Whitepaper · november 2026



DATA-ONDERSTEUND WERKEN · DEEL 1 VAN 2

Werkt werken met data?

Wat onderzoek zegt over datagebruik en schoolkwaliteit, en wat een betere school eigenlijk is. Voor schoolleiders en bestuurders in PO, VO en (V)SO. In deel 2 lees je hoe je een verband in je eigen gegevens eerlijk toetst en wat er nodig is om de hypothese van deze whitepaper te toetsen.

Inhoud

01 In het kort

02 Presteren scholen die data-ondersteund werken beter?

03 Wat is een betere school?

04 Wat zegt Nederlands onderzoek?

05 Voorwaarden en valkuilen

06 Eén pijler zegt te weinig

07 Verantwoording

08 Bronnen

IN HET KORT

1 **Een systeem of verplichting alleen maakt een school niet aantoonbaar beter.** Op 79 Nederlandse basisscholen hing het gebruik van een zelfevaluatie-instrument over vijf jaar niet samen met betere leerprestaties (Schildkamp e.a., 2009). Mogelijk komt dat doordat zo'n instrument wel gegevens oplevert, maar niet bepaalt wat een team er in de klas mee doet. Stuur je op één cijfer of registreer je meer dan je gebruikt, dan kan dat bijeffecten hebben, zoals onnodige werkdruk of minder aandacht voor de zwakste leerlingen (Onderwijsraad, 2026; Neal & Schanzenbach, 2010; Data Management Review Group, 2016).

2 **In Nederlandse projecten in het PO leverde leren werken met data soms bescheiden winst op en soms gemiddeld geen effect.** Het Focusproject van de Universiteit Twente is drie keer geëvalueerd. Twee evaluaties vergeleken met eerdere jaargroepen op dezelfde scholen en vonden ongeveer een maand extra leerwinst bij rekenen, de tweede ook bij spelling (Van Geel e.a., 2016; Keuning e.a., 2019). De derde, van een eerdere ronde, vergeleek met soortgelijke scholen zonder de aanpak en vond gemiddeld geen effect op rekenen (Staman e.a., 2017). In het leesproject Streef liepen leerlingen bij begrijpend lezen ruim een half jaar voor op vergelijkbare klassen (Van Kuijk, 2014). Geen van deze studies werkte met loting, dus je kunt de winst niet met zekerheid aan de aanpak toeschrijven. In een experiment met loting kregen leraren van groep 6 een intensieve training in werken met data. Gemiddeld had dat geen effect op de rekenprestaties, wel voor leerlingen die verlengde instructie kregen (Van der Scheer & Visscher, 2018). Voor VO en (V)SO vonden we nauwelijks Nederlands effectonderzoek.

3 **Werken met data is één onderdeel van evidence-informed werken en vraagt meer dan analyseren.** Bij evidence-informed werken combineer je schoolgegevens met inzichten uit onderzoek en met de ervaring van je team (Brown e.a., 2017). Op Nederlandse basisscholen bleek de stap van analyse naar doel en instructie lastig (Inspectie van het Onderwijs, 2010; Staman e.a., 2017). Of een team data gebruikt, hangt het sterkst samen met de schoolorganisatie: visie, leiderschap en ondersteuning. Ook de samenwerking en kennis in het team en de kwaliteit en toegankelijkheid van de gegevens tellen mee (Schildkamp e.a., 2017; Hoogland e.a., 2016; Schildkamp e.a., 2014b). TRIPS

brengt gegevens uit verschillende systemen samen in één overzicht, met als doel dat het team minder tijd kwijt is aan zoeken en meer aan bespreken. Onderzoek naar die beoogde tijdwinst vonden we niet.

4 **Een betere school laat leerlingen uit alle groepen meer vooruitgaan dan je op grond van hun startpunt en achtergrond mag verwachten, en meer dan leerlingen op dezelfde school in eerdere jaren deden.** De definitie is onze keuze. We vergelijken met de instroom, zoals onderzoek naar schooleffectiviteit doet (Scheerens, 2000). Uit het bredere begrip onderwijskwaliteit nemen we de ontwikkeling als burger en als persoon en een veilig klimaat mee, met de basiskwaliteit van de Inspectie als ondergrens (Onderwijsraad, 2016; Staatscourant 2026, 24891). Omdat het om alle groepen gaat, kijk je ook of de kloof tussen groepen kleiner wordt; in onderzoek zijn gemiddelde kwaliteit en kansengelijkheid twee aparte maten (Kyriakides & Creemers, 2011).

5 **Eén soort gegevens zegt te weinig.** In het onderwijs blijft datagebruik vaak beperkt tot toetsresultaten (Schildkamp, 2019). Die laten zien dat er iets speelt; de verklaring kan in andere gegevens liggen. Een verzonnen voorbeeld: in groep 6 daalt de tevredenheid over de lessen en blijft de groei in begrijpend lezen achter. Pas naast de bezetting zie je dat er van november tot maart drie verschillende invallers voor de klas stonden. Ook die verklaring blijft een vermoeden tot je haar toetst.

6 **De hypothese van deze whitepaper is nog niet met een sterke opzet getoetst.** Ze luidt: *scholen die data gebruiken om hun beleid te onderbouwen, presteren beter dan scholen die dat niet doen*. Het meeste onderzoek gaat over leraren en teams die leren werken met data (Van Geel e.a., 2016; Ansyari e.a., 2022; Spiele e.a., 2025). Naar datagebruik voor schoolbeleid en door besturen vonden we nauwelijks onderzoek. Hoe je haar wel kunt toetsen, met gegevens van veel scholen en een meting van hun datagebruik, lees je in deel 2.

1 · AANLEIDING

Presteren scholen die data-ondersteund werken beter?

Scholen leggen veel vast: toetsresultaten, verzuim, incidenten, tevredenheidspeilingen, ontwikkelingsperspectieven en gespreksverslagen. Achter al dat registreren zit de verwachting dat beleid beter wordt als het op die gegevens rust. Verzamelen en bespreken kost je team tijd. Of die verwachting op jouw school uitkomt, is lastig te zien.

In deze whitepaper leggen we die verwachting naast het onderzoek. We werken met één hypothese: *scholen die data gebruiken om hun beleid te onderbouwen, presteren beter dan scholen die dat niet doen.*

Wat we onder werken met data verstaan

Schildkamp en Kuiper (2010) omschrijven werken met data als het systematisch analyseren van gegevens die de school al heeft, om haar onderwijs, curriculum en prestaties te verbeteren. Die verbeteringen voert de school uit en evalueert ze. Wij volgen die omschrijving; data-ondersteund werken betekent hier hetzelfde.

Schildkamp (2019) beschrijft het als een cyclus. Een team stelt eerst een doel. Dan verzamelt het formele gegevens zoals toetsresultaten, informele zoals wat een leraar in de les opmerkt, en uitkomsten van onderzoek. Daarna geeft het team er samen betekenis aan, voert het verbeteracties uit en evalueert die. Begin als school bij een eigen vraag en zoek daar de gegevens bij.

Evidence-informed werken is ruimer. Volgens Brown e.a. (2017) combineer je dan gegevens van je eigen school met inzichten uit onderzoek en met de professionele ervaring van je team. Dat is een conceptueel voorstel waarvan het effect nog niet is getoetst. Je eigen gegevens laten zien wat er op school gebeurt; onderzoek helpt verklaren waarom. Daarom zien we werken met data als één onderdeel van evidence-informed werken.

Waarom de vraag nu speelt

De Inspectie werkt aan een volledig herzien onderzoekskader voor PO, VO en (V)SO, dat volgens planning op 1 augustus 2027 ingaat (Inspectie van het Onderwijs, z.j.-c; Inspectie van het Onderwijs, 2026b). In december 2025 schreef de staatssecretaris van OCW aan de Tweede Kamer dat daarin de directe koppeling vervalt tussen een onvoldoende voor onderwijsresultaten (standaard OR1) en het eindoordeel over een school. De criteria voor het eindoordeel Onvoldoende werkt de Inspectie nog uit (Kamerstukken II 31 293; Inspectie van het Onderwijs, 2026b). Je weet dus nog niet langs welke meetlat ze je kwaliteitszorg legt. Dan helpt het als je als school en bestuur zelf al hebt uitgewerkt wat je onder een goede school verstaat.

> 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 beperken we ons tot uitspraken die het onderzoek draagt. Hoe we AI (Claude, van Anthropic) hebben gebruikt, lees je in de Verantwoording.

2 · DEFINITIE

Wat is een betere school?

Stel, je legt als bestuurder de resultaten van twee scholen naast elkaar. De ene staat in een welvarende wijk en haalt hoge scores. De andere heeft veel leerlingen met laagopgeleide ouders en scoort rond het gemiddelde. Welke het beter doet, lees je maar ten dele uit die scores af.

Hoge scores zeggen voor een groot deel iets over de leerlingen die een school binnenkrijgt. In Australisch onderzoek werden de verschillen tussen scholen veel kleiner zodra de onderzoekers rekening hielden met eerdere toetsresultaten van leerlingen (Marks, 2015). De tweede school kan

meer toevoegen. Wil je toetsen of werken met data een school beter maakt, leg dan eerst vast wat 'beter' is en waarmee je vergelijkt.

Bij schooleffectiviteit zet je wat een school bereikt af tegen wat je op grond van haar instroom mag verwachten (Scheerens, 2000). Zo kom je uit bij wat de school zelf toevoegt: de toegevoegde waarde. Onderwijskwaliteit is breder. De Inspectie beoordeelt naast resultaten ook het onderwijsproces, de veiligheid en het schoolklimaat, en de sturing, kwaliteitszorg en ambitie van een school (Staatscourant 2026, 24891). Een school die veel toevoegt aan leerresultaten, kan op andere doelen of in haar onderwijsproces toch tekortschieten.

Wij combineren drie vergelijkingen, want elke vergelijking heeft een zwakke plek. De eerste zet wat leerlingen bereiken af tegen scholen met vergelijkbare leerlingen. Die toegevoegde waarde is geen eenduidige maat: bij Nederlandse VO-scholen hingen de uitkomsten van verschillende rekenmodellen sterk samen, maar kwamen afzonderlijke scholen per model anders uit (Timmermans e.a., 2011). De tweede vergelijkt een school met zichzelf in eerdere jaren, maar zegt niets over wat de leerlingen op een andere school hadden bereikt. De derde neemt de normen van de Inspectie als ondergrens: een school die aan de wettelijke eisen voldoet, heeft basiskwaliteit. Die normen zeggen weinig over wat een school daarboven toevoegt.

Een betere school is een school waar leerlingen meer vooruitgaan dan je op grond van hun startpunt en achtergrond mag verwachten, en meer dan leerlingen op die school in eerdere jaren deden. Die vooruitgang gaat over kennis en vaardigheden, over hun ontwikkeling als burger en als persoon en over hun welbevinden. Ze moet voor alle groepen leerlingen gelden en meerdere jaren standhouden, in een veilig schoolklimaat en met de basisnormen van de Inspectie als ondergrens.

'Wat je mag verwachten' is wat leerlingen met een vergelijkbaar startpunt en een vergelijkbare achtergrond, zoals het opleidingsniveau van de ouders, op andere scholen gemiddeld bereiken. De maatstaf is er om scholen eerlijk te vergelijken. Hij betekent niet dat je van sommige leerlingen minder verwacht. Door de vergelijking met eerdere jaren gaat de definitie over beter worden. Een school die al jaren ruim boven verwachting presteert en dat niveau vasthoudt, is volgens onze definitie niet beter geworden. Een sterke school blijft ze wel.

De definitie is onze keuze. Ze bouwt voort op het interdepartementale beleidsonderzoek *Sturing op kwaliteit van onderwijs*. Daarin is kwaliteit de mate waarin het onderwijs eraan bijdraagt dat elke leerling, ongeacht afkomst, zijn of haar mogelijkheden in de volle breedte kan benutten (Interdepartementaal beleidsonderzoek, 2022). Die breedte verdelen we in vijf onderdelen.

1. Groei boven verwachting

Je vergelijkt de groei van je leerlingen met die van vergelijkbare leerlingen elders (Melhuish e.a., 2006; Scheerens, 2000). Zo kom je in de buurt van wat de school zelf toevoegt, al is het geen zuivere meting van haar effect (Raudenbush & Willms, 1995).

2. Voor alle leerlingen

Een stijgend gemiddelde kan samengaan met een groeiende kloof tussen groepen leerlingen. Daarom behandelen we kwaliteit en kansengelijkheid als twee aparte maten. In twee studies in Cyprus ging vooruitgang op de ene maat niet ten koste van de andere (Kyriakides & Creemers, 2011). Kijk naast het gemiddelde ook of de verschillen in groei tussen groepen kleiner worden.

3. Brede doelen

Biesta (2009) onderscheidt drie doelen van onderwijs. De Onderwijsraad noemt ze kwalificatie (kennis en vaardigheden), socialisatie (ingroeien in de samenleving, waaronder burgerschap) en persoonsvorming (bijvoorbeeld zelfstandigheid, zelfvertrouwen en een eigen oordeel) (Biesta, 2009; Onderwijsraad, 2016). Burgerschap is in het PO, VO en (V)SO een wettelijke opdracht (Staatsblad 2021, 320).

4. Veilig klimaat en welbevinden

In PISA 2015 bleken 15-jarigen in sommige landen met hoge gemiddelde scores relatief ontevreden over hun leven (OECD, 2017). Goede cijfers zeggen dus niet automatisch iets over hoe leerlingen zich voelen; welbevinden meet je apart. Scholen in het PO, VO en (V)SO moeten sociale veiligheid jaarlijks monitoren, en de Inspectie ziet daarop toe (VO-raad, z.j.).

5. Volhouden

In het Australische onderzoek verschilde wat een school toevoegde van jaar tot jaar en van vak tot vak (Marks, 2015). Kijk daarom naar de ontwikkeling over minstens drie jaar. De Inspectie werkt in het PO en VO ook met driejaarsgemiddelden (Staatscourant 2026, 26023; Inspectie van het Onderwijs, 2025).

Wat we wel en niet als uitkomst tellen

Welbevinden tellen we als uitkomst: net als kennis of zelfvertrouwen is het een toestand of ontwikkeling van de leerling, geen oordeel over de school. Ervaren autonomie, het gevoel dat je zelf invloed hebt op wat je doet, rekenen we tot persoonsvorming. Volgens de zelfdeterminatietheorie gaat onderwijs dat de behoefte aan autonomie, competentie en verbondenheid ondersteunt, samen met meer motivatie en welzijn (Ryan & Deci, 2020). Een Groningse review kwam voor jonge tieners tot een vergelijkbaar beeld (Stroet e.a., 2013). Een dalende ervaren autonomie is ook een signaal dat de motivatie van leerlingen onder druk kan staan.

Tevredenheid van leerlingen, ouders en medewerkers tellen we niet als uitkomst. Het is een oordeel over de school, geen ontwikkeling van de leerling. Als bron telt tevredenheid wel: daalt in één groep de tevredenheid over de lessen, kijk dan ook naar de leergroei en het welbevinden in die groep.

Kwaliteitszorg, personeelsbeleid en de gesprekscyclus tellen we evenmin als uitkomst, al beoordeelt de Inspectie kwaliteitszorg wel als onderdeel van onderwijskwaliteit (Staatscourant 2026, 24891). Het zijn werkwijzen. Tel je ze mee, dan noem je een school al beter zodra ze een systeem heeft, ook als haar leerlingen er niet op vooruitgaan. Werken met data is ook zo'n werkwijze, en de vraag is juist of die leerlingen iets oplevert.

Meten per sector

Welke gegevens je per onderdeel kunt gebruiken, verschilt per sector, want de Inspectie werkt in het PO, VO en (V)SO met andere maten.

Onderdeel	PO	VO	(V)SO
Groei boven verwachting	Percentage leerlingen op of boven 1F en 1S/2F, over drie jaar. De signaleringswaarde voor 1S/2F hangt af van de schoolweging van het CBS (Staatscourant 2026, 26023; CBS, z.j.-a). Leergroei uit het leerlingvolgsysteem. Rapportage van het Nationaal Cohortonderzoek Onderwijs (NCO) (CBS, z.j.-b).	Onderwijspositie in leerjaar 3 ten opzichte van het advies, onderbouwsnelheid, bovenbouwsucces en examencijfers (Inspectie van het Onderwijs, 2025). NCO-rapportage.	Behaald uitstroomniveau ten opzichte van het ontwikkelingsperspectief (OPP). De school stelt het OPP zelf vast (Inspectie van het Onderwijs, z.j.-b).
Voor alle leerlingen	Verschil in groei tussen groepen leerlingen.	Onderwijspositie per groep.	Aandeel leerlingen dat het OPP-niveau haalt, per groep.
Brede doelen	Eigen doelen voor burgerschap en persoonsvorming, met een meting daarvan, bijvoorbeeld een vragenlijst naar zelfvertrouwen of ervaren autonomie. Voor persoonsvorming vonden we geen landelijke maat.		
Veilig klimaat en welbevinden	Jaarlijkse monitor sociale veiligheid (VO-raad, z.j.), eventueel aangevuld met een eigen peiling naar welbevinden.		
Volhouden	Driejaarscijfers van de Inspectie (Staatscourant 2026, 26023).	Jaarscores en driejaarsgemiddelden (Inspectie van het Onderwijs, 2025).	Uitstroom over meerdere jaren, omdat het per jaar om kleine aantallen gaat.

Mogelijke indicatoren per onderdeel en sector, door TRIPS bij elkaar gezet en niet gevalideerd als meetmodel.

Voor kansengelijkheid, persoonsvorming en welbevinden ben je grotendeels aangewezen op wat de school zelf verzamelt. Ook bij groei boven verwachting heeft alleen de school of het bestuur een deel van de gegevens, zoals de vertrouwelijke NCO-rapportages (CBS, z.j.-b). Leg daarom als eerste stap per onderdeel vast welke gegevens je gebruikt en waarmee je ze vergelijkt: scholen met vergelijkbare leerlingen, eerdere jaren of de norm van de Inspectie.

3 · HET BEWIJS

Wat zegt Nederlands onderzoek?

Steek je tijd en geld in datateams, een nieuw leerlingvolgsysteem of een dashboard, dan wil je weten wat je leerlingen eraan hebben. Hier zetten we op een rij wat Nederlands onderzoek zegt en hoe stevig dat bewijs is.

De hoofdvraag is: helpt werken met data een school om leerlingen meer te laten vooruitgaan dan je op grond van hun startpunt en achtergrond mag verwachten? We beginnen bij Nederlandse scholen, want ons stelsel van toetsen en toezicht wijkt af van dat in andere landen. Internationaal onderzoek vult aan waar Nederlands onderzoek ontbreekt. Het meeste Nederlandse effectonderzoek gaat over leerresultaten in het PO. Over VO en (V)SO weten we veel minder, net als over welbevinden, veiligheid, burgerschap en persoonsvorming.

Zo lees je het onderzoek in deze whitepaper

In een gerandomiseerd experiment bepaalt het lot welke scholen, leraren of leerlingen de aanpak krijgen. De groepen lijken dan gemiddeld op elkaar, dus een verschil kun je met meer zekerheid aan de aanpak toeschrijven. Een quasi-experiment vergelijkt zonder loting, met andere scholen of met eerdere jaren. Dan kunnen de groepen ook op andere punten verschillen: scholen die zich vrijwillig aanmelden, zijn misschien al extra gemotiveerd.

Veel onderzoek laat alleen samenhang zien. Neem opbrengstgericht werken: stelselmatig toetsresultaten analyseren en er doelen en instructie op afstemmen. Presteren zulke scholen beter, dan kan dat door de werkwijze komen. Het kan ook andersom: goed lopende scholen gaan eerder zo werken. Of een derde factor, zoals de leerlingpopulatie, verklaart allebei.

Effecten staan vaak in standaarddeviaties (SD), een maat voor de spreiding van toetsscores. Bij 0,10 SD scoort de groep met de aanpak gemiddeld een tiende standaarddeviatie hoger dan de vergelijkingsgroep. Dat is ongeveer het mediane effect op gestandaardiseerde toetsen in gerandomiseerde experimenten met onderwijsinterventies, vooral Amerikaanse en Britse (Kraft, 2020). Het is een grof ijkpunt: op toetsen die onderzoekers zelf maakten, zijn effecten meestal groter dan op brede gestandaardiseerde toetsen (Kraft, 2020). Nederlandse projecten drukken de winst soms uit in maanden onderwijs (Van Geel e.a., 2016). Die kun je niet direct vergelijken met een effect in SD.

Elke studie in de tabellen krijgt een label voor de opzet: ■ **sterk** voor een gerandomiseerd experiment of een meta-analyse (een samenvoeging van veel studies tot één gemiddeld effect) van studies met een causale opzet, ■ **matig** voor een quasi-experiment en ■ **beperkt** voor onderzoek naar samenhang, casusonderzoek, evaluatie of literatuuronderzoek. Het label zegt iets over de opzet, niets over de omvang van het effect.

De bewijskaart vat per vakje, een combinatie van sector en thema, samen welk Nederlands onderzoek we vonden. Let vooral op ■ **geen effect aangetoond**: dan is een effect gezocht en niet gevonden. Dat is gebrek aan bewijs, iets anders dan bewijs dat er geen effect is. De andere labels staan in de legenda onder de kaart.

De Nederlandse bewijskaart

De bewijskaart is een eigen samenvatting van TRIPS, op basis van de Nederlandse studies in de tabel na de kaart. Niet elk vakje gaat over werken met data in strikte zin. Bij uitval, sociale veiligheid en het (V)SO gaat het om evaluaties van afspraken en wetten waarin gegevens een rol spelen. Bij 'niet onderzocht' heb je weinig houvast, dus wees daar voorzichtig met verwachtingen.

Thema	PO	VO	(V)SO
Leerresultaten na leren werken met data	■ **gemengd** Twee evaluaties van het Focusproject en het leesproject Streef vonden winst, zonder loting (Van Geel e.a., 2016; Keuning e.a., 2019; Van Kuijk, 2014). Een eerdere Focusronde vond, vergeleken met vergelijkbare scholen, geen hoofdeffect (Staman e.a., 2017). Een experiment met loting vond gemiddeld geen effect op rekenen, wel voor leerlingen met verlengde instructie (Van der Scheer & Visscher, 2018).	■ **casusonderzoek** Vier van de negen datateams losten hun zelfgekozen probleem op, een vijfde deels (Poortman & Schildkamp, 2016).	■ **beschrijvend** Professionals vinden dat het OPP helpt bij planmatig werken. Leerlingresultaten mat de evaluatie niet (KBA Nijmegen & ResearchNed, 2024).
Uitstroom, uitval en verzuim	□ *niet onderzocht*	■ **geen effect aangetoond** Afspraken met 14 van de 39 regio's over minder uitval hadden in het eerste jaar geen meetbaar effect (Van der Steeg e.a., 2008).	■ **beschrijvend** Of passend onderwijs leerlingen vooruit heeft geholpen, kon de evaluatie niet vaststellen (Ledoux e.a., 2020).
Sociale veiligheid en welbevinden	■ **beschrijvend** De verplichte monitor uit de Wet veiligheid op school zette veiligheid op de agenda. De wet zelf vinden betrokkenen niet doorslaggevend voor een veiliger klimaat (MOOZ & Ecorys, 2021). Nederlands onderzoek naar het effect van werken met data op welbevinden ontbreekt.		
Lesgeven en professionele dialoog	■ **samenhang** Op 53 basisscholen hing hechtere samenwerking tussen leraren via gedeeld vertrouwen samen met betere leerresultaten (Moolenaar e.a., 2012).	■ **gemengd** Een begeleidingsprogramma voor startende leraren versterkte hun lesvaardigheid (Helms-Lorenz e.a., 2016). Een feedback-app voor leraren had geen blijvend effect (Bijlsma e.a., 2019).	□ *niet onderzocht*
Kwaliteitszorg en toezicht	■ **samenhang** Opbrengstgericht werken en sommige vormen van zelfevaluatie gaan samen met betere resultaten (Inspectie van het Onderwijs, 2010; Hofman e.a., 2009). Voor het gebruik van een zelfevaluatie-instrument gold dat niet (Schildkamp e.a., 2009).	■ **geen effect aangetoond** Een kennisoverzicht vond geen effect van een kwaliteitszorgsysteem op zich (NRO-Kennisrotonde, 2018).	□ *niet onderzocht*

■ **gemengd** deels positieve uitkomsten in experimenten of quasi-experimenten ■ **geen effect aangetoond** effect gezocht en niet gevonden ■ **samenhang** , ■ **casusonderzoek** , ■ **beschrijvend** alleen onderzoek zonder experimentele vergelijking □ *niet onderzocht* geen Nederlands onderzoek gevonden

De Nederlandse studies

Voor het PO vonden we één Nederlands experiment met loting naar werken met data (Van der Scheer & Visscher, 2018). De andere effectstudies in het PO zijn quasi-experimenten, waarvan drie over het Focusproject. Twee andere Nederlandse experimenten met loting gaan over lesgeven in het VO (Helms-Lorenz e.a., 2016; Bijlsma e.a., 2019).

Studie	Wat is onderzocht	Opzet	Uitkomst	Bewijskracht
PRIMAIR ONDERWIJS (PO)				
Van der Scheer & Visscher (2018)	Intensieve training in werken met data voor leraren van groep 6	Gerandomiseerd experiment	Gemiddeld geen effect op de rekenprestaties. Voor leerlingen die verlengde instructie kregen, was het effect wel positief.	■ sterk
Staman e.a. (2017), Focus, eerste ronde	Training in het gebruik van gegevens uit het leerlingvolgsysteem voor instructie op maat, 42 basisscholen	Quasi-experimenteel, vergeleken met 42 vergelijkbare scholen zonder de aanpak	Geen hoofdeffect op rekenen. Gunstiger voor leerlingen met een lage begintoets of een lage sociaaleconomische status. Leraren vonden het moeilijk om voortgangsgegevens te vertalen naar instructie op maat.	■ matig
Van Geel e.a. (2016), Focusproject	Twee jaar schoolbrede training in werken met data, 53 basisscholen	Quasi-experimenteel, vergeleken met eerdere jaargroepen op dezelfde scholen, zonder externe controlegroep	Ongeveer een maand extra leerwinst bij rekenen, het enige gemeten vak. De auteurs concluderen dat vooral scholen met veel leerlingen met een lage sociaaleconomische status profiteerden. In hun eigen analyse was dat verschil niet significant.	■ matig
Keuning e.a. (2019), Focusproject	Teamtraining in werken met data, 39 basisscholen	Quasi-experimenteel, vergeleken met eerdere jaargroepen op dezelfde scholen, zonder externe controlegroep	Ongeveer een maand extra leerwinst bij rekenen en bij spelling. Bij rekenen was de winst het grootst op scholen met veel leerlingen met een lage sociaaleconomische status.	■ matig
Van Kuijk (2014), Streef	Doelen stellen, toetsgegevens gebruiken en de leesinstructie verbeteren, 33 leraren van groep 4 en 5 op 19 scholen	Quasi-experimenteel, vergeleken met vergelijkbare klassen	Leerlingen liepen bij begrijpend lezen ruim een half jaar voor op vergelijkbare klassen.	■ matig

Studie	Wat is onderzocht	Opzet	Uitkomst	Bewijskracht
Keuning e.a. (2017)	Waarom werkt de Focusaanpak op de ene school wel en op de andere niet?	Vergelijking achteraf van de tien scholen met het grootste en de tien met het kleinste of geen effect	Succes ging samen met beter lesgeven, een positieve houding tegenover data en een sterke datacultuur.	■ beperkt
Inspectie van het Onderwijs (2010)	Opbrengstgericht werken bij rekenen	Samenhang	Scholen die opbrengstgericht werkten, presteerden beter. Ze analyseerden hun resultaten wel, maar stelden er zelden vooraf ambitieuze doelen bij.	■ beperkt
Hofman e.a. (2009)	Zelfevaluatie, 81 basisscholen	Samenhang	Sommige vormen van zelfevaluatie gingen samen met betere leerprestaties.	■ beperkt
Moolenaar e.a. (2012)	Samenwerkingsnetwerken van leraren en gedeeld vertrouwen, 53 basisscholen	Samenhang	Hechtere netwerken gingen via gedeeld vertrouwen samen met betere leerresultaten.	■ beperkt
Schildkamp e.a. (2009)	Gebruik van het zelfevaluatie-instrument ZEBO, 79 basisscholen, vijf jaar gevolgd	Longitudinaal, zonder loting	Het gebruik van de uitkomsten hing (nog) niet samen met betere leerprestaties, wel met de professionele ontwikkeling van leraren.	■ beperkt
Blok e.a. (2015)	Wat weten we over opbrengstgericht werken?	Literatuurverkenning	De effectiviteit voor betere leerresultaten is 'aan twijfel onderhevig'. De auteurs raden realistische verwachtingen aan.	■ beperkt
VOORTGEZET ONDERWIJS (VO)				
Poortman & Schildkamp (2016)	Negen datateams van leraren en schoolleiders	Meervoudige casestudie, zonder controlegroep	Vier van de negen teams losten hun zelfgekozen probleem op en een vijfde deels. Het ging bijvoorbeeld om lage examencijfers of veel zittenblijvers.	■ beperkt
Helms-Lorenz e.a. (2016)	Begeleiding van startende leraren met lesobservaties en coaching, 338 leraren op 71 scholen	Gerandomiseerd experiment	Begeleide leraren groeiden sterker in lesvaardigheid. Leerlingresultaten zijn niet gemeten.	■ sterk
Bijlsma e.a. (2019)	Leerlingfeedback na de les via de app Impact!, 60 leraren	Gerandomiseerd experiment	Volgens leerlingen werden de lessen eerst iets beter, maar dat hield geen stand.	■ sterk

Studie	Wat is onderzocht	Opzet	Uitkomst	Bewijskracht
NRO-Kennisrotonde (2018)	Draagt een kwaliteitszorgsysteem bij aan onderwijskwaliteit?	Literatuurverkenning	Geen effect gevonden.	■ beperkt
(V)SO EN PASSEND ONDERWIJS				
KBA Nijmegen & ResearchNed (2024)	Evaluatie van de Wet kwaliteit (V)SO, met een ontwikkelingsperspectief (OPP) per leerling	Beleidsevaluatie op basis van ervaringen	Professionals vinden dat het OPP helpt bij planmatig werken, maar ervaren het ook als administratieve last. Leerlingresultaten zijn niet gemeten.	■ beperkt
Ledoux e.a. (2020)	Evaluatie van passend onderwijs, 2014-2020	Monitor zonder controlegroep	Of leerlingen erop vooruitgingen, kon niet worden vastgesteld.	■ beperkt
UITVAL EN VEILIGHEID				
Van der Steeg e.a. (2008), CPB	Afspraken over voortijdig schoolverlaten met 14 van de 39 regio's, met een beloning per uitvaller minder	Quasi-experimenteel	Geen meetbaar effect in het eerste jaar. In regio's zonder afspraken daalde de uitval ongeveer even sterk.	■ matig
MOOZ & Ecorys (2021)	Evaluatie van de Wet veiligheid op school, met de jaarlijkse monitor sociale veiligheid	Wetsevaluatie op basis van ervaringen	Veiligheid staat op de agenda van alle scholen. De wet is volgens betrokkenen niet doorslaggevend.	■ beperkt

■ **sterk** gerandomiseerd experiment of meta-analyse van studies met een causale opzet ■ **matig** quasi-experiment (vergelijkingsgroep zonder loting) ■ **beperkt** onderzoek naar samenhang, casusonderzoek, evaluatie of literatuuronderzoek

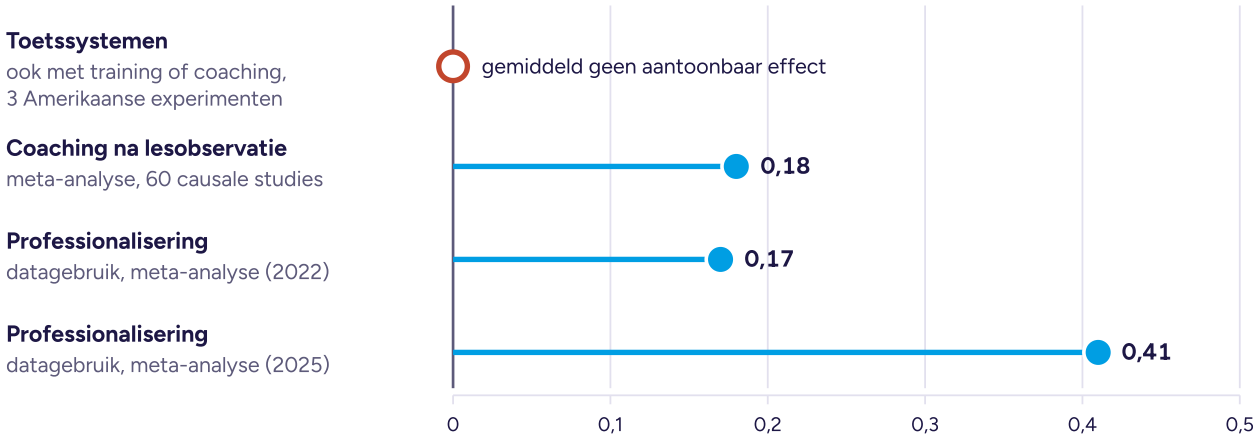
Wat buitenlands onderzoek laat zien

Buitenlands onderzoek, vooral uit de Verenigde Staten, telt meer experimenten met loting. De uitkomsten gelden niet zonder meer voor een Nederlandse school, want het Amerikaanse stelsel verschilt van het onze. Ze helpen wel waar we hier geen experiment vonden, zoals bij de vraag wat een toetssysteem op zich oplevert.

Drie Amerikaanse gerandomiseerde experimenten testten toetssystemen waarmee leraren tussentijds de voortgang volgen, soms met training of coaching erbij. Gemiddeld vonden ze geen aantoonbaar effect op leerprestaties (Konstantopoulos e.a., 2016; Cordray e.a., 2013; West e.a., 2016). Ook voor dashboards met leergegevens, vooral onderzocht in het hoger onderwijs, vond een internationale review geen bewijs dat ze leerprestaties verbeteren (Kaliisa e.a., 2024).

Twee Amerikaanse experimenten vonden wel een effect, en daar kregen scholen naast de gegevens ook begeleiding of een vaste werkwijze. In 59 districten hadden kwartaaltoetsen met training in het gebruik van de uitkomsten na een jaar een klein positief effect op rekenen (Carlson e.a., 2011). Een vroegsignaleringssysteem met een vaste teamroutine verlaagde het chronisch verzuim en de vakuitval (Faria e.a., 2017).

Volgens twee internationale meta-analyses gaan programma's waarin leraren leren werken met data gemiddeld samen met betere resultaten van leerlingen: 0,17 SD in die uit 2022 en 0,41 SD in die uit 2025 (Ansyari e.a., 2022; Spiele e.a., 2025). Ze telden ook minder strenge studies mee, dus de gemiddelden vallen mogelijk te hoog uit. Verwacht op je eigen school niet zonder meer een winst van die omvang.



Effectgrootte op leerprestaties in internationaal onderzoek, in SD. Effectgroottes hangen af van keuzes in het onderzoek, zoals vergelijkingsgroep, leerlingen en toets (Simpson, 2017; Cheung & Slavin, 2016). Lees de stippen niet als ranglijst. De open stip bij nul staat voor de toetssystemen.

Studie	Wat is onderzocht	Opzet	Uitkomst	Bewijskracht
Konstantopoulos e.a. (2016); Cordray e.a. (2013); West e.a. (2016), VS	Toetssystemen voor het tussentijds volgen van de voortgang, deels met training of coaching	Drie gerandomiseerde experimenten	Gemiddeld geen aantoonbaar effect op leerprestaties. In een van de experimenten gingen leraren ook niet vaker gedifferentieerd lesgeven.	■ sterk
Carlson e.a. (2011), VS	Kwartaaltoetsen met training in het gebruik van de uitkomsten, 59 districten	Gerandomiseerd experiment	Na een jaar een klein positief effect op rekenen.	■ sterk
Faria e.a. (2017), VS	Vroegsignalering met een vaste teamroutine, 73 middelbare scholen	Gerandomiseerd experiment	Na één jaar was het chronisch verzuim 10%, tegen 14% in de controlegroep. Een of meer vakken niet gehaald: 21%, tegen 26%.	■ sterk
Kraft e.a. (2018), internationaal	Coaching van leraren op basis van lesobservaties	Meta-analyse van 60 studies met een causale opzet	Effect op de kwaliteit van de instructie 0,49 SD, op leerprestaties 0,18 SD. Bij grote programma's ongeveer half zo groot.	■ sterk

Studie	Wat is onderzocht	Opzet	Uitkomst	Bewijskracht
Kaliisa e.a. (2024), internationaal	Dashboards met leergegevens, vooral voor studenten in het hoger onderwijs	Systematische review van 38 studies	Geen bewijs dat dashboards de leerprestaties verbeteren.	■ beperkt

Wat leer je hiervan?

In het Nederlandse PO vonden twee evaluaties van het Focusproject en één van het leesproject Streef bescheiden winst, en niet op elke school (Van Geel e.a., 2016; Keuning e.a., 2019; Van Kuijk, 2014; Keuning e.a., 2017). Die studies werkten zonder loting, en de twee Focusevaluaties vergeleken alleen met eerdere jaargroepen op dezelfde scholen. Je kunt de winst dus niet met zekerheid aan de aanpak toeschrijven. De Focusevaluatie met vergelijkbare scholen zonder de aanpak vond gemiddeld geen effect (Staman e.a., 2017), net als het experiment met loting, behalve voor leerlingen die verlengde instructie kregen (Van der Scheer & Visscher, 2018). Een eerdere verkenning van Nederlands en buitenlands onderzoek noemde de effectiviteit voor betere leerresultaten al 'aan twijfel onderhevig' (Blok e.a., 2015).

Waar scholen vooral een nieuw instrument of een verplichting kregen, is weinig effect aangetoond, al verschillen de studies sterk in opzet. Het gebruik van een zelfevaluatie-instrument hing over vijf jaar niet samen met betere leerprestaties (Schildkamp e.a., 2009). Een kennisoverzicht vond geen effect van een kwaliteitszorgsysteem (NRO-Kennisrotonde, 2018). Een feedback-app had in een Nederlands experiment geen blijvend effect (Bijlsma e.a., 2019), Amerikaanse toetssystemen gemiddeld ook niet (Konstantopoulos e.a., 2016; Cordray e.a., 2013; West e.a., 2016). Misschien zit het hierin: een instrument levert wel gegevens op, maar bepaalt niet wat een team er in de klas mee doet.

Meer houvast geeft onderzoek naar aanpakken die direct ingrijpen op het lesgeven, met lesobservaties als gegevens. In een Nederlands experiment met loting groeiden beginnende leraren met zulke begeleiding en coaching sterker in lesvaardigheid (Helms-Lorenz e.a., 2016). Of hun leerlingen daardoor beter presteerden, is niet gemeten. Internationaal verbeterde zulke coaching wel de prestaties van leerlingen (Kraft e.a., 2018).

Verwacht van werken met data dus geen snelle of zekere winst. In de projecten met winst leerden leraren en teams van gegevens naar doelen en instructie te komen (Van Geel e.a., 2016; Van Kuijk, 2014). Of juist die stap de winst verklaart, weten we niet. Hoe een aanpak op jouw school uitpakt, kun je niet voorspellen. Leg vooraf vast wat je verwacht, bijvoorbeeld hoeveel groei in begrijpend lezen je na een jaar wilt zien, en meet daarna of dat gebeurt.

Voorwaarden en valkuilen

Twee basisscholen beginnen tegelijk met dezelfde aanpak om met data te werken. Na twee jaar gaan de leerlingen op de ene school vooruit. Op de andere verandert er weinig. Als schoolleider wil je weten waar dat verschil vandaan komt, en welk deel ervan de school zelf in de hand heeft.

Keuning e.a. (2017) zagen dit patroon in het Focusproject op Nederlandse basisscholen. Dezelfde training in werken met toetsgegevens sloeg op de ene school aan en op de andere niet. Een deel van zo'n verschil kan toeval zijn. Schoolgemiddelden schommelen van jaar tot jaar, en volgens Amerikaans onderzoek sterker naarmate een school minder leerlingen per leerjaar heeft (Kane & Staiger, 2002).

Welke schoolkenmerken verder meespelen, is in het Nederlandse onderzoek dat we vonden vooral bekeken met vragenlijsten, casestudies en vergelijkingen achteraf (Schildkamp e.a., 2017; Hubers e.a., 2017; Keuning e.a., 2017). Dat onderzoek laat meestal alleen samenhang zien, bijvoorbeeld tussen veel leerwinst en een sterke datacultuur: een team waarin het gewoon is om gegevens te gebruiken en te bespreken. Wat oorzaak is en wat gevolg, is dan niet getoetst. Neem je een kenmerk van succesvolle scholen over, dan is het de vraag of het op je eigen school hetzelfde oplevert.

Wat samengaat met succes

- **De stap van analyse naar doel en instructie** blijkt in Nederlands onderzoek telkens lastig. Volgens de Inspectie analyseerden basisscholen in 2010 hun rekenresultaten vaak wel. Doelen verbonden ze er vaak niet aan (Inspectie van het Onderwijs, 2010; Kamerstukken II 31 293, nr. 96). In een quasi-experiment op 42 basisscholen kregen schoolleiders en intern begeleiders twee jaar en leraren één jaar training in het gebruik van toetsgegevens, zonder algemeen effect op de rekenprestaties. Leraren vinden het moeilijk om voortgangsgegevens om te zetten in instructie op maat, concludeerden de onderzoekers (Staman e.a., 2017). Oefenen blijft wel verstandig: van cijfer naar doel, en van doel naar wat je in de klas anders doet.
- **Goed lesgeven, een positieve houding tegenover data en een sterke datacultuur** gingen samen met de meeste winst. Bij het Focusproject vergeleken de onderzoekers achteraf de tien basisscholen waar de aanpak het meest opleverde met tien scholen met het kleinste of geen effect (Keuning e.a., 2017). Die verschillen waren er volgens de onderzoekers al vóór de aanpak. De scholen met de meeste winst

Waar het misgaat

- Een app voor **feedback van leerlingen**, waarmee zij direct na afloop de les beoordeelden, had in een Nederlands experiment in het VO geen blijvend effect (Bijlsma e.a., 2019). Volgens de leerlingen werden de lessen eerst iets beter, maar dat hield geen stand, en de leraren reflecteerden niet meer dan die in de controlegroep. Als mogelijke verklaringen noemen de onderzoekers dat leraren hun aanpassingen niet op een grondige probleemanalyse baseerden, en dat complexe lesvaardigheden zonder verdere ondersteuning lastig te verbeteren zijn. Feedback van leerlingen krijgt waarschijnlijk pas waarde als je er samen over praat en er vervolgstappen aan koppelt.
- Bij sociale veiligheid zie je wat er gebeurt als je **meet omdat het moet**. Sinds 2015 moeten scholen in PO, VO en (V)SO elk jaar de sociale veiligheid van hun leerlingen monitoren (MOOZ & Ecorys, 2021). Volgens de evaluatie van die wet staat veiligheid daardoor op de agenda, maar vinden betrokkenen de wet niet doorslaggevend voor een veiliger schoolklimaat. Sommige monitors gaven een te grof beeld of werden

presteerden bij de start vaak laag en hadden dus meer ruimte om te groeien. Of lesgeven en datacultuur de winst verklaren, toont deze vergelijking niet aan.

- **Lesobservatie en coaching** hoorden bij een begeleidingsprogramma voor beginnende leraren. In een Nederlands experiment op 71 VO-scholen versterkte dat programma hun lesvaardigheid (Helms-Lorenz e.a., 2016). Het lot bepaalde welke scholen het programma kregen, dus die groei kun je aan het programma als geheel toeschrijven. De onderdelen zijn niet apart door loting toegewezen, al hingen coaching en lesobservatie in aanvullende analyses sterk samen met de lesvaardigheid in het derde jaar. De resultaten van leerlingen zijn niet gemeten. Een internationale meta-analyse van 60 studies met een causale opzet vond dat coaching na lesobservaties ook de prestaties van leerlingen verbeterde, vooral bij taal en lezen in de voorschoolse educatie en het PO (Kraft e.a., 2018).
- Lesobservaties kun je alleen gebruiken als je op de scores kunt vertrouwen. **Betrouwbaar observeren** vraagt volgens Dobbelaer en Visscher (2023) een goed scoringsinstrument, getrainde observatoren met een handleiding en afspraken over welke en hoeveel lessen je per leraar bekijkt. Dat weegt extra zwaar als de scores meetellen in gesprekken met leraren of in beleid.
- Wil je je team leren werken met data, dan kun je steunen op onderzoek naar **professionalisering** van leraren in het algemeen. Volgens een Nederlandse reviewstudie levert professionalisering het meest op als ze over het vak en het leren van leerlingen gaat, langer duurt, samen met collega's gebeurt en aansluit bij het beleid van de school (Van Veen e.a., 2010). Losse workshops laten weinig effect zien. De review vat vooral buitenlands onderzoek samen dat vaak alleen samenhang laat zien. Op 53 Nederlandse basisscholen hing hechtere samenwerking tussen leraren samen met betere leerresultaten, via een sterker gedeeld vertrouwen in wat het team kan bereiken (Moolenaar e.a., 2012).
- **Een schoolleider met tijd voor het onderwijs** heb je niet vanzelf. Veel schoolleiders komen door dagelijkse drukte en regeldruk te weinig toe aan onderwijskundig leiderschap, stelt de Inspectie (Inspectie van het Onderwijs, 2026a). Onderzoekers bedoelen daarmee een schoolleider die doelen stelt en het onderwijs plant, volgt en evalueert. In een internationale meta-analyse, die grotendeels steunt op onderzoek naar samenhang, hing die vorm van leiderschap sterker

maar bij een deel van de leerlingen afgenomen. Zo'n monitor voldoet aan de plicht, maar een school kan er weinig gericht mee verbeteren.

- **Registreer je meer dan je gebruikt**, dan kost dat onnodig tijd, waarschuwde een Engelse adviesgroep (Data Management Review Group, 2016). Professionals in het (V)SO ervaren het OPP, het verplichte plan per leerling, als hulp bij planmatig werken (KBA Nijmegen & ResearchNed, 2024). Tegelijk voelen ze het als administratieve last, en vaak staat er meer in dan de wet vraagt.
- **Stuur je op één cijfer**, let dan op een waarschuwing van de Onderwijsraad uit 2026. Toetsen die het leren moeten ondersteunen, worden ook gebruikt voor selectie en om scholen te beoordelen, stelt de raad vast (Onderwijsraad, 2026). Te veel nadruk op resultaten kan volgens de raad het onderwijs versmallen en de kansenongelijkheid vergroten. Houd daarom naast het cijfer waarop je school wordt beoordeeld de andere onderdelen van kwaliteit in beeld.
- Op Nederlandse basisscholen zagen onderzoekers een vorm van **strategisch omgaan met toetsen**. Ongeveer een derde van de scholen sloot een of meer leerlingen uit van de eindtoets (Ehren & Swanborn, 2012). Een deel daarvan viel binnen de regels. Het onderzoek is van vóór de doorstroomtoets, maar laat zien dat een gemiddelde ook verandert door wie er meedoet. Een stijging op een stuurcijfer bewijst dus nog geen beter onderwijs.
- **Sturen op het percentage leerlingen boven een norm** kan de zwakste leerlingen uit beeld laten raken. Na invoering van zo'n maat in het Amerikaanse Chicago gingen vooral leerlingen in het midden van de verdeling vooruit. De zwakste leerlingen scoorden niet hoger (Neal & Schanzenbach, 2010). De onderzoekers verklaren dat uit extra aandacht voor leerlingen die net rond de norm zitten. In het PO werkt de Inspectie zelf met normpercentages: het percentage leerlingen op of boven 1F en het percentage op of boven 1S/2F (Staatscourant 2026, 26023). Kijk daarom ook naar de leergroei van elke leerling in het leerlingvolgsysteem, vooral bij leerlingen die ver onder die niveaus blijven.
- Jones e.a. (2017) vroegen schoolleiders uit zeven Europese inspectiestelsels, waaronder het Nederlandse, naar de **druk van buitenaf** die ze ervoeren. Wie meer druk van de inspectie ervoor, zag vaker dat curriculum en didactiek smaller werden. Een

samen met leerresultaten dan leiderschap dat vooral inspireert en motiveert (Robinson e.a., 2008). Juist dat vraagt tijd om gegevens te bekijken en te bespreken.

minderheid gaf toe de school in data en documenten gunstiger voor te stellen. Het gaat om zelfrapportage van vóór het huidige Nederlandse onderzoekskader.

- Op zes Amerikaanse scholen die met data werken, verklaarden leraren uitkomsten het vaakst vanuit hun eigen instructie. Soms wezen ze naar **kenmerken van leerlingen die ze als vast zagen**, zoals een tweede taal of een speciale onderwijsbehoefte (Bertrand & Marsh, 2015). Volgens de onderzoekers remt dat de reflectie op het eigen onderwijs en kan het lage verwachtingen voor die groepen versterken. Het gaat om een kleine beschrijvende studie. Toch helpt het om bij elke uitkomst ook te vragen wat het team zelf anders kan doen.

Staan de voorwaarden eenmaal, dan moet je ze nog vasthouden. Hubers e.a. (2017) volgden datateams in het VO tijdens een project en ongeveer een jaar daarna. Op dat tweede moment was het werken met data nog niet blijvend verankerd. De teams hadden moeite om wat ze geleerd hadden over te brengen op collega's, waardoor het datagebruik vaak beperkt bleef tot het datateam zelf. Wil je dat het werken met data een project overleeft, maak het dan deel van schoolbrede routines, met vaste momenten en rollen in de jaarplanning en een plek in de visie en de gesprekscyclus. Zo'n routine vraagt om gegevens die voor elk overleg op tijd en compleet klaarliggen, zonder dat iemand ze eerst bij elkaar moet zoeken.

Of een team data gebruikt, hangt ook samen met de kwaliteit en toegankelijkheid van de gegevens (Schildkamp e.a., 2014b; Schildkamp e.a., 2017). Daarom brengt TRIPS gegevens uit bronsystemen als ParnasSys, Somtoday en Magister samen met eigen peilingen en lesobservaties in één overzicht. Onderzoek naar de tijd die dat bespaart, vonden we niet.

5 · GEGEVENS COMBINEREN

Eén pijler zegt te weinig

Wil je als schoolleider of bestuurder weten hoe een school of groep ervoor staat, dan haal je gegevens uit verschillende systemen. Toetsresultaten staan in het leerlingvolgsysteem, de tevredenheid in een losse peiling en de verslagen van lesbezoeken in een map. Verzuim en bezetting staan weer ergens anders. Naast elkaar zie je ze zelden.

Kijk je per bron, dan zie je wel *dat* er iets speelt, bijvoorbeeld een dalende tevredenheid of een groep die achterblijft. *Waarom* het speelt, ontgaat je dan makkelijk, want de verklaring kan in een andere bron staan. En juist de verklaring bepaalt welke maatregel je kiest.

De Amerikaanse onderwijsonderzoeker Victoria Bernhardt waarschuwde eind jaren negentig al dat alleen metingen van leerresultaten te weinig zeggen om een school als geheel te verbeteren, en je zelfs op het verkeerde been kunnen zetten (Bernhardt, 1998). Zij onderscheidt vier soorten gegevens, die we in deze whitepaper pijlers noemen. De bruikbare vragen ontstaan volgens haar waar die pijlers elkaar kruisen. Haar indeling is een denkkader uit de praktijk van schoolverbetering en berust niet op onderzoek naar wat zo werken oplevert. We volgen haar, in een vrije vertaling, omdat zij het duidelijkst uitlegt waarom één soort gegevens niet genoeg is.

LEERLINGEN EN CONTEXT

Wie je leerlingen en medewerkers zijn, en wie er instroomt, aanwezig is en uitstroomt.

leerlingkenmerken, instroom, verzuim, OPP

ERVARINGEN EN BELEVING

Hoe leerlingen, ouders en medewerkers de school en het onderwijs ervaren.

tevredenheid, veiligheidsbeleving, samengevoegde uitkomsten van peilingen

LEREN EN RESULTATEN

Wat leerlingen leren, gemeten met toetsen, cijfers en andere beoordelingen.

toetsen, cijfers, voortgang op OPP-doelen

ONDERWIJS EN ORGANISATIE

Wat de school doet aan lessen, programma's, rooster en inzet van personeel. Dit is het deel dat de school zelf in de hand heeft.

lesobservaties, rooster, bezetting, voortgang van de gesprekscyclus

Hoe zo'n kruising werkt, laten we zien met een verzonnen voorbeeld: begrijpend lezen in groep 6 van een basisschool binnen een PO-bestuur met zes scholen.

EEN VERZONNEN VOORBEELD · GEEN ECHTE SCHOOL

ERVARINGEN EN BELEVING

De leerlingen van groep 6 worden dit schooljaar minder tevreden over de lessen.

Los bekeken lijkt het aan de leraar te liggen.

LEREN EN RESULTATEN

De groei in begrijpend lezen blijft in groep 6 achter. De andere groepen 6 van het bestuur laten een stabiel beeld zien.

Los bekeken lijkt groep 6 een lastige groep.

LEERLINGEN EN CONTEXT

De samenstelling van groep 6 is niet veranderd.

Los bekeken ligt het niet aan de leerlingen.

ONDERWIJS EN ORGANISATIE

Van november tot maart had groep 6 geen vaste leraar en stonden er drie verschillende invallers voor de klas. Sinds maart is er weer een vaste leraar.

Los bekeken gaat het om een bezettingsprobleem, en dat is inmiddels opgelost.

Samen bekeken wijzen de gegevens vooral naar de maanden zonder vaste leraar. Zijn die maanden de verklaring, dan kijk je eerst hoe vervanging en overdracht naar invallers geregeld zijn. Extra ondersteuning kan helpen om de achterstand in begrijpend lezen in te halen. Een verbetertraject voor de huidige leraar ligt dan minder voor de hand. Het blijft wel een hypothese die je nog moet toetsen met gegevens van meer groepen.

Bij één groep kom je niet verder. Je kunt de invallers niet losmaken van wat er verder in die maanden speelde, en het gemiddelde van zo'n kleine groep schommelt sterk door toeval, zoals Amerikaanse onderzoekers lieten zien (Kane & Staiger, 2002).

Wat Nederlands en Vlaams onderzoek zegt

Schildkamp (2019) stelt vast dat datagebruik in het onderwijs zich vaak richt op toetsresultaten, en pleit voor meer soorten gegevens en bewijs. Welke pijlers je naast elkaar legt, volgt dan uit het doel dat een team vooraf stelt. Zie je een opvallende uitslag, vraag dan meteen welke andere gegevens iets over de oorzaak kunnen zeggen.

Ander Nederlands onderzoek vond verbanden tussen de pijlers zelf. Hoe leerlingen in derdejaarsklassen in het VO het gedrag van hun leraar ervoeren, hing samen met hun toetsprestaties en hun houding tegenover het vak (Den Brok e.a., 2004). Klassen die hun leraar als sturend en betrokken zagen, scoorden gemiddeld beter. In een studie onder ruim 40.000 VO-leerlingen in zes landen, waaronder Nederland, noemden leerlingen zich betrokkener bij het leren naarmate ze positiever over het lesgeven oordeelden. In Nederland was dat verband het sterkst (Maulana e.a., 2023). Beide studies laten samenhang zien. Toch maken ze aannemelijk dat het oordeel van leerlingen over de lessen iets zegt over het leren in de klas, en dus naast de toetsuitslag hoort.

Elke pijler heeft ook eigen zwakke plekken. Zelfs verzuim is lastig eenduidig te meten. In Vlaanderen hing bij ruim 4.000 leerlingen van 15 en 16 jaar het geregistreerde ongeoorloofde verzuim maar zwak samen met wat de leerlingen zelf meldden (Keppens e.a., 2019). Neem een signaal uit één bron serieuzer als een andere bron dezelfde kant op wijst.

Wat Amerikaans onderzoek toevoegt

Het meeste onderzoek naar het combineren van bronnen komt uit de Verenigde Staten en gaat over het beoordelen van individuele leraren. In het MET-project, een groot onderzoek naar het meten van goed lesgeven, gaven lesobservaties, leerlingenquêtes en de toetsgroei van leerlingen samen een betrouwbaarder beeld van een leraar dan observaties alleen (Kane & Staiger, 2012).

Hoe je zulke verschillende bronnen weegt, is nog een open vraag. Onafhankelijke reviewers van het MET-project vinden het niet te onderbouwen om ze zomaar even zwaar mee te tellen in één score (Rothstein & Mathis, 2013). Bespreek de pijlers daarom naast elkaar in plaats van ze op te tellen tot één cijfer. Dan blijft zichtbaar welke bron wat zegt en waar bronnen elkaar tegenspreken.

Geen van de studies in dit hoofdstuk, en ook geen ander onderzoek dat we vonden, toont aan dat een school beter gaat presteren doordat ze bronnen combineert.

Begin bij een concrete vraag, dan kies je makkelijker welke gegevens nodig zijn. Het bestuur uit het voorbeeld wil bijvoorbeeld weten of groepen met veel invallers op zijn zes scholen minder groeien in begrijpend lezen. Dan legt het de gegevens van groep 5 tot en met 8 op die scholen naast elkaar.

VERDER MET JE EIGEN GEGEVENS

1. **Lees deel 2, Van verband naar bewijs.** Daarin toetst het bestuur uit het voorbeeld stap voor stap of groepen met veel invallers minder groeien in begrijpend lezen.
2. **Vraag de zelfcheck op.** Met uitspraken die we uit het onderzoek in deze whitepaper hebben afgeleid, bespreek je met je team hoe jullie met data werken. De uitspraken zijn bedoeld om dat gesprek te openen en zijn niet gevalideerd als meetinstrument.
3. **Plan een demo.** Neem een vraag van je school of bestuur mee. Met voorbeeldgegevens laten we zien wat je daarvoor in TRIPS naast elkaar zet, en we bekijken welke van jouw systemen je kunt koppelen.

Deel 2, de zelfcheck en een demo vind je op trips-edu.nl/whitepaper.

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 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 conceptteksten 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. Ansyari, M.F., Groot, W. & De Witte, K. (2022). A systematic review and meta-analysis of data use professional development interventions. Journal of Professional Capital and Community, 7(3), 256-289. emerald.com
2. Bernhardt, V.L. (1998). Multiple Measures. California ASCD, Invited Monograph No. 4. www.oregon.gov
3. Bertrand, M. & Marsh, J.A. (2015). Teachers' sensemaking of data and implications for equity. American Educational Research Journal, 52(5). doi.org
4. Biesta, G. (2009). Good education in an age of measurement. Educational Assessment, Evaluation and Accountability, 21(1), 33-46. doi.org
5. 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
6. Blok, H., Ledoux, G. & Roeleveld, J. (2015). Opbrengstgericht werken in het primair onderwijs: een effectieve weg naar onderwijsverbetering? Pedagogische Studiën, 92(3), 167-178. pedagogischestudien.nl
7. Brown, C., Schildkamp, K. & Hubers, M.D. (2017). Combining the best of two worlds: a conceptual proposal for evidence-informed school improvement. Educational Research, 59(2), 154-172. doi.org
8. Carlson, D., Borman, G.D. & Robinson, M. (2011). A multistate district-level cluster randomized trial of the impact of data-driven reform. Educational Evaluation and Policy Analysis, 33(3), 378-398. doi.org
9. CBS. Schoolweging po (dataset). data.overheid.nl
10. CBS. Woo-verzoek inzake NRO-rapportages (vertrouwelijkheid NCO-rapportages). www.cbs.nl
11. Cheung, A.C.K. & Slavin, R.E. (2016). How methodological features affect effect sizes in education. Educational Researcher, 45(5), 283-292. doi.org
12. Cordray, D. e.a. (2013). Impact of the Measures of Academic Progress (MAP) program on student reading achievement. NCEE 2013-4000. ies.ed.gov
13. Data Management Review Group (2016). Eliminating unnecessary workload associated with data management. Department for Education. www.gov.uk
14. Den Brok, P., Brekelmans, M. & Wubbels, T. (2004). Interpersonal teacher behaviour and student outcomes. School Effectiveness and School Improvement, 15(3-4), 407-442.
15. Dobbelaer, M.J. & Visscher, A.J. (2023). Hoofdstuk 20. In: Onderwijs aan het werk 2023. Den Haag: CAOP. www.caop.nl
16. Ehren, M.C.M. & Swanborn, M.S.L. (2012). Strategic data use of schools in accountability systems. School Effectiveness and School Improvement, 23(2). discovery-pp.ucl.ac.uk
17. Faria, A.-M. e.a. (2017). Getting students on track for graduation: impacts of the Early Warning Intervention and Monitoring System after one year. REL 2017-272. ies.ed.gov
18. Helms-Lorenz, M., Van de Grift, W. & Maulana, R. (2016). Longitudinal effects of induction on teaching skills and attrition rates of beginning teachers. School Effectiveness and School Improvement, 27(2), 178-204. doi.org
19. Hofman, R.H., Dijkstra, N.J. & Hofman, W.H.A. (2009). School self-evaluation and student achievement. School Effectiveness and School Improvement, 20(1). www.tandfonline.com
20. Hoogland, I., Schildkamp, K., Van der Kleij, F., Heitink, M., Kippers, W., Veldkamp, B. & Dijkstra, A.M. (2016). Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations. Teaching and Teacher Education, 60, 377-386. doi.org
21. Hubers, M.D. e.a. (2017). Capacity building by data team members to sustain schools' data use. research.utwente.nl
22. Inspectie van het Onderwijs (2010). Opbrengsten: maak er werk van! www.onderwijsinspectie.nl
23. Inspectie van het Onderwijs (2025). Technische toelichting onderwijsresultaten voortgezet onderwijs 2025. www.onderwijsinspectie.nl
24. Inspectie van het Onderwijs (2026). De Staat van het Onderwijs 2026. www.lowan.nl
25. Inspectie van het Onderwijs (2026). Ontwikkelingen in het toezicht. www.onderwijsinspectie.nl
26. Inspectie van het Onderwijs. Onderwijsresultaten (v)so. www.onderwijsinspectie.nl
27. Inspectie van het Onderwijs. Onderzoekskader 2027 (themapagina, geraadpleegd oktober 2026). www.onderwijsinspectie.nl
28. Interdepartementaal beleidsonderzoek (2022). Sturing op kwaliteit van onderwijs. open.overheid.nl
29. Jones, K.L. e.a. (2017). Unintended consequences of school inspection. Oxford Review of Education. doi.org
30. Kaliisa, R. e.a. (2024). Have learning analytics dashboards lived up to the hype? LAK '24. arxiv.org
31. Kamerstukken II, 31 293. Kamerbrief over het toezicht, december 2025. www.tweedekamer.nl
32. Kamerstukken II, 31 293, nr. 96. Kamerbrief over opbrengstgericht werken, met bevindingen van de Inspectie. zoek.officielebekendmakingen.nl
33. 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
34. Kane, T.J. & Staiger, D.O. (2012). Gathering feedback for teaching. MET Project. cepr.harvard.edu
35. KBA Nijmegen & ResearchNed (2024). Evaluatie Wet kwaliteit (v)so. In opdracht van OCW. zoek.officielebekendmakingen.nl
36. 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
37. 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
38. 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
39. Konstantopoulos, S. e.a. (2016). Effects of interim assessments on student achievement. Journal of Research on Educational Effectiveness, 9. ies.ed.gov
40. Kraft, M.A. (2020). Interpreting effect sizes of education interventions. Educational Researcher, 49(4), 241-253. doi.org
41. Kraft, M.A., Blazar, D. & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement. Review of Educational Research. doi.org

42. Kyriakides, L. & Creemers, B. (2011). Can schools achieve both quality and equity? Journal of Education for Students Placed at Risk, 16(4), 237-254. eric.ed.gov
43. Ledoux, G. & Waslander, S., m.m.v. Eimers, T. (2020). Evaluatie passend onderwijs. Eindrapport. zoek.officielebekendmakingen.nl
44. Marks, G.N. (2015). The size, stability, and consistency of school effects. School Effectiveness and School Improvement, 26(3). acuresearchbank.acu.edu.au
45. Maulana, R., Van der Lans, R., Helms-Lorenz, M. e.a. (2023). The mediated relationship between secondary school student perceptions of teaching behaviour and self-reported academic engagement across six countries. In: Effective Teaching Around the World, 439-471. Springer. doi.org
46. Melhuish, E. e.a. (2006). Effectiveness of primary schools in England in Key Stage 2 for 2002, 2003 and 2004 (EPPE 3-11). dera.ioe.ac.uk
47. 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
48. MOOZ & Ecorys (2021). Evaluatie Wet veiligheid op school. Aangehaald in de toelichting bij de concept-amvb monitor sociale veiligheid. internetconsultatie.nl
49. Neal, D. & Schanzenbach, D.W. (2010). Left behind by design. Review of Economics and Statistics. www.nber.org
50. NRO-Kennisrotonde (2018). Draagt het hanteren van een kwaliteitszorgsysteem bij aan onderwijskwaliteit, en welke sturing is het meest effectief? KR.430. www.caop.nl
51. OECD (2017). PISA 2015 Results, Volume III: Students' well-being. oecd-ilibrary.org
52. Onderwijsraad (2016). De volle breedte van onderwijskwaliteit. Van smal beoordelen naar breed verantwoorden. Den Haag: Onderwijsraad. www.onderwijsraad.nl
53. Onderwijsraad (2026). Kijk anders naar toetsing. Samenvatting via de VO-raad. vo-raad.nl
54. Poortman, C.L. & Schildkamp, K. (2016). Solving student achievement problems with a data use intervention for teachers. Teaching and Teacher Education, 60, 425-433. www.sciencedirect.com
55. Raudenbush, S.W. & Willms, J.D. (1995). The estimation of school effects. Journal of Educational and Behavioral Statistics, 20(4), 307-335. doi.org
56. Robinson, V., Lloyd, C. & Rowe, K. (2008). The impact of leadership on student outcomes. Educational Administration Quarterly. journals.sagepub.com
57. Rothstein, J. & Mathis, W.J. (2013). Review of two culminating reports from the MET Project. NEPC. nepc.colorado.edu
58. Ryan, R.M. & Deci, E.L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. Contemporary Educational Psychology, 61, 101860. doi.org
59. Scheerens, J. (2000). Improving school effectiveness. Fundamentals of Educational Planning, 68. Parijs: UNESCO International Institute for Educational Planning.
60. 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
61. Schildkamp, K. (2019). Data-based decision-making for school improvement: research insights and gaps. Educational Research, 61(3), 257-273. www.tandfonline.com
62. Schildkamp, K., Karbautzki, L. & Vanhoof, J. (2014). Exploring data use practices around Europe. Studies in Educational Evaluation, 42, 15-24. doi.org
63. 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
64. Schildkamp, K., Visscher, A. & Luyten, H. (2009). The effects of the use of a school self-evaluation instrument. eric.ed.gov
65. Simpson, A. (2017). The misdirection of public policy: comparing and combining standardised effect sizes. Journal of Education Policy, 32(4), 450-466. doi.org
66. Spiele, E., Janssen, E., Schildkamp, K. & Poortman, C. (2025). The effect of teacher data use professional development interventions on student achievement. Studies in Educational Evaluation, 87. research.utwente.nl
67. Staatsblad 2021, 320. Wet verduidelijking burgerschapsopdracht. zoek.officielebekendmakingen.nl
68. Staatscourant 2026, 24891. Onderzoekskader 2021, versie per 1 augustus 2026. zoek.officielebekendmakingen.nl
69. Staatscourant 2026, 26023. Signaleringswaarden leerresultaten primair onderwijs. zoek.officielebekendmakingen.nl
70. Staman, L., Timmermans, A.C. & Visscher, A.J. (2017). Effects of a data-based decision making intervention on student achievement. Studies in Educational Evaluation, 55. research.utwente.nl
71. Stroet, K., Opdenakker, M.-C. & Minnaert, A. (2013). Effects of need supportive teaching on early adolescents' motivation and engagement: A review of the literature. Educational Research Review, 9, 65-87. doi.org
72. 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
73. Van der Scheer, E.A. & Visscher, A.J. (2018). Effects of a data-based decision-making intervention for teachers on students' mathematical achievement. Journal of Teacher Education, 69(3), 307-320. doi.org
74. Van der Steeg, M., Van Elk, R. & Webbink, D. (2008). Did the 2006 covenants reduce school dropout in the Netherlands? CPB Document 177. www.cpb.nl
75. Van Geel, M., Keuning, T., Visscher, A. & Fox, J.-P. (2016). Assessing the effects of a school-wide data-based decision-making intervention on student achievement growth in primary schools. American Educational Research Journal, 53(2), 360-394. doi.org
76. 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
77. Van Veen, K., Zwart, R., Meirink, J. & Verloop, N. (2010). Professionele ontwikkeling van leraren: een reviewstudie naar effectieve kenmerken van professionaliseringsinterventies van leraren. Leiden: ICLON. www.poraad.nl
78. VO-raad. Wat betekent de uitbreiding monitoring sociale veiligheid voor uw school? vo-raad.nl
79. West, M.R., Morton, B.A. & Herlihy, C.M. (2016). Achievement Network's Investing in Innovation expansion: impacts on educator practice and student achievement. CEPR, Harvard University. cepr.harvard.edu