TRIPS

Whitepaper · november 2026



DATA-ONDERSTEUND WERKEN · DEEL 2 VAN 2

Van verband naar bewijs

Een verband in je eigen gegevens is een hypothese. Zo toets je die eerlijk en ga je na of werken met data je school beter maakt. Voor schoolleiders en bestuurders in PO, VO en (V)SO.

Inhoud

01 In het kort

02 Van vermoeden naar toets

03 Van verband naar hypothese

04 Zo toets je of data-ondersteund werken loont

05 Hoe TRIPS helpt

06 Verantwoording

07 Bronnen

IN HET KORT

1 **Behandel een verband in je eigen gegevens als hypothese.** Minder lestijd, verzuim van leerlingen en een slechte relatie tussen leraar en leerling gaan samen met lagere resultaten of minder betrokkenheid (Engzell e.a., 2021; Keppens, 2023; Roorda e.a., 2011). Waar onderzoekers een grootte noemen, zijn de verbanden met leerresultaten meestal klein. En vaak gaat het alleen om samenhang. Groeien groepen met veel invallers minder, dan kan dat ook komen door wie er in die groepen zit. Toets je hypothese door een verandering gefaseerd in te voeren en te loten welke scholen of groepen eerst beginnen. De andere dienen tot hun beurt als vergelijking (Coe e.a., 2013).

2 **Een verband toets je in zeven stappen. De eerste is opschrijven wat je verwacht.** Zo begint ook de Nederlandse datateammethode: een team toetst eerst een hypothese over de oorzaak en kiest pas dan maatregelen (Schildkamp e.a., 2014a). Een analyse over invallers of lesuitval is niet geschikt om een individuele leraar te beoordelen.

3 **Met weinig groepen is een verschil al snel toeval.** In het verzonnen voorbeeld groeien groepen met veel invallers 2 punten minder in begrijpend lezen. Bij de 24 groepen van één bestuur in één jaar kan dat toeval zijn. Bij honderden groepen wordt toeval minder waarschijnlijk, maar een oorzaak is dan nog niet bewezen.

4 **Of werken met data een school beter maakt, toets je niet met één school.** In gerandomiseerde experimenten in het onderwijs ligt het middelste effect rond 0,10 SD (Kraft, 2020). Bij een handvol scholen verdwijnt zo'n effect in de schommelingen van jaar tot jaar (Kane & Staiger, 2002). Binnen een bestuur kun je wel nagaan of je eigen aanpak werkt, met een vooraf vastgelegde uitkomst en een opzet die oorzaak en samenhang zo goed mogelijk scheidt.

1 · AANLEIDING

Van vermoeden naar toets

Leg je gegevens uit verschillende bronnen naast elkaar, dan zie je al snel verbanden, bijvoorbeeld dat groepen met veel invallers minder groeien. Wat doe je daarmee? En maakt werken met data je school beter?

Met werken met data bedoelen we dat een school haar eigen gegevens systematisch analyseert, er verbeteringen op baseert en die evalueert (Schildkamp & Kuiper, 2010). We gaan uit van deze hypothese: *scholen die data gebruiken om hun beleid te onderbouwen, presteren beter dan scholen die dat niet doen.*

Met beter bedoelen we dat leerlingen meer vooruitgaan dan je op grond van hun startpunt en achtergrond mag verwachten, en meer dan leerlingen op dezelfde school in eerdere jaren. Het gaat om kennis en vaardigheden, hun ontwikkeling als burger en als persoon en hun welbevinden, bij alle groepen leerlingen en over meerdere jaren.

We delen schoolgegevens in vier pijlers in, vrij vertaald naar Bernhardt (1998): leerlingen en context, ervaringen en beleving, leren en resultaten, en onderwijs en organisatie. Waarom je ze naast elkaar legt, liet deel 1 zien met een verzonnen basisschool. In groep 6 daalde de tevredenheid over de lessen en bleef de groei in begrijpend lezen achter. Pas naast de bezetting bleek dat er van november tot maart drie invallers voor de klas stonden. Het bleef een vermoeden. Nu wil haar bestuur weten of zoiets vaker voorkomt.

> TRIPS maakt een dataplatform voor scholen en besturen en heeft dus belang bij de uitkomst. Daarom laten we ook de onderzoeken zien die tegen de hypothese ingaan en doen we alleen uitspraken die het onderzoek draagt. Bij het zoeken en samenvatten van een deel van de bronnen hebben we AI gebruikt (Claude, van Anthropic).

2 · VERBANDEN TOETSEN

Van verband naar hypothese

Leg je gegevens uit de pijlers naast elkaar, dan vallen al snel verbanden op. Misschien groeien leerlingen minder in begrijpend lezen in groepen die lang met invallers werkten. Of VO-leerlingen halen lagere cijfers bij toetsen in het eerste uur. Als schoolleider of bestuurder wil je dan weten wat zo'n verband over de oorzaak zegt, en of je er beleid op kunt bouwen.

Neem een verzonnen basisschool van een PO-bestuur met zes scholen. In groep 6 daalt de tevredenheid van leerlingen over de lessen en blijft de groei in begrijpend lezen achter. De andere groepen 6 van het bestuur blijven stabiel. De samenstelling van de groep is niet veranderd, maar van

november tot maart stonden er van november tot maart drie verschillende invallers voor de klas; sinds maart is er weer een vaste leerkracht. Of groepen met veel invallers vaker minder groeien, kan het bestuur nagaan met de gegevens van al zijn zes scholen.

Wat onderzoek over verbanden vond

De tabel zet per pijler verbanden met leerresultaten, uitval of betrokkenheid op een rij, met Nederlands onderzoek voorop. Waar we de grootte konden nagaan, staat die erbij, vaak als effectgrootte in standaarddeviaties (SD): de gebruikelijke spreiding van scores rond het gemiddelde. Ter vergelijking: onderzoekers die de schoolsluiting van 2020 bestudeerden, vergeleken een achterstand van 0,08 SD bij basisschoolleerlingen met ongeveer een vijfde schooljaar aan leergroei (Engzell e.a., 2021). In gerandomiseerde experimenten met onderwijsinterventies ligt het middelste effect rond 0,10 SD (Kraft, 2020). Effectgroottes uit verschillende studies kun je niet zomaar rangschikken (Simpson, 2017).

Verband	Wat onderzoek vond	Grootte	Soort bewijs	Waar
ONDERWIJS EN ORGANISATIE				
Schoolsluiting en leergroei	Tijdens acht weken schoolsluiting in 2020 maakten basisschoolleerlingen weinig of geen vooruitgang.	0,08 SD, bij laagopgeleide ouders tot 60% groter	■ **oorzaak aannemelijk**	Nederland, PO (Engzell e.a., 2021)
Extra lestijd en prestaties	In een experiment met loting op zeven basisscholen leverde een verlengde schooldag van drie maanden geen meetbare verbetering op voor taal of rekenen.	geen meetbaar effect	■ **oorzaak aannemelijk**	Nederland, PO (Meyer & Van Klaveren, 2013)
Ervaring van de leerkracht en prestaties	Bij tweelingen die op dezelfde basisschool in verschillende klassen zaten, scoorden leerlingen beter naarmate hun leerkracht meer ervaring had.	ongeveer 1% van een SD per extra jaar ervaring	■ **oorzaak aannemelijk**	Nederland, PO (Gerritsen e.a., 2017)
Tijdstip van de toets en cijfers	Leerlingen die van nature laat actief zijn, haalden op toetsen in de ochtend lagere cijfers dan vroege types. In de vroege middag was het verschil niet significant.	geen grootte vermeld	■ **samenhang**	Nederland, VO, 741 leerlingen (Van der Vinne e.a., 2015)
Geobserveerde lesgeefkwaliteit en betrokkenheid	Scores op ICALT, een observatie-instrument voor lesgeven, hangen samen met de betrokkenheid van leerlingen, en in een eerdere studie met hun houding, gedrag en resultaten.	geen grootte vermeld	■ **samenhang**	Nederland en andere landen, PO en VO (Van de Grift, 2007; Van de Grift e.a., 2017)
Samenwerking, professionele cultuur en prestaties	Op 53 basisscholen ging hechtere samenwerking tussen leraren samen met betere leerresultaten. Een meta-analyse van Groningse onderzoekers vond een klein verband tussen professionele gemeenschap en prestaties.	in de meta-analyse 0,25 SD (klein)	■ **samenhang**	Nederland, PO (Moolenaar e.a., 2012); internationaal (Lomos e.a., 2011a)

Verband	Wat onderzoek vond	Grootte	Soort bewijs	Waar
Onderwijstijd en prestaties	Meer onderwijstijd hing gemiddeld maar zwak samen met betere prestaties.	zwak	■ samenhang	internationaal, overzicht door Nederlandse onderzoekers (Hendriks e.a., 2014)
Instroom van nieuwe leraren en examenresultaten	Meer instroom van nieuwe leraren ging samen met iets lagere eindexamenresultaten.	klein	■ samenhang	Engeland, VO (Gibbons e.a., 2021)
Verzuim van de leraar en rekenscores	Per tien verzuimdagen van de eigen leraar scoorden leerlingen lager op rekenen.	ongeveer 3% van een SD per 10 dagen	■ oorzaak aannemelijk	VS, PO (Miller e.a., 2008)
ERVARINGEN EN BELEVING				
Hoe leerlingen hun leraar ervaren en prestaties	Als leerlingen hun leraar als sturend en betrokken ervaren, gaat dat samen met betere toetsprestaties van de klas en een positievere houding tegenover het vak.	voor sommige uitkomsten meer dan de helft van de verschillen tussen klassen	■ samenhang	Nederland, VO (Den Brok e.a., 2004)
Relatie tussen leraar en leerling	Een warme relatie gaat samen met meer betrokkenheid en iets betere prestaties.	middelgroot tot groot voor betrokkenheid, klein tot middelgroot voor prestaties	■ samenhang	internationaal, meta-analyse van 99 studies (Roorda e.a., 2011)
Gepest worden en prestaties	Leerlingen die gepest worden, presteren gemiddeld lager.	geen grootte vermeld	■ samenhang	internationaal, meta-analyse van 33 studies (Nakamoto & Schwartz, 2010)
LEERLINGEN EN CONTEXT				
Spijbelen en voortijdig schoolverlaten	Leerlingen die spijbelen, verlaten vaker de school zonder startkwalificatie.	3,4 procentpunt meer kans	■ samenhang	Nederland, VO (De Witte & Csillag, 2014)
Verzuim melden en uitval	Betere verzuimmelding verlaagde volgens een schatting de uitval. Dat zegt iets over signaleren, niet over verzuim zelf.	ongeveer 5 procentpunt minder uitval	■ oorzaak aannemelijk	Nederland, VO (De Witte & Csillag, 2014)
Verzuim van leerlingen en resultaten	Op een Nederlandse school haalden leerlingen met meer verzuim lagere cijfers. In Vlaanderen hing ook geoorloofd verzuim samen met lagere examenresultaten.	geen grootte vermeld	■ samenhang	Nederland, VO, één school (Zerbini e.a., 2017); Vlaanderen, VO (Keppens, 2023)

■ **oorzaak aannemelijk** experiment, natuurlijk experiment (een toevallige omstandigheid of regel werkt als loting) of vergelijking binnen dezelfde leraar, leerling of school, ook als er geen effect uitkwam ■ **samenhang** verband op één moment of tussen groepen, waarbij de oorzaak ook elders kan liggen

De verbanden met leerresultaten zijn meestal klein. Een effect als dat van tien verzuimdagen van een leraar in het Amerikaanse onderzoek (Miller e.a., 2008) valt in één groep of klas weg in de gewone schommelingen van jaar tot jaar.

Achter een samenhang kan een derde factor zitten, zoals de samenstelling van de leerlingpopulatie. Ook de registratie speelt mee. Volgens Vlaams onderzoek kan geregistreerd ongeoorloofd verzuim afwijken van wat leerlingen zelf melden, en die afwijking verschilt tussen soorten leerlingen (Keppens e.a., 2019).

Ervaringen hangen sterker samen met betrokkenheid dan met cijfers, bijvoorbeeld bij de relatie tussen leraar en leerling (Roorda e.a., 2011). De dalende tevredenheid in groep 6 zegt dus vermoedelijk meer over betrokkenheid, maar is wel een reden om te kijken wat er in de les gebeurt.

Wat bekend is over invallers en leesgroei

We vonden geen Nederlands onderzoek dat invallers of wisselende leerkrachten binnen een schooljaar koppelt aan leesgroei. Tijdens de schoolsluiting van 2020 ging minder lestijd samen met minder leergroei (Engzell e.a., 2021). Maar een groep met invallers krijgt gewoon les; de vraag is of die lessen even goed en samenhangend blijven.

Meer lestijd alleen lijkt weinig uit te maken (Hendriks e.a., 2014; Meyer & Van Klaveren, 2013). Wat er in de les gebeurt, telt meer. Leerkrachten in groep 4 en 5 leerden werken met doelen, toetsgegevens en betere leesinstructie, en hun leerlingen groeiden meer in begrijpend lezen dan leerlingen in vergelijkbare klassen (Van Kuijk, 2014). Ook de ervaring van de leerkracht telt mee (Gerritsen e.a., 2017). Een invaller die de groep en de leesaanpak nog niet kent, kan juist op die punten tekortschieten. Dat maakt het een redelijke hypothese dat maanden met wisselende invallers de leesgroei drukken. Rechtstreeks getoetst is dat niet.

Neem ook een andere verklaring mee. Er zijn aanwijzingen dat het lerarentekort zich ophoopt op scholen met veel leerlingen met een onderwijsachterstand (Jonk & De Wolf, 2025), al betwisten andere onderzoekers dat (Van Wickeren e.a., 2025). Klopt het eerste beeld, dan meet een vergelijking van invallers en leesgroei mogelijk deels het verschil in leerlingpopulatie.

Een verband zelf toetsen

De datateammethode van de Universiteit Twente biedt een stappenplan om zo'n hypothese te toetsen (Schildkamp e.a., 2014a). Een team van leraren en schoolleiders stelt een meetbare hypothese op over de oorzaak van een probleem en toetst die met eigen gegevens, voordat het maatregelen kiest. De zeven stappen hieronder werken voor elk verband; wij nemen de invallers als voorbeeld.

1 Formuleer de hypothese vooraf

Schrijf de hypothese op voordat je de gegevens bekijkt. In het voorbeeld luidt ze zo. *In groepen waar op meer dan een kwart van de lesdagen een invaller of wisselende leerkracht voor de klas stond, groeien leerlingen minder in begrijpend lezen dan je op grond van hun eerdere vaardigheidsscores mag verwachten.*

Leg ook vast welke groepen en welk schooljaar je bekijkt, wat je als veel invallers rekent en welke uitkomst telt. Spreek af hoe groot een verschil moet zijn om ertoe te doen. Wie dat achteraf doet, kiest al snel de grens met het mooiste resultaat.

2 Maak vooraf afspraken over de gegevens

- **Welke gegevens?** Gebruik voor de bezetting alleen het rooster of de vervangingsregistratie. Je wilt weten wie er voor de groep stond, niet waarom, dus de verzuimregistratie van personeel blijft buiten de analyse. Leg ook vast dat de uitkomsten niet dienen om een individuele leerkracht te beoordelen.
- **Openheid en kleine aantallen.** Vertel leerkrachten en directeuren vooraf wat het doel is, welke gegevens je gebruikt en wie de uitkomsten ziet. In het PO wijst zelfs een gemiddelde per groep naar één of twee leerkrachten. Rapporteer daarom alleen over meerdere groepen samen, met een vooraf afgesproken minimumaantal.
- **Wie analyseert?** Iemand die leerkrachten niet beoordeelt, bijvoorbeeld een kwaliteitsmedewerker van het bestuur.

3 Kies eenheid en maat

- **Eenheid.** Analyseer per groep, want alle leerlingen maken dezelfde wisselingen mee. Tel je elke leerling apart, dan oogt de uitkomst te zeker. In het VO kan dat per lesgroep.
- **Maat.** Het aantal invallerdagen blijft een benadering. Een duocollega die extra dagen overneemt, is voor de groep iets anders dan een onbekende invaller. Leg vooraf vast hoe je zulke gevallen telt.

4 Leg de andere pijlers ernaast

Groepen verschillen ook in startniveau (eerdere vaardigheidsscores), samenstelling, het aantal leerlingen met extra ondersteuning, het verzuim van leerlingen en de leesaanpak van de school. Neem die factoren mee in de analyse.

5 Kijk naar grootte en zekerheid

Vergelijk de leesgroei van groepen met en zonder veel invallers met wat je op grond van eerdere scores mag verwachten. Een eindscore alleen zegt vooral waar een groep begon.

Met enkele tientallen groepen kan een verschil van een paar punten makkelijk toeval zijn. Hoe zeker je bent, hangt vooral af van het aantal groepen (zie de figuur na het voorbeeld).

6 Zoek de andere verklaringen

- **Omgekeerde richting.** Een zware groep kan er ook toe leiden dat een leerkracht uitvalt of vertrekt. Dan brengt de groep de invallers mee.
- **Toeval.** Wie veel verbanden tegelijk bekijkt, vindt er altijd wel een paar die toevallig sterk lijken.
- **Terugval naar het gemiddelde.** Een groep die uitzonderlijk laag scoorde, scoort de volgende keer vaak vanzelf hoger, ook zonder maatregel.
- **Kleine aantallen.** Gemiddelden van kleine groepen schommelen sterk van jaar tot jaar, zo liet Amerikaans onderzoek naar schoolgemiddelden op basisscholen zien (Kane & Staiger, 2002). Bekijk liever meerdere jaren. Ook de Inspectie beoordeelt de leerresultaten in het PO over de laatste drie schooljaren (Staatscourant 2026, 26023).
- **Dezelfde school of leerkracht.** Groepen op dezelfde school delen vaak directeur, leesmethode en soms invallers. Vier groepen op één school vertellen je minder dan vier groepen op vier scholen. Dat geldt ook voor groepen met dezelfde leerkracht in verschillende jaren.

7 Toets door te veranderen

Een verband in bestaande gegevens blijft samenhang. Verander daarom iets bij een deel van de groepen of scholen en vergelijk met groepen waar de verandering nog niet is ingevoerd.

Loten wie de nieuwe aanpak krijgt, is de sterkste manier om een eerlijke vergelijkingsgroep te maken (Coe e.a., 2013). Een gefaseerde invoering is meestal haalbaarder: iedereen krijgt de verandering, je loot alleen wie begint. Lukt ook dat niet, vergelijk dan met zo gelijk mogelijke groepen; de uitkomst is dan minder zeker (Coe e.a., 2013). Kies de deelnemers in elk geval niet uit omdat ze laag scoorden, want dan speelt terugval naar het gemiddelde mee.

Loot in het PO liever per school dan per groep. Anders neemt een vergelijkingsgroep op dezelfde school de nieuwe aanpak al snel deels over, en lijkt het verschil kleiner dan het is.

Ga ook na of de verandering echt in de klas plaatsvond. In een Amerikaans experiment gingen leraren ondanks training en coaching niet vaker gedifferentieerd lesgeven (Cordray e.a., 2013).

Zie je bij een handvol groepen geen verschil, dan bewijst dat nog niet dat er geen effect is.

UITGEWERKT VOORBEELD · VERZONNEN BESTUUR

Groeien groepen met veel invallers minder in begrijpend lezen?

Het verzonnen PO-bestuur bekijkt groep 5 tot en met 8 op al zijn zes scholen, samen 24 groepen. In zes daarvan stond vorig schooljaar op meer dan een kwart van de lesdagen een invaller of wisselende leerkracht voor de klas. Daar groeiden leerlingen gemiddeld 5 punten op de vaardigheidsschaal van het leerlingvolgsysteem; in de andere 18 groepen was dat 9 punten. De maanden met invallers in groep 6 vallen in dit schooljaar en tellen niet mee. Zo toetst het bestuur de hypothese op andere gegevens dan die waarin het patroon opviel.

De kwaliteitsmedewerker van het bestuur doet de analyse. Directeuren en teams zien alleen uitkomsten over meerdere groepen samen.

Wat je eerst ziet

4 punten

minder groei in de zes groepen met veel invallers

Rekening gehouden met startniveau en samenstelling

2 punten

verschil in groei dat overblijft

Na een proef met vervanging

?

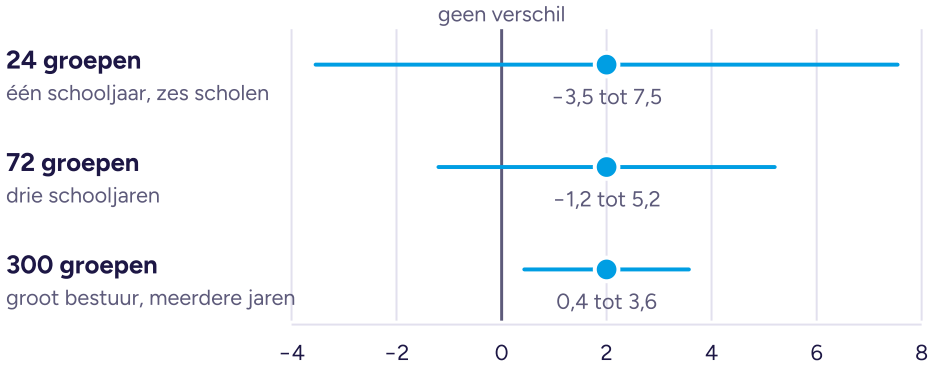
pas dan weet je of het helpt

1. **Wat verschilt er nog meer?** De groepen met veel invallers zitten vaker op scholen met veel leerlingen met een onderwijsachterstand, en hun leerlingen begonnen het jaar lager. Na correctie voor startniveau en samenstelling blijft van de 4 punten verschil in groei 2 punten over.
2. **Kan het toeval zijn?** Met 24 groepen schommelen gemiddelden makkelijk een paar punten, zeker omdat groepen op dezelfde school op elkaar lijken. Ook over drie schooljaren, dus 72 groepen, kan een verschil van 2 punten nog toeval zijn.
3. **Kan het andersom zijn?** Misschien viel een vaste leerkracht juist uit in een groep die al zwaar was. Het bestuur legt die vraag voor aan directeuren en intern begeleiders, die weten hoe de groepen het jaar begonnen.

4. **Helpt een vaste aanpak?** Het volgende schooljaar voert het bestuur een vaste aanpak voor vervanging in: per school een vaste invaller en een overdracht van de leesinstructie bij elke wissel. Het loot welke drie scholen in het eerste halfjaar beginnen; de andere drie volgen een halfjaar later. Na dat halfjaar vergelijkt het de leesgroei op de scholen die begonnen met die op de scholen die nog wachtten.

Groeien de groepen op de scholen die eerst begonnen duidelijk meer, dan heeft het bestuur een reden om de aanpak te houden. Ook zo'n verschil kan bij zes scholen nog toeval zijn.

Minder leesgroei bij veel invallers, in punten op de vaardigheidsschaal



Verzonnen voorbeeld. De stip is het verschil van 2 punten in leesgroei dat overblijft na correctie voor startniveau en samenstelling. De lijn is het betrouwbaarheidsinterval: de waarden voor het echte verschil die nog goed bij de gegevens passen. Hoe meer groepen, hoe korter de lijn. Raakt de lijn de nul, zoals bij 24 en 72 groepen, dan kan het verschil ook toeval zijn. Omdat groepen op dezelfde school op elkaar lijken, is de echte onzekerheid groter dan de figuur laat zien. Ook bij 300 groepen bewijst het verschil nog geen oorzaak.

Staan gegevens uit de vier pijlers op één plek, dan verken je zo'n hypothese zonder ze eerst uit verschillende systemen te halen. De afspraken en controles uit de stappen blijven nodig. Begin bij één verband dat in je team of bestuur al leeft, zoals de vraag over invallers. Dezelfde redenering geldt voor de grotere vraag van deze reeks: maakt werken met data scholen zelf beter?

3 · ONDERZOEKSOPZET

Zo toets je of data-ondersteund werken loont

Een bestuur dat jaren in werken met data investeert, wil op een dag weten of dat iets oplevert. Maar resultaten stijgen en dalen ook om andere redenen. Wie te snel conclusies trekt, schrijft een toevallige stijging toe aan de nieuwe aanpak, of stopt met een aanpak die wel werkte.

Om de hypothese van deze reeks te toetsen, heb je aan één school of één bestuur niet genoeg. Effecten in het onderwijs zijn meestal klein: in gerandomiseerde experimenten ligt het middelste effect rond 0,10 SD (Kraft, 2020). Bij een handvol scholen valt zo'n effect weg in de schommelingen die schoolgemiddelden volgens Amerikaans onderzoek van jaar tot jaar vertonen (Kane & Staiger, 2002).

Het verzonnen PO-bestuur met zes scholen merkte dat al bij de vraag over invallers: met 24 groepen bleef onzeker of het verschil meer was dan toeval. Wil je weten of datagebruik scholen beter maakt, dan heb je gegevens van veel scholen nodig, met per school een meting van hoe het team met data werkt. Binnen één bestuur kun je wel nagaan of een eigen aanpak werkt. De vier stappen hieronder gelden voor allebei.

1 Leg vooraf vast wat 'beter' is

Bepaal vooraf per sector welke uitkomsten tellen. Doe je dat pas achteraf, dan vind je al snel iets dat verbeterd lijkt. Ga uit van de definitie van een betere school: groei boven verwachting, voor alle groepen leerlingen, op brede doelen en over meerdere jaren.

In het PO is dat het percentage leerlingen dat in groep 8 1F en 1S/2F haalt (Staatscourant 2026, 26023), aangevuld met de leergroei op de vaardigheidsschaal van het leerlingvolgsysteem. In het VO liggen de resultaatindicatoren van de Inspectie voor de hand (Inspectie van het Onderwijs, 2025), in het (V)SO het behaalde uitstroomniveau ten opzichte van het OPP (Inspectie van het Onderwijs, z.j.-b). Gebruik steeds de onderliggende scores, zoals percentages of gemiddelden. Een oordeel als 'voldoende' of 'onvoldoende' verbergt veel van de verandering.

Kies per sector één hoofduitkomst en leg die vast in een gedateerd onderzoeksplan dat anderen kunnen inzien, samen met de vraag en de manier van analyseren. Doe dat voordat je de resultaten bekijkt; onderzoekers noemen dat preregistratie. Bekijk je twintig uitkomsten, dan valt er vaak wel één gunstig uit zonder dat er iets veranderd is. Andere uitkomsten mag je als aanvulling rapporteren, juist ook als ze tegenvallen.

Een gemiddelde kan stijgen terwijl de kloof tussen groepen leerlingen groeit. Volgens onze definitie is dat niet beter. Een aanpak kan ook per leerlingpopulatie anders uitpakken: in een evaluatie van het Focusproject boekten bij rekenen juist scholen met veel leerlingen met een lage sociaaleconomische status de grootste winst (Keuning e.a., 2019). Benoem de kloof daarom vooraf als uitkomst.

Middel over minstens drie jaargangen, zoals de Inspectie in het PO ook doet (Staatscourant 2026, 26023), en ga na of je conclusie overeind blijft met een iets ander rekenmodel. Toen Nederlandse onderzoekers de toegevoegde waarde van VO-scholen met verschillende modellen berekenden, kwamen individuele scholen per model anders uit (Timmermans e.a., 2011).

Tel tot slot de werkwijze die je toetst niet ook als uitkomst, want dan meet je twee keer hetzelfde. Het eindoordeel van de Inspectie is daarom lastig als uitkomst, want het omvat ook sturing, kwaliteitszorg en ambitie (Staatscourant 2026, 24891).

2 Meet datagebruik op meer dan één manier

Werken met data loopt van het openen van een overzicht tot het bijstellen van een doel in een teamoverleg, en elke meetwijze ziet daar maar een deel van. Combineer daarom drie invalshoeken.

- De eerste invalshoek is **wat mensen doen**. De gebruiksgegevens van een dataplatform laten zien wie welke overzichten opent en hoe vaak, niet of het team de gegevens daarna analyseert, er verbeteringen op baseert en die evalueert. Laat gegevens van de leverancier bij voorkeur verwerken door een onafhankelijke onderzoekspartner, want de leverancier heeft belang bij de uitkomst.

- De tweede invalshoek is **wat mensen zelf rapporteren** in een vragenlijst die al eerder in onderzoek is gebruikt, zoals de lijst waarmee Nederlandse onderzoekers ruim 1000 leraren in het VO vroegen hoeveel zij met data werkten (Schildkamp e.a., 2017). Laat een onafhankelijke onderzoekspartner de lijst anoniem afnemen. Leraren die weten dat de schoolleiding of de leverancier meekijkt, antwoorden misschien gunstiger dan de praktijk rechtvaardigt.
- De derde invalshoek is **wat een buitenstaander ziet**, zoals het oordeel van de Inspectie over kwaliteitszorg, waar dat er is (Staatscourant 2026, 24891), of een observatie van een databespreking door de onderzoekspartner.

  Controleer eerst of de meetwijzen samenhangen. Scoren scholen die het platform veel gebruiken ook hoog op de vragenlijst? Zo niet, dan meten ze iets anders en kun je ze niet zomaar optellen tot één score. Gebruiksgegevens heb je bovendien alleen van scholen met het platform. Voor een vergelijking met andere scholen is een maat nodig die overal af te nemen is, zoals de vragenlijst.

3 Kies een opzet die oorzaak en samenhang zo goed mogelijk scheidt

Presteren scholen met veel datagebruik beter, dan is dat op zichzelf alleen samenhang. Van een oorzaak spreek je pas als de betere prestaties door het datagebruik komen, en niet door iets anders waarin die scholen verschillen. De Engelse Education Endowment Foundation raadt aan te vergelijken met een groep die de aanpak (nog) niet krijgt, het liefst gevormd door loting. Lukt dat niet, dan met groepen die zo veel mogelijk op elkaar lijken, met een minder zekere uitkomst (Coe e.a., 2013).

- Een eerste mogelijkheid is **voor en na meten, met vergelijkingsscholen**: een quasi-experiment. Vergelijk de ontwikkeling van scholen vóór en na de start met die van vergelijkbare scholen die niet startten. Stegen de startende scholen vóór de start al sneller, dan kun je het verschil daarna niet aan het datagebruik toeschrijven.
- Sterker is **loting** over welke scholen de aanpak of extra begeleiding krijgen. Dan lijken de groepen gemiddeld op elkaar, ook op kenmerken die je niet kunt meten. Wil geen school de aanpak mislopen, laat het lot dan de volgorde van een gefaseerde invoering bepalen.
- Het zwakst is een vergelijking van **dosis en effect**: gaan scholen meer vooruit naarmate ze intensiever met data werken? Dat laat alleen samenhang zien, want die scholen kunnen ook op andere punten verschillen. Binnen het Focusproject ging succes samen met beter lesgeven (Keuning e.a., 2017). Volgens de onderzoekers was dat betere lesgeven er al vóór de aanpak. Of het de winst verklaart, toont zo'n vergelijking achteraf niet aan.

  Ga bij elke opzet na of de bedoelde verandering in de praktijk plaatsvond. In een Amerikaans experiment leidde een toetssysteem met training en coaching niet tot meer gedifferentieerd lesgeven (Cordray e.a., 2013). Meet je dat niet, dan weet je bij een tegenvallende uitkomst niet of je moet stoppen of de uitvoering moet verbeteren.

4 Benoem wat de uitkomst kan vertekenen

Ook een goede opzet kan een vertekend beeld geven. Benoem de belangrijkste risico's vooraf in het onderzoeksplan en schrijf erbij hoe je ermee omgaat.

- **Selectie.** Scholen die zelf voor werken met data kiezen, verschillen misschien al van andere scholen. Vergelijk met scholen die vóór de start op hen leken, of loot.
- **Omgekeerde richting.** Soms leidt de uitkomst tot het datagebruik. Een school met een onvoldoende oordeel van de Inspectie kan juist intensiever met data gaan werken, en scholen die het al goed doen kiezen misschien eerder voor een dataplatform. De ontwikkeling vóór de start laat zien of de resultaten al daalden of stegen.
- **Terugval naar het gemiddelde.** Na een uitzonderlijk slecht jaar gaan resultaten gemiddeld vanzelf omhoog, ook als de nieuwe aanpak niets doet. Vergelijk dan met scholen die een even slecht jaar hadden.
- **Nieuwigheid.** Teams die net met iets nieuws beginnen, doen soms tijdelijk extra hun best. In een Nederlands experiment met een feedback-app voor docenten in het VO verbeterden de lessen volgens de leerlingen eerst licht. Die verbetering hield geen stand (Bijlsma e.a., 2019). Waarom, staat niet vast, dus een nieuwigheidseffect is niet aangetoond. Wel blijkt dat een vroege winst nog geen blijvend effect is. Volg de uitkomsten daarom meerdere jaren.

- **Meting.** Uitkomsten die alleen in de eigen schoolgegevens staan, heb je van vergelijkingsscholen niet. Kies uitkomsten die van alle scholen bekend zijn.
- **Breuk in de reeks.** Vooral in het PO kan de maat zelf veranderen. De signaleringswaarden voor 1S/2F zijn per 1 augustus 2026 opnieuw vastgesteld, en vanaf 2026-2027 kijkt de Inspectie alleen nog naar de doorstroomtoets (Staatscourant 2026, 26023). Per 1 augustus 2027 wil de Inspectie een volledig herzien onderzoekskader invoeren (Inspectie van het Onderwijs, z.j.-c). Valt zo'n verandering samen met de start van het datagebruik, dan lijkt een sprong in de cijfers een effect. Vergelijk dan met scholen die dezelfde breuk doormaakten.
- **Kleine aantallen.** Op kleine scholen en in het (V)SO verschuiven een paar leerlingen meer of minder het gemiddelde al flink. Middel daar over meer jaren en trek geen conclusies uit één meting.

Je hoeft niet op een groot onderzoek te wachten. Het verzonnen bestuur kiest bij zijn vaste aanpak voor vervanging de leesgroei als uitkomst en loot welke drie van zijn zes scholen beginnen; de andere drie volgen een halfjaar later. Wat het zo leert, geldt alleen voor de eigen scholen. Toch helpt het om onderbouwd te beslissen over doorgaan of bijstellen.

4 · HOE TRIPS HELPT

Je gegevens op één plek, als basis voor het gesprek

Scholen in vijf Europese landen, waaronder Nederland, noemden weinig toegang tot goede gegevens, te weinig scholing en te weinig samenwerking rond data als belemmeringen (Schildkamp e.a., 2014b). TRIPS richt zich vooral op de toegang tot gegevens en wil het makkelijker maken om gegevens samen te bespreken. Scholing en vaste tijd voor samenwerking blijven werk van de school zelf.

TRIPS brengt de gegevens van een school of bestuur samen in één omgeving, voor PO, VO en (V)SO (TRIPS, z.j.). Het platform heeft drie modules.

Explore

Explore zet gegevens uit bronsystemen als ParnasSys, Somtoday en Magister in dashboards, met je eigen peilingen en observaties ernaast. Zo leg je de vier pijlers naast elkaar, voor zover je bronsystemen en peilingen die gegevens bevatten. De school bepaalt per rol wie welke gegevens ziet.

Guide

Guide ondersteunt de ontwikkelgesprekken met medewerkers, met afspraken en meetpunten in een digitaal ontwikkeldossier. Welke gegevens meetellen, bepaalt de school zelf.

Quest

Met enquêtes, lesobservaties, 360-gradenfeedback en zelfreflecties meet je wat niet in je bronsystemen staat, zoals tevredenheid, welbevinden en de kwaliteit van lessen.

In het Focusproject op Nederlandse basisscholen ging de grootste winst samen met kenmerken van het team, zoals beter lesgeven en een sterke datacultuur (Keuning e.a., 2017). Die levert geen platform. TRIPS wil zorgen dat de gegevens op tijd en compleet bij elkaar staan, zodat je minder zoekt en meer tijd hebt om te analyseren en te bespreken. Leg je bij een besluit de gegevens uit alle pijlers naast elkaar, dan kun je ook betere, data-informed keuzes maken.

De onderzoeken in deze reeks evalueren TRIPS zelf niet: ze gaan over werken met data in het algemeen. Wat hierboven over TRIPS staat, is wat we met het platform willen bereiken.

Meer over TRIPS lees je op trips-edu.nl.

VERANTWOORDING

Hoe deze whitepaper tot stand kwam

Hier lees je waar de uitspraken in deze whitepaper vandaan komen en hoe zeker ze zijn.

Hoe we de literatuur selecteerden

We zochten in meerdere rondes naar Nederlandse effectstudies en evaluaties, ook in documenten van de Inspectie, de Onderwijsraad, het CPB, het NRO, DUO en het CBS. Internationaal zochten we naar meta-analyses, gerandomiseerde experimenten en theoretische kaders over schoolkwaliteit en datagebruik. Na feedback op een eerdere versie zochten we vooral naar Nederlands of Europees onderzoek dat Amerikaanse studies kon vervangen.

Een bron kwam in aanmerking als ze over datagebruik of schoolkwaliteit ging en relevant was voor PO, VO of (V)SO. We kozen bij voorkeur studies met een sterke opzet, die met meer zekerheid laten zien of een aanpak iets oplevert. Nederlands onderzoek staat voorop, omdat een uitkomst uit een ander stelsel niet zomaar op een Nederlandse school geldt; internationaal onderzoek vult aan.

Hoe we de labels toekenden

Alle labels in de tabellen en op de bewijskaart kenden we zelf toe, op basis van de publicatie, een samenvatting ervan of, bij de laatst toegevoegde bronnen, zoekresultaten. De bewijskaart in deel 1 is een eigen samenvatting van TRIPS. Het label □ *niet onderzocht* betekent dat wij geen Nederlands onderzoek vonden; er kan onderzoek bestaan dat ons ontging. De kwaliteit van de uitvoering beoordeelden we niet systematisch per studie. Haken bij een experiment met loting veel deelnemers af, dan is de uitkomst minder zeker dan het label doet vermoeden.

Gebruik van AI

We gebruikten het taalmodel Claude van Anthropic om bronnen te zoeken en samen te vatten en om concepteksten te schrijven. Die concepten hebben we zelf in meerdere rondes bijgestuurd: wij bepaalden de vraag, de opbouw en de inhoudelijke keuzes, en gaven per hoofdstuk aan wat eruit moest, wat anders moest en wat ontbrak. Collega's met een onderzoeksachtergrond gaven feedback op een eerdere versie. De concepttekst die zo ontstond, hebben we zelf herschreven en tot de definitieve tekst gemaakt. De belangrijkste cijfers en conclusies hebben we daarna vergeleken met de oorspronkelijke publicaties, voor zover die openbaar toegankelijk waren.

BRONNEN

Gebruikte bronnen

1. Bernhardt, V.L. (1998). Multiple Measures. California ASCD, Invited Monograph No. 4. www.oregon.gov
2. Bijlsma, H.J.E., Visscher, A.J., Dobbelaer, M.J. & Veldkamp, B.P. (2019). Does smartphone-assisted student feedback affect teachers' teaching quality? Technology, Pedagogy and Education, 28(2). ris.utwente.nl
3. Coe, R., Kime, S., Nevill, C. & Coleman, R. (2013). The DIY Evaluation Guide. Londen: Education Endowment Foundation. educationendowmentfoundation.org.uk
4. Cordray, D. e.a. (2013). Impact of the Measures of Academic Progress (MAP) program on student reading achievement. NCEE 2013-4000. ies.ed.gov
5. De Witte, K. & Csillag, M. (2014). Does anybody notice? On the impact of improved truancy reporting on school dropout. Education Economics, 22(6), 549-568. doi.org
6. Den Brok, P., Brekelmans, M. & Wubbels, T. (2004). Interpersonal teacher behaviour and student outcomes. School Effectiveness and School Improvement, 15(3-4), 407-442.
7. Engzell, P., Frey, A. & Verhagen, M.D. (2021). Learning loss due to school closures during the COVID-19 pandemic. Proceedings of the National Academy of Sciences, 118(17), e2022376118. doi.org
8. Gerritsen, S., Plug, E. & Webbink, D. (2017). Teacher quality and student achievement: Evidence from a sample of Dutch twins. Journal of Applied Econometrics, 32(3), 643-660. doi.org
9. Gibbons, S., Scrutinio, V. & Telhaj, S. (2021). Teacher turnover: Effects, mechanisms and organisational responses. Labour Economics, 73, 102079. doi.org
10. Hendriks, M., Luyten, H., Scheerens, J. & Sleegers, P. (2014). Meta-analyses. In J. Scheerens (red.), Effectiveness of time investments in education: Insights from a review and meta-analysis (pp. 55-142). Cham: Springer. doi.org
11. Inspectie van het Onderwijs (2025). Technische toelichting onderwijsresultaten voortgezet onderwijs 2025. www.onderwijsinspectie.nl
12. Inspectie van het Onderwijs. Onderwijsresultaten (v)so. www.onderwijsinspectie.nl
13. Inspectie van het Onderwijs. Onderzoekskader 2027 (themapagina, geraadpleegd oktober 2026). www.onderwijsinspectie.nl
14. Jonk, A. & De Wolf, I. (2025). Lerarentekort vooral bij scholen met veel achterstandskinderen. Economisch Statistische Berichten, 110(4850). esb.nu
15. Kane, T.J. & Staiger, D.O. (2002). The promise and pitfalls of using imprecise school accountability measures. Journal of Economic Perspectives, 16(4), 91-114. doi.org
16. Keppens, G. (2023). School absenteeism and academic achievement: Does the timing of the absence matter? Learning and Instruction, 86, 101769. doi.org
17. Keppens, G., Spruyt, B. & Dockx, J. (2019). Measuring school absenteeism: administrative attendance data collected by schools differ from self-reports in systematic ways. Frontiers in Psychology, 10, 2623. doi.org
18. Keuning, T., Van Geel, M. & Visscher, A. (2017). Why a data-based decision-making intervention works in some schools and not in others. Learning Disabilities Research & Practice, 32(1), 32-45. cris.maastrichtuniversity.nl
19. Keuning, T., Van Geel, M., Visscher, A. & Fox, J.-P. (2019). Assessing and validating effects of a data-based decision-making intervention on student growth for mathematics and spelling. Journal of Educational Measurement. research.utwente.nl
20. Kraft, M.A. (2020). Interpreting effect sizes of education interventions. Educational Researcher, 49(4), 241-253. doi.org
21. Lomos, C., Hofman, R.H. & Bosker, R.J. (2011). Professional communities and student achievement: a meta-analysis. School Effectiveness and School Improvement, 22(2), 121-148. doi.org
22. Meyer, E. & Van Klaveren, C. (2013). The effectiveness of extended day programs: Evidence from a randomized field experiment in the Netherlands. Economics of Education Review, 36, 1-11. www.sciencedirect.com
23. Miller, R.T., Murnane, R.J. & Willett, J.B. (2008). Do teacher absences impact student achievement? Longitudinal evidence from one urban school district. Educational Evaluation and Policy Analysis, 30(2), 181-200. doi.org
24. Moolenaar, N.M., Sleegers, P.J.C. & Daly, A.J. (2012). Teaming up: Linking collaboration networks, collective efficacy, and student achievement. Teaching and Teacher Education, 28(2), 251-262. doi.org
25. Nakamoto, J. & Schwartz, D. (2010). Is peer victimization associated with academic achievement? A meta-analytic review. Social Development, 19(2), 221-242. doi.org
26. Roorda, D.L., Koomen, H.M.Y., Spilt, J.L. & Oort, F.J. (2011). The influence of affective teacher-student relationships on students' school engagement and achievement: a meta-analytic approach. Review of Educational Research, 81(4), 493-529. doi.org
27. Schildkamp, K. & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. Teaching and Teacher Education, 26(3), 482-496. doi.org
28. Schildkamp, K., Handelzalts, A., Poortman, C.L., Leusink, H., Meerdink, M., Smit, M., Ebbeler, J. & Hubers, M.D. (2014). De datateam-methode: een concrete aanpak voor onderwijsverbetering. Antwerpen/Apeldoorn: Garant.
29. Schildkamp, K., Karbautzki, L. & Vanhoof, J. (2014). Exploring data use practices around Europe. Studies in Educational Evaluation, 42, 15-24. doi.org
30. Schildkamp, K., Poortman, C.L., Luyten, H. & Ebbeler, J. (2017). Factors promoting and hindering data-based decision making in schools. School Effectiveness and School Improvement, 28(2), 242-258. doi.org
31. Simpson, A. (2017). The misdirection of public policy: comparing and combining standardised effect sizes. Journal of Education Policy, 32(4), 450-466. doi.org
32. Staatscourant 2026, 24891. Onderzoekskader 2021, versie per 1 augustus 2026. zoek.officielebekendmakingen.nl
33. Staatscourant 2026, 26023. Signaleringswaarden leerresultaten primair onderwijs. zoek.officielebekendmakingen.nl
34. Timmermans, A., Doolaard, S. & De Wolf, I. (2011). Conceptual and empirical differences among various value-added models. School Effectiveness and School Improvement, 22(4), 393-413. dare.uva.nl
35. TRIPS. Ons product. trips-edu.nl
36. Van de Grift, W. (2007). Quality of teaching in four European countries: a review of the literature and application of an assessment instrument. Educational Research, 49(2), 127-152. doi.org

37. Van de Grift, W.J.C.M., Chun, S., Maulana, R., Lee, O. & Helms-Lorenz, M. (2017). Measuring teaching quality and student engagement in South Korea and The Netherlands. School Effectiveness and School Improvement, 28(3), 337-349. doi.org
38. Van der Vinne, V., Zerbini, G., Siersema, A., Pieper, A., Merrow, M., Hut, R.A., Roenneberg, T. & Kantermann, T. (2015). Timing of examinations affects school performance differently in early and late chronotypes. Journal of Biological Rhythms, 30(1), 53-60. doi.org
39. Van Kuijk, M.F. (2014). Raising the bar for reading comprehension. The effects of a teacher professional development program targeting goals, data use, and instruction (proefschrift). Rijksuniversiteit Groningen. www.rug.nl
40. Van Wickeren e.a. (2025). Het lerarentekort is geen landelijk maar een schoolspecifiek probleem. Economisch Statistische Berichten.
41. Zerbini, G., Van der Vinne, V., Otto, L.K.M., Kantermann, T., Krijnen, W.P., Roenneberg, T. & Merrow, M. (2017). Lower school performance in late chronotypes: underlying factors and mechanisms. Scientific Reports, 7, 4385. doi.org