg beter het model gebaseerd op het omgekeerd evenredig verband gebruiken.



**Figuur 8** – Boyle: spreidingsdiagram en regressierechte



**Figuur 9** – Boyle: spreidingsdiagram en gebogen trendlijn

### Zuiver kwadratisch verband

We laten een kogel verschillende keren van een helling afrollen. Telkens noteren we de afgelegde tijd  $t$  in seconden en de afstand  $s$  in meter. De gegevens van een concreet experiment zijn te vinden in volgende tabel. (Voorbeeld meetrapport, 2013)

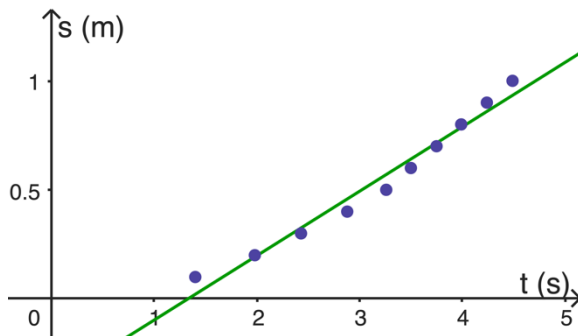
$t$ (s)	$s$ (m)	$t$ (s)	$s$ (m)
1,40	0,10	3,50	0,60
1,98	0,20	3,75	0,70
2,43	0,30	3,99	0,80
2,88	0,40	4,24	0,90
3,26	0,50	4,49	1,00

De puntenwolk en de correlatiecoëfficiënt  $r = 0,9857$  kunnen de eerste indruk opwekken dat er hier sprake is van een lineair verband. Ook hier zien we echter dat de punten niet willekeurig rond de regressierechte liggen. Bijgevolg zoeken we naar een gebogen trendlijn. Ditmaal is een omgekeerd

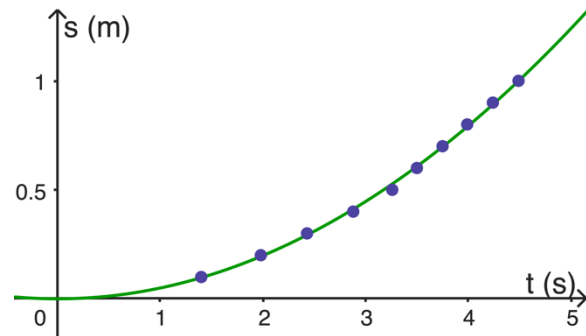
evenredig verband niet aan de orde. Het spreekt bovendien voor zich dat de kogel na 0 s geen afstand afgelegd heeft. De gebogen trendlijn moet dus door de oorsprong gaan. Intuïtief denken we hier aan een zuiver kwadratisch verband. Met het commando *PassendeKromme* $((x_1, y_1), a x^2)$  zoeken we naar de functie  $f$  met  $f(x) = ax^2$  ( $a \in \mathbb{R}$ ) die het best aansluit bij de gegeven datapunten. We vinden dus het volgende verband tussen de grootheden  $s$  en  $t$ :

$$s = 0,0496 \cdot t^2$$

De grafiek die hierbij hoort, lijkt inderdaad dicht bij de punten in de puntenwolk te liggen.



**Figuur 10** – Eenparig versnelde beweging: spreidingsdiagram en regressierechte



**Figuur 11** – Eenparig versnelde beweging: spreidingsdiagram en gebogen trendlijn

## Conclusie

De grafische rekenmachine van GeoGebra biedt de mogelijkheid om leerlingen van de tweede graad doorstroomfinaliteit op intuïtieve wijze de basisprincipes van correlatie en regressie mee te geven. Dankzij de gebruiksvriendelijke interface van GeoGebra Suite kunnen ook leerkrachten en leerlingen met weinig ervaring in GeoGebra vrij snel een spreidingsdiagram, een correlatiecoëfficiënt en trendlijnen bij gegeven data laten verschijnen.

## Bronnen

Callaert, H., Bekaert, H., Goethals, C., Provoost, L., & Vancaudeberg, M. (2023). *Lesmateriaal statistiek: werkteksten*. (UHasselt) Opgeroepen op 11 januari 2023, van Aanbod voor scholen en leerkrachten: <https://www.uhasselt.be/nl/info-voor/aanbod-voor-scholen-en-leerkrachten/lesmateriaal/statistiek/werkteksten>

Coussement, E. (2022). *Correlatie en regressie met GeoGebra Suite*. Opgeroepen op 11 januari 2023, van GeoGebra: <https://www.geogebra.org/m/nmxcp3t8>

*De wet van Boyle (CoachLab Experiment)*. (2004). Opgeroepen op 11 januari 2023, van <http://www.thuisexperimenteren.nl/science/coachlabexp/wetvanboyle/boyle.htm>

Deprez, J., & Moons, F. (2022). Spreidingsdiagrammen en trendlijnen. *Statistiek in de 2de graad. Uitwiskeling 38/1*, 21-39.

*Voorbeeld meetrapport*. (2013). Opgeroepen op 11 januari 2023, van FysiKarel: <https://www.fysikarel.nl/Voorbeeldmeetrapport.htm>

## GeoGebra Vlaanderen

Het landschap van GeoGebra is nog steeds in ontwikkeling. Op regelmatige basis verschijnen nieuwe elementen. Als je graag op de hoogte blijft van alle mogelijkheden van GeoGebra, hou dan zeker onze *facebookpagina 'GeoGebra Vlaanderen'* in de gaten:

<https://www.facebook.com/GeoGebraVlaanderen>

Volgers van deze pagina kunnen rekenen op nieuwtjes die heet van de naald zijn maar ook op linken naar artikels, nascholingen, nieuwe of gewijzigde initiatiehandleidingen, online cursussen en introductiefilmpjes.