



Het ministerie van Onderwijs, Cultuur en Wetenschap wil scholen in het voortgezet onderwijs beter in staat stellen om de ontwikkeling van de basisvaardigheden te volgen. Daarom is het voornemen uitgesproken om het gebruik van een leerlingvolgsysteem (lvs) landelijk verplicht te stellen in de onderbouw. Dit voornemen roept veel vragen op: is het wenselijk, haalbaar en bruikbaar in de praktijk? Deze ontwerpstudie biedt inzicht in kansen en knelpunten, en geeft richting aan scholen, ontwikkelaars en beleidsmakers voor een succesvolle invoering en gebruik van een lvs.

Rapport

Ontwerpstudie leerlingvolg- systemen in het voortgezet onderwijs

Bernard Veldkamp
Monique Dijks
Suzanne Gerritsen

UNIVERSITY
OF TWENTE.

C&TO Lab

Oberon
onderzoek | advies

Rapport

Ontwerpstudie

leerlingvolgsystemen in het voortgezet onderwijs

In opdracht van het ministerie van Onderwijs, Cultuur en Wetenschap

Universiteit Twente

Bernard Veldkamp
Hans Luyten
Lyset Rekers
Maaïke Heitink

CitoLab

Monique Dijks
Feline de Wit
Remco Feskens
Lientje Maas
Senne van den Berg

Oberon Onderzoek en Advies

Suzanne Gerritsen
Eef Jacobs
Anniek van den Hurk
Sjerp van der Ploeg
Kevin Postma

Voor meer informatie kunt u contact opnemen met Bernard Veldkamp via b.p.veldkamp@utwente.nl

© Universiteit Twente

Enschede, 15 mei 2025
www.utwente.nl
info@utwente.nl

Disclaimer

U mag delen uit deze publicatie overnemen op voorwaarde van bronvermelding: Auteurs, de titel van de publicatie en het jaar van uitgave.
Afbeeldingen in deze publicatie kunnen auteursrechtelijk beschermd zijn. Neem geen afbeeldingen over zonder toestemming van de rechthebbende(n).

Managementsamenvatting

Om leerlingen optimaal voor te bereiden op de maatschappij wordt in het voortgezet onderwijs (vo) ingezet op de basisvaardigheden taal, rekenen-wiskunde, burgerschap en digitale geletterdheid. Met het Masterplan basisvaardigheden werkt het ministerie van Onderwijs, Cultuur en Wetenschap (OCW) samen met scholen aan een duurzame verbetering van het onderwijs op de gebieden taal, rekenen-wiskunde, burgerschap en digitale geletterdheid. Om zicht te krijgen op de ontwikkeling van de leerlingen is het voornemen uitgesproken tot een verplichtstelling van een leerlingvolgsysteem in de onderbouw van het vo¹. Door wettelijke kaders te formuleren zou, zo zal de gedachte zijn, de kwaliteit van de monitoring kunnen worden gegarandeerd. Met een leerlingvolgsysteem (lvs) wordt een systeem bedoeld waarin de ontwikkeling van leerlingen wordt bijgehouden door middel van gestandaardiseerde toetsen, dit in tegenstelling tot een leerlingadministratiesysteem (las), waarin (alleen) de administratieve gegevens van een leerling (zoals aanwezigheid en persoonlijke gegevens) worden bijgehouden. Rond genoemd voornemen leefden nog vragen met betrekking tot het draagvlak binnen scholen, de wenselijkheid en de (on)mogelijkheden voor de onderwijspraktijk. Om deze vragen te beantwoorden is door Universiteit Twente, Oberon en Stichting Cito een ontwerpstudie uitgevoerd. Doel van deze studie was om inzicht te krijgen in hoe lvs'en momenteel worden ingezet, wat de ervaringen zijn van verschillende betrokkenen en welke richtlijnen bijdragen aan succesvolle invoering en gebruik. De studie is gebaseerd op documentanalyse, interviews en raadpleging van aanbieders, scholen, beleidsmakers, experts en leerlingen. De studie resulteert in aanbevelingen voor ontwikkelaars/aanbieders en voor beleid.

Uit dit onderzoek blijkt dat 81% van de scholen in het vo reeds gebruik maakt van een lvs, maar dat de toepassing sterk verschilt. Scholen zetten het systeem bijvoorbeeld in op leerlingniveau voor feedback of overgangsbepalingen, of op schoolniveau ter ondersteuning van onderwijsverbetering. De vakken waarin lvs'en vooral worden gebruikt zijn Nederlands en rekenen-wiskunde. De gegevens worden voornamelijk door leraren benut. Zij gebruiken het lvs met name op leerlingniveau, bijvoorbeeld voor het geven van feedback over onderwijsresultaten aan individuele leerlingen. Op ongeveer een derde van de scholen gebruiken schoolleiders een lvs, met name op schoolniveau voor bijvoorbeeld het verbeteren van onderwijsprocessen. De meerwaarde van lvs'en ligt volgens de gebruikers vooral in de mogelijkheid om leerlingprestaties te spiegelen aan landelijke standaarden. De lvs'en sluiten goed aan bij de bestaande referentieniveaus. De aansluiting op de recent ontwikkelde conceptkerndoelen van SLO is wisselend: de dekking voor wiskunde is toegenomen, maar voor de talen is deze aanzienlijk lager. Vooral hogere-orde cognitieve doelen (zoals analyseren, evalueren en creëren) blijken lastig toetsbaar via standaard lvs-toetsen, met name bij Nederlands en burgerschap. Bij burgerschap speelt bovendien dat attitudevragen moeilijk objectief te meten zijn en risico's kennen zoals sociaal-wenselijke antwoorden of waardeoordelen.

De implementatie van een lvs op schoolniveau brengt zowel voordelen als uitdagingen met zich mee. Motivatie voor gebruik neemt toe wanneer het systeem goed aansluit bij het lesprogramma en inzetbaar is voor directe onderwijsverbetering. Tegelijkertijd ervaren scholen nadelen, zoals toets stress bij leerlingen, verhoogde werkdruk voor leraren en het gevoel van wantrouwen als het gebruik verplicht wordt gesteld. Oneigenlijk gebruik van lvs-data, zoals voor verantwoording richting externe toezichthouders of het nemen van determinatiebeslissingen, wordt door scholen en experts als risicovol gezien, omdat de toetsen daar niet voor ontwikkeld zijn. Op landelijk niveau is een aandachtspunt de mogelijke koppeling met gegevens

¹ Paul, M. L. J. (2024). Kamerbrief 'Brief van de minister voor primair en voortgezet onderwijs' (nr. 732). 30 mei 2024.

uit het primair onderwijs en doorstroomtoetsen. Hoewel deze koppeling de leerloopbaan kan ondersteunen, is de psychometrische basis hiervoor momenteel onvoldoende. Bovendien zijn er zorgen over privacy en de wenselijkheid van een continue datalijn, aangezien dit kan botsen met het idee dat leerlingen in het vo met een schone lei moeten kunnen beginnen.

Deze ontwerpstudie diende te onderzoeken in hoeverre een wettelijke verplichting voor het gebruiken van een lvs in de onderbouw van het vo wenselijk en nodig is, en zo ja; onder welke voorwaarden? Uit de resultaten van dit onderzoek blijkt dat er weinig draagkracht is bij verschillende stakeholders, waaronder beleidsmakers, leraren en leerlingen. Daarnaast gebruikt een groot deel van de scholen al een lvs. Voor hen zal de invoering weinig meerwaarde bieden. Bij de scholen waar geen lvs ingezet wordt, wordt meer weerstand verwacht dan dat het de gewenste resultaten oplevert. Het zou daarom zinvoller zijn om scholen de kans te geven om, naast de mogelijkheid om een lvs te gebruiken, ook de kans te geven op hun eigen manier de vaardigheden van leerlingen formatief en systematisch te evalueren. Mocht de verplichting toch worden ingevoerd, is het essentieel om vooraf duidelijkheid te geven over het doel (bij voorkeur formatief), het gebruik (welke vakken, frequentie) en de rapportagevorm. Aansluiting bij de nieuwe kerndoelen, eigenaarschap van het onderwijsveld, gebruiksvriendelijke systemen en heldere communicatie zijn daarbij randvoorwaarden. Inspiratie kan worden gehaald uit het primair onderwijs, maar implementatie moet zorgvuldig en contextspecifiek plaatsvinden. Met betrekking tot het verplichte karakter van de invoering wordt geconcludeerd dat de meerwaarde van een verplichting wordt betwijfeld en dat het risico's oproept met betrekking tot weerstand vanuit het veld, toegenomen toets druk en verlies van autonomie van scholen en docenten. Als alternatief zou overwogen kunnen worden om scholen weliswaar te verplichten om basisvaardigheden formatief te evalueren, maar hen vrij te laten in de manier waarop.

De studie heeft geresulteerd in een aantal aanbevelingen voor het implementeren en ontwikkelen van lvs'en en voor wanneer de landelijke verplichting voor het gebruik van een lvs in de onderwijs van het vo wel wordt ingevoerd.

Een lvs zou niet gebruikt moeten worden voor

- Verantwoording voor de onderwijskwaliteit door een toezichthouder (zowel intern als extern); Dit leidt af van het voornaamste doel van het lvs en kan zorgen voor druk om hoge prestaties te leveren. Dit kan leiden tot teaching-to-the-test.
- Summatieve oordelen (bijv. determinatie); Lvs-toetsen zijn ontwikkeld voor formatieve doeleinden en zijn dus niet geschikt voor summatief gebruik.
- Een vergelijking van resultaten tussen docenten. Ook dit is voor docenten een vorm voor externe verantwoording, waardoor ambivalentie in doelstellingen ontstaat. Dit kan ten koste gaan van een valide beeld van de ontwikkeling van de leerlingen.

Advies beleidsmakers

- Besluit in hoeverre vaardigheden (digitale geletterdheid) en/of attitudes (burgerschap) een plek moeten krijgen in het lvs en wees bewust van de verschillende voor- en nadelen en draagkracht die hiermee gepaard gaan
- Beperk het gebruik tot formatieve doeleinden op individueel of klasniveau. Zo voorkom je dat met het lvs meerdere doelen worden nagestreefd en behoud je de validiteit.
- Investeer in heldere communicatie rond een verplichte invoering. Scholen hebben behoefte aan duidelijke kaders, zodat ze weten wat wel en niet van hen verwacht wordt.

- Zie voorlopig af van het koppelen van lvs-gegevens uit po en vo. Deze koppeling kan momenteel nog niet betekenisvol en valide gelegd worden.
- Introduceer een kwaliteitsborging van lvs-toetsen. Dit garandeert de kwaliteit van de toetsen voor scholen.

Advies aanbieders

- Neem de nieuwe conceptkerndoelen als uitgangspunt voor een actuele invulling van de inhoud. Deze zijn actueler dan de referentieniveaus.
- Concretiseer de kerndoelen zodat ze geschikt zijn voor operationalisering binnen lvs-toetsen. Momenteel is het bij sommige kerndoelen nog onduidelijk hoe dit tot concreet uiting kan komen in de praktijk en hoe dit in verschillende leerjaren en niveaus naar voren kan komen. Concretisering van de kerndoelen draagt bij aan het valide ontwikkelen van de toetsen.
- Beperk de inhoud van het lvs tot objectief meetbare doelen. Dit komt ten goede van de betrouwbaarheid en validiteit van de toetsen.
- Verzorg een externe borging van de inhoudelijke en psychometrische kwaliteit van de lvs-toetsen.
- Onderzoek welke randvoorwaarden er zijn voor een rapportage die direct toepasbaar en inzichtelijk is. Scholen hebben namelijk baat bij een direct toepasbare rapportage om het zinvol in te kunnen zetten in hun onderwijspraktijk.
- Stimuleer dat het lvs dusdanig gebruiksvriendelijk vormgegeven is dat docenten gemotiveerd raken het te gebruiken.
- Neem wensen van docenten als uitgangspunt bij het ontwikkelen van de rapportages.

Advies leraren

- Verhoog motivatie van leerlingen door uit te leggen wat een lvs hen oplevert. Intrinsieke motivatie van leerlingen zal bijdragen aan het valide en betrouwbaar meten van hun ontwikkelingen.
- Integreer de afname van het lvs in het leerproces ter voorkoming van toets stress. Toets stress is namelijk al een probleem in het voortgezet onderwijs en het is onwenselijk om deze verder te vergroten.

Beperkingen

Wel zijn er beperkingen waar rekening mee gehouden moet worden:

- Een lvs is niet geschikt om de meer cognitieve vaardigheden te meten.
- Het draagvlak voor het verplicht invoeren van een lvs vo is laag bij schoolleiders, docenten en leerlingen

Tenslotte, om recht te doen aan de toegenomen aandacht voor praktijkgericht onderwijs binnen het vo:

- Verken hoe praktijkgerichte vaardigheden gemeten kunnen worden binnen een lvs.

Inhoudsopgave

1	Inleiding	9
1.1	Aanleiding.....	9
1.2	Leeswijzer	10
1.2.1	Woordenlijst	10
2	Onderzoeksopzet.....	11
3	Huidig gebruik van lvs in het vo	12
3.1	Inleiding	12
3.2	Dekkingsgraad	12
3.2.1	Redenen voor niet-gebruik	13
3.3	Inzet van een leerlingvolgsysteem	13
3.3.1	Redenen voor gebruik lvs	13
3.3.2	Survey: opbrengsten van het gebruiken van een lvs	15
3.3.3	Survey: gebruik van lvs-rapportages.....	16
3.3.4	Vakgebieden	17
3.3.5	Gebruikers	17
3.4	Aanbod lvs	19
3.4.1	Dekking kerndoelen Nederlands	20
3.4.2	Dekking kerndoelen Engels	22
3.4.3	Dekking kerndoelen rekenen-wiskunde	23
3.4.4	Motivatatie voor dekking kerndoelen door toetsaanbieders	24
3.4.5	Vergelijking referentiekaders taal en rekenen met inhoud volgroepen	25
3.4.6	Motivatatie voor dekking van referentiekader door toetsaanbieders	32
3.5	Knelpunten huidig gebruik	33
3.5.1	Survey: knelpunten bij interpretatie en implementatie van lvs-resultaten	34
4	Toekomstig gebruik van lvs in het vo	36
4.1	Inleiding	36
4.2	Wenselijkheid verplichting lvs.....	36
4.2.1	Ontwikkeling volgen	36
4.2.2	Curriculum	38
4.2.3	Toetsdruk	39
4.2.4	Werkdruk docenten	39
4.2.5	Eigenaarschap en visie.....	40
4.2.6	Marktwerking, kwaliteit en databeheer	41
4.2.7	Survey: wenselijkheid verplichting lvs'en voor kernvakken	41
4.2.8	Survey: nadelen en voordelen van gebruik lvs'en	42
4.3	Oneigenlijk gebruik	43
4.3.1	Survey: voorbeelden van oneigenlijk gebruik	44
4.4	Onderwijsinhouden toekomstig lvs	44
4.4.1	Wenselijkheid van het opnemen van specifieke vakgebieden	44
4.4.2	Meetbaarheid van de onderwijsdoelen in een lvs.....	45
4.4.3	Burgerschap en digitale geletterdheid	46
4.4.4	Samenhang tussen vakgebieden	47
4.5	Koppeling met het lvs po	51
4.5.1	Aansluiting onderwijsinhoud volgroepen po en vo	51

4.5.2 (Psychometrische) mogelijkheden voor een koppeling.....	52
4.5.3 Wenselijkheid	56
4.5.4 Aandachtspunten	56
4.6 Voorwaarden voor succesvolle implementatie	57
4.6.1. Voorwaarden in de implementatie van de verplichting	57
4.6.2 Voorwaarden in het eigenaarschap en de vaardigheden van het onderwijsveld	58
4.6.3 Voorwaarden in de rapportage van de lvs-resultaten	58
5 Vervolgonderzoek	59
5.1 Praktische vaardigheden	59
5.1.1 Conceptualisering van praktische vaardigheden	59
5.1.2 Praktische vaardigheden in beroepscontext.....	60
5.1.3 Draagkracht voor het opnemen van praktische vaardigheden in een lvs	60
5.1.4 Concluderend	61
5.2 Cognitief Diagnostische Modellen	61
5.2.1 Achtergrond en vraagstelling	61
5.2.2 Methode	63
5.2.3 Resultaten.....	64
5.2.4 Conclusie	69
5.3 Omgaan met de resultaten.....	69
5.3.1 Het terugkoppelen van resultaten aan leerkrachten en scholen	70
5.3.2 Rapportages in leerlingvolgsystemen	70
6 Conclusies en aanbevelingen	72
Bijlage 1 Onderzoeksverantwoording	76
Interviews	76
Enquête	76
Literatuurstudie.....	77
Focusgroepen	78
Survey met vignetten	78
Gesprekken en documentenanalyse toetsaanbieders.....	80
Technische achtergrond Cognitive Diagnostic Models	80
Bijlage 2 Inhoudelijke dekking referentiekader taal en rekenen met doelomschrijvingen	85
Bijlage 3 Meetbaarheid van de kerndoelen voor Nederlands, Engels, wiskunde, digitale geletterdheid en burgerschap	90
Bijlage 4 Wat zijn referentieniveaus?	96
Bijlage 5 Bronnenlijst	100

Inleiding

1.1 Aanleiding

Om leerlingen optimaal voor te bereiden op de maatschappij, vervolgonderwijs en de arbeidsmarkt wordt in het voortgezet onderwijs (vo) ingezet op de basisvaardigheden taal, rekenen-wiskunde, burgerschap en digitale geletterdheid. Met het Masterplan basisvaardigheden werkt het ministerie van Onderwijs, Cultuur en Wetenschap (OCW) samen met scholen aan een duurzame verbetering van het onderwijs op de gebieden taal, rekenen-wiskunde, burgerschap en digitale geletterdheid. Om de Tweede Kamer op de hoogte te houden van de vorderingen van het plan stuurde de demissionair minister voor primair en voortgezet onderwijs op 30 mei 2024 een brief naar de Tweede Kamer over de voortgang van het masterplan. De brief gaat in op een groot aantal initiatieven waaronder het gebruik van leerlingvolgsystemen. Met een leerlingvolgsysteem (lvs) bedoelen we voor dit onderzoek een systeem waarin de ontwikkeling van leerlingen wordt bijgehouden door middel van gestandaardiseerde toetsen. We kijken niet naar een systeem waarin de administratieve gegevens van een leerling (zoals aanwezigheid en persoonlijke gegevens) worden bijgehouden. Dit is namelijk een leerlingadministratiesysteem (las).

Scholen in het primair onderwijs (po) werken verplicht met een lvs voor taal en rekenen om de prestaties van hun leerlingen goed in het oog te houden. Hiermee kan de school de voortgang en resultaten goed bijhouden van zowel individuele leerlingen als van groepen en de school in het geheel. In het vo is het gebruik van een lvs nog niet verplicht. In 2018 is een inventarisatie gedaan naar het gebruik van een lvs in het vo en toen bleek de dekkingsgraad 92% te zijn. Echter ging het in deze meting niet alleen om lvs'en, maar ook om las'en (zoals Magister en Somtoday). We kunnen dus niet met zekerheid zeggen dat een groot deel van de scholen al gebruikmaakt van een lvs. Daarnaast kunnen er door het ontbreken van wettelijke kaders geen eisen gesteld worden aan de systemen, waardoor de kwaliteit niet gewaarborgd kan worden. Daarnaast is het belangrijk dat alle vo-scholen op deze manier goed zicht hebben op de ontwikkeling van hun leerlingen. Daarom wordt in de onderbouw van het vo ook de verplichting ingevoerd voor een lvs. Om te kunnen bepalen hoe dat eruit moet komen te zien, is een ontwerpstudie uitgevoerd naar de benodigde kwaliteitseisen en gebruikscriteria.

Deze ontwerpstudie moet leiden tot een 'ontwerpkader' of 'programma van eisen' en moet antwoord geven op vragen als:

1. Welke onderwijsinhouden in (de onderbouw van) het voortgezet onderwijs lenen zich ervoor om met een landelijk lvs systematisch gevolgd te worden?
2. Hoe verhouden zich de in leerlingvolgsystemen getoetste (basis)vaardigheden zich tot landelijke regelgeving met betrekking tot deze inhouden, en tot de actualisatie daarvan in het kader van de integrale curriculumherziening?
3. Hoe verhouden deze inhouden zich tot de inhouden die in de momenteel meest gebruikte leerlingvolgsystemen worden gevolgd.
4. Wat is het marktaandeel van deze meest gebruikte lvs'en (welk aandeel van de scholen gebruikt welke lvs'en)?
5. In hoeverre zijn de leerlingvolgsystemen die gebruikt worden in het voortgezet onderwijs te koppelen aan de in het primair onderwijs gebruikte systemen, en in hoeverre is het gegeven dat

- scholen (zowel in po als vo) verschillende leerlingvolgsystemen (kunnen) gebruiken hierbij een complicerende factor – en (hoe) is dat te ondervangen?
6. Hoe kan worden gewaarborgd dat informatie die wordt verkregen uit leerlingvolgsystemen aansluit bij de intrinsieke motivatie van leraren (en leerlingen) om zicht te krijgen op leervorderingen en dat deze informatie niet ‘oneigenlijk’ gebruikt wordt?
 7. Waar lopen scholen tegenaan bij de implementatie en het gebruik van leerlingvolgsystemen en wat kunnen scholen hierover leren van scholen die leerlingvolgsystemen succesvol hebben geïmplementeerd en succesvol gebruiken?

1.2 Leeswijzer

Hoofdstuk 2 van dit rapport geeft een kort overzicht van de onderzoeksopzet. In hoofdstuk 3 geven we een beschrijving van het huidige gebruik van een lvs in het vo. Daarbij kijken we naar onderwerpen als dekkingsgraad, inzet, gebruik en knelpunten bij gebruik van een lvs. Vervolgens zoomen we in hoofdstuk 4 in op de consequenties van het verplicht stellen van een lvs in de toekomst. Hierbij beschrijven we de wenselijkheid, de risico's van oneigenlijk gebruik, de gewenste inhoud, de koppeling met lvs po en de voorwaarden voor een succesvolle implementatie. hoofdstuk 5 beschrijft een aantal onderwerpen die in deze studie zijn verkend, maar waar een verdieping wenselijk is. In hoofdstuk 6, ten slotte, worden de conclusies weergegeven en aanbevelingen gedaan voor een succesvolle implementatie van lvs in het vo.

1.2.1 Woordenlijst

Er worden een aantal termen gebruikt in het rapport die verschillend geïnterpreteerd zouden kunnen worden. In Tabel 1.1 staat uitgewerkt wat bedoeld wordt met de termen.

Tabel 1.1 Woordenlijst

Term	Definitie
Gespecialiseerd onderwijs	Scholen in het speciaal onderwijs, speciaal basisonderwijs, voortgezet speciaal onderwijs en het praktijkonderwijs
Onderwijssoort	Het onderscheid tussen vmbo, havo en vwo in het voortgezet onderwijs
Sector	Het primair ofwel voortgezet onderwijs
Vakgebieden	Bijvoorbeeld Nederlands, Engels, rekenen-wiskunde, burgerschap en digitale geletterdheid
Leerlingadministratiesysteem (las)	Een systeem waarin de administratieve gegevens worden bijgehouden (zoals roosters of absentie leerlingen). Een las bevat geen eigen meetinstrument. In de volksmond worden deze systemen soms ook leerlingvolgsystemen genoemd, maar in dit onderzoek maken we een duidelijk onderscheid tussen een lvs en een las.

Onderzoeksopzet

Om de verschillende onderzoeksvragen te beantwoorden en te komen tot aanbevelingen voor het succesvol implementeren van een verplicht lvs in het vo, is het project opgedeeld in een aantal werkpakketten. Elk werkpakket richt zich op één van de onderzoeksvragen. Vanwege de nauwe verbinding zijn de tweede en derde onderzoeksvraag daarbij gecombineerd. Voor elk van de werkpakketten is een specifieke aanpak geformuleerd. Voor een compleet overzicht van de onderzoeksaanpak verwijzen we naar bijlage 1.

Bij het beantwoorden van onderzoeksvragen 1, 5 en 6 is gestart met een literatuurstudie en met het opvragen van informatie bij aanbieders van bestaande lvs'en. Op deze manier, aangevuld met bureauonderzoek, is informatie verzameld over de beschikbaarheid van lvs'en, over hoe ze gebruikt worden, over hoe vaak ze gebruikt worden, over hoe lvs'en uit po en vo op elkaar aansluiten en over wat al bekend is over wenselijkheid van verschillende manieren van gebruik.

Deze informatie is vervolgens gebruikt als start voor interviews en focusgroepen met aanbieders, stakeholders, scholen en leerlingen. De aanbieders zijn bevraagd voor het beantwoorden van onderzoeksvragen 2 en 3. Stakeholders zijn bevraagd voor het beantwoorden van de onderzoeksvragen 1 en 6. Scholen zijn bevraagd voor onderzoeksvragen 1, 4 en 6 en leerlingen voor onderzoeksvraag 6. In een focusgroep met toetsexperts is verdere informatie verzameld voor de extra deelvraag over praktische vaardigheden binnen onderzoeksvraag 1. Voor inhoud onderzoeksvragen zie paragraaf 1.1.

Met een simulatiestudie is onderzocht in hoeverre het mogelijk is lvs-toetsen te ontwikkelen die op het niveau van deelvaardigheden betrouwbare resultaten kunnen rapporteren, als onderdeel van de optionele studie in onderzoeksvraag 6.

Ten slotte is een vignettenstudie uitgevoerd om onderzoeksvraag 7 te beantwoorden. Bij het maken van de vignetten is ook rekening gehouden met vragen die nog open stonden na de gehouden interviews en focusgroepen. Voor de gebruikte vignetten zie bijlage 1.

Huidig gebruik van lvs in het vo

3.1 Inleiding

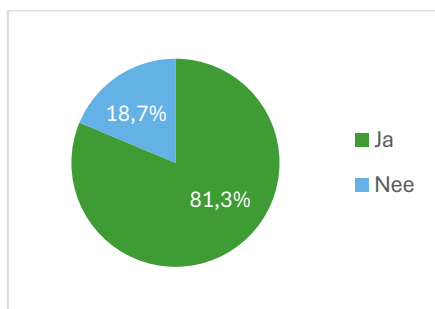
Voor het voortgezet onderwijs worden verschillende leerlingvolgsystemen aangeboden die aansluiten bij de in deze studie gebruikte definitie (zie paragraaf 1.2): Cito Leerling in beeld VO (Cito), JIJ! (ICE), Diataal (Diaatoetsen) en AMN Taal- en Rekentoetsen. Enkele jaren geleden (2018) meldde OCW aan de Tweede Kamer dat 92% van de vo-scholen een lvs gebruikte en dat 88% dat ‘voldoende tot goed’ deed.²

Voor deze studie hebben we het huidige gebruik van het lvs in het vo in kaart gebracht. In dit hoofdstuk gaan we achtereenvolgens in op het aandeel scholen dat gebruikmaakt van een lvs, voor welke vakgebieden en waarom scholen (geen) gebruik maken van een lvs. Ook gaan we in op het huidige aanbod van lvs'en en knelpunten in het huidige gebruik.

3.2 Dekkingsgraad

In een vragenlijst onder alle vo-vestigingen³ hebben we scholen bevraagd op het huidige gebruik van lvs'en. Daaruit komt naar boven dat op het merendeel van de vestigingen (81%) wordt gebruikgemaakt van een leerlingvolgsysteem. Dit is ofwel voor alle leerlingen van de school, ofwel voor een deel van de leerlingen. Op de overige 19% van de vestigingen wordt geen gebruikgemaakt van een leerlingvolgsysteem, zie Figuur 3.1. Een aantal vestigingen heeft aangegeven dat zij wel gebruikmaken van een lvs, maar zij benoemden dan een las (bijv. Magister). Deze respondenten zijn hercodeerd naar ‘geen gebruik van een lvs’ en vallen in de onderstaande figuur dus in de groep ‘Nee’. Administratiesystemen zijn in deze studie bewust niet meegenomen. We zien wel dat de meeste vestigingen die een lvs gebruiken, deze *combineren* met een las (namelijk 81%).

Figuur 3.1 Wordt er op uw vestiging gebruik gemaakt van een lvs? (n = 545)



Ongeveer een kwart van de vestigingen gebruikt een lvs voor een deel van de leerlingen op de vestiging. In de meeste gevallen gaat dit dan om leerlingen uit de onderbouw. Op enkele vestigingen wordt aangegeven dat het lvs alleen gebruikt wordt voor een specifieke groep leerlingen, bijvoorbeeld NT2-leerlingen of leerlingen die een leerachterstand hebben.

² https://www.tweedekamer.nl/kamerstukken/brieven_regering/detail?id=2018D02166&did=2018D02166

³ Respons = 37%

Over het algemeen gebruiken scholen één lvs. Van de vier aanbieders bedienen twee aanbieders een groot aandeel van de scholen (ong. 40%), één een middelgrote aanbieder (ong. 20%) is en de laatste aanbieder levert aan een klein aantal scholen het lvs (ong. 1%). Een vierde van de vestigingen heeft aangegeven meerdere lvs'en te gebruiken op de vestiging.

Vestigingen die een lvs gebruiken, waarderen het lvs gemiddeld met een 6,9 op een schaal van 1 tot 10. Dertig vestigingen (8,4%) geven aan dat ze verwachten in de komende drie jaar over te stappen op een ander lvs. De redenen hiervoor variëren. Zo willen acht vestigingen met één systeem gaan werken voor verschillende toetsen en administratieve functies in plaats van met verschillende systemen. Ook zijn zeven vestigingen nog in de oriëntatiefase en is het dus nog niet zeker of zij gaan overstappen. Ze willen eerst op zoek naar een beter alternatief dan wat ze nu hebben. Andere redenen om over te stappen zijn de gebruiksvriendelijkheid en kwaliteit van een ander systeem (genoemd door zeven vestigingen) of de aansluiting op het type onderwijs (genoemd door vijf vestigingen). Tot slot geven twee vestigingen aan dat ze in de transitiefase zitten van het ene lvs naar het andere lvs.

3.2.1 Redenen voor niet-gebruik

In de vragenlijst hebben we aan de scholen die *geen lvs* gebruiken, gevraagd wat hiervoor de reden is. Voor 102 respondenten (18,7%) geldt dat er op hun vestiging geen gebruik wordt gemaakt van een lvs. Het deel van de respondenten dat we hebben hercodeerd naar 'geen gebruik van een lvs' heeft deze vervolgvraag over de reden voor niet-gebruik niet gekregen. Ook in de verkorte vragenlijst (zie Onderzoeksverantwoording in Bijlage 1) is deze vraag niet gesteld. Uiteindelijk hebben 31 respondenten de vraag beantwoord. De meest genoemde redenen voor niet-gebruik zijn dat het werken met een lvs niet past bij het onderwijsconcept of dat het leerlingadministratiesysteem voldoende functionaliteiten biedt om leerlingen te kunnen volgen. Daarvan verwachten 22 respondenten (71%) niet dat er op de vestiging in de komende drie jaar met een lvs gewerkt gaat worden. Naast de eerder genoemde redenen wordt bij deze vraag ook nog genoemd dat lvs'en vaak te duur zijn. Redenen om wel te gaan werken met een lvs in de komende drie jaar is omdat respondenten het idee hebben dat het verplicht gaat worden of dat ze zelf vinden dat ze hun leerlingen beter kunnen volgen dan op de manier die nu op de vestiging gehanteerd wordt.

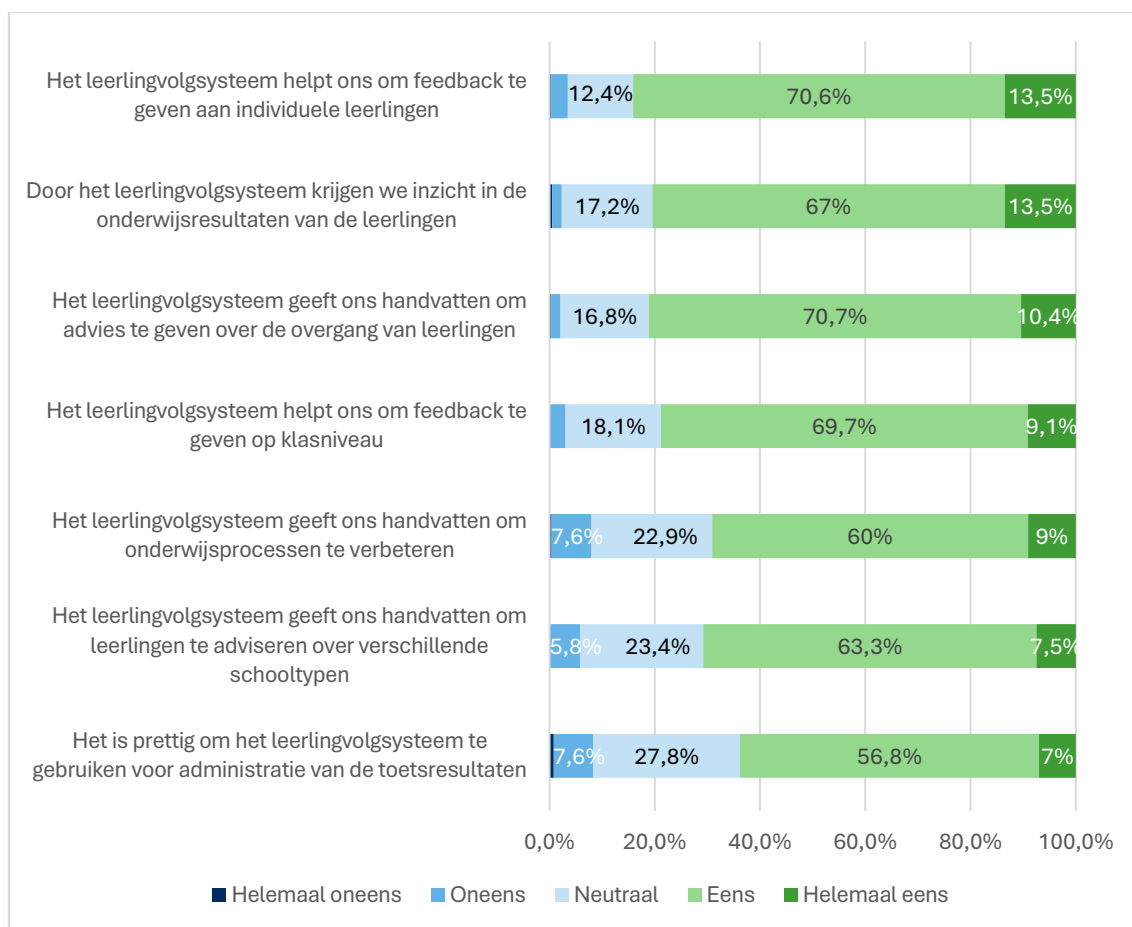
3.3 Inzet van een leerlingvolgsysteem

In de vragenlijst vroegen we de respondenten van vestigingen die *wel* met een lvs werken naar de manier waarop ze het lvs inzetten op school. De respondenten vulden hierbij de vragen in over één lvs: ofwel het enige lvs dat gebruikt wordt op de vestiging, ofwel het meest gebruikte lvs op de vestiging. Ook in de gesprekken met scholen en vakverenigingen hebben we informatie opgehaald over de inzet van het lvs in de praktijk.

3.3.1 Redenen voor gebruik lvs

Een lvs kan worden ingezet voor het geven van onderwijs, het bieden van ondersteuning aan leerlingen en het monitoren van de kwaliteit van het onderwijs. We vroegen respondenten naar de mate waarin het lvs de school hierbij helpt (zie Figuur 3.2).

Figuur 3.2 In hoeverre zijn respondenten het eens met de volgende stellingen met betrekking tot functionaliteiten van het lvs? (n = 371)



Noot. Omwille van leesbaarheid zijn percentages <5% niet in de figuur weergegeven.

Het meest positief zijn de respondenten over de functionaliteiten van het lvs op *leerlingniveau*. Zo is het merendeel van de respondenten het er (helemaal) mee eens dat het lvs hen helpt om feedback te geven aan individuele leerlingen, om inzicht te krijgen in onderwijsresultaten van leerlingen en dat het lvs hen handvatten biedt om advies te geven over de overgang van leerlingen.

Uit de gesprekken met scholen halen we een vergelijkbaar beeld op. De scholen gebruiken het lvs voornamelijk om te bepalen of leerlingen extra begeleiding nodig hebben. De uitkomsten worden dan bijvoorbeeld gebruikt om leerlingen in te delen voor steunlessen. Het op klasniveau analyseren van de lvs-data lukt niet overal even goed. De reden hiervoor is dat dit tijd kost die de docenten en andere betrokkenen in de school niet voldoende hebben.

Een school benoemt het voordeel van de 0-meting die het door hen gebruikte lvs biedt. Zij zetten de 0-meting in om zo het niveau van alle brugklasleerlingen met een gelijke toets te meten. De leerlingen komen binnen met informatie van het lvs en de doorstroomtoets uit het po, maar doordat er in het po allerlei verschillende aanbieders van lvs'en zijn en leerlingen van allerlei verschillende scholen komen, is het voor vo-scholen lastig de resultaten te interpreteren en te vergelijken.

Scholen geven ook aan dat ze de lvs-resultaten gebruiken voor determinatie: leerlingen indelen in groepen. Een voorbeeld is een school voor praktijkonderwijs, waar ze werken met verschillende niveaugroepen op basis van de lvs-resultaten. Een andere school met een brede brugklas gebruikt de resultaten om te bepalen naar welke onderwijssoort leerlingen na klas 3 gaan. Overigens merkt deze school hierbij op dat nooit alleen naar de lvs-data wordt gekeken. Ook resultaten van de schooleigen toetsen en andere input worden gebruikt voor de besluitvorming.

Zoals uit Figuur 3.2 op te maken valt, wordt het lvs ook op *schoolniveau* ingezet om onderwijsprocessen te verbeteren. Het lijkt erop dat het lvs daarvoor minder geschikt bevonden wordt, dan voor de inzet op leerlingniveau. Ook uit de gesprekken met scholen horen we wisselende signalen over het gebruik van het lvs op schoolniveau. We hebben gesproken met scholen die het lvs *alleen* op schoolniveau inzetten, om de ontwikkeling van klassen en leerjaren te volgen en de vergelijking met een landelijk beeld te kunnen maken. Ook zijn er scholen die het lvs juist niet inzetten op schoolniveau, en alleen focussen op het leerlingniveau. Enkele andere scholen geven aan wel de intentie te hebben om het lvs in te gaan zetten op schoolniveau, maar dat dat nog in ontwikkeling is of dat dit afhankelijk is van het vakgebied, bijvoorbeeld voor het opzetten van een schoolbreed, vakoverstijgend rekenprogramma of taalbeleid.

Van leerlingen horen we terug dat het voor hen over het algemeen niet duidelijk is waarom ze de lvs-toetsen moeten doen. Er wordt vaak gezegd dat het niet mee telt voor een cijfer of voor het overgaan naar het volgende jaar. Ook hoeven leerlingen zich er niet op voor te bereiden. Het is voor deze leerlingen weinig motiverend om een lvs-toets te maken. Enkele docenten beamen dit en vragen zich af of leerlingen hun best doen voor de toets, en zo niet, wat dan de waarde van het resultaat is.

3.3.2 Survey: opbrengsten van het gebruiken van een lvs

In een survey is aan schoolleiders en vakdocenten gevraagd wat de voordelen zijn voor hun school van het inzetten van lvs'en voor de kernvakken. Veel respondenten antwoorden dat ze een methode-onafhankelijk inzicht geven in het niveau van leerlingen. We hebben specifiek doorgevraagd wat de opbrengsten zijn voor hun school, docenten en leerlingen/ouders.

Op schoolniveau helpen lvs'en te beoordelen of het beeld van het prestatieniveau van leerlingen klopt, om te bepalen welke leerlingen extra ondersteuningslessen of begeleidingsuren nodig hebben en welke leerlingen extra uitgedaagd kunnen worden. Daarnaast bieden lvs'en de mogelijkheid om klas of schoolprestaties te vergelijken met benchmarks/normscores. Hierdoor wordt duidelijk waarop klassen/scholen relatief goed presteren en waar nog verbetering mogelijk is. Lvs'en kunnen ook gebruikt worden om te peilen in hoeverre interventies effect hebben.

De op docentniveau genoemde opbrengsten van lvs'en overlappen met de opbrengsten op schoolniveau. Lvs'en helpen bij het vergelijken van prestaties tussen leerlingen, klassen en scholen. Ze maken zichtbaar op welke onderdelen leerlingen uitvallen, waardoor docenten beter kunnen begrijpen waar lage cijfers vandaan komen en hierop kunnen inspelen. Ook ondersteunen lvs'en docenten bij het bepalen welke leerlingen extra hulp of juist meer uitdaging nodig hebben. De resultaten worden gedeeld met mentoren en opgeslagen in administratieve systemen zoals Magister, zodat docenten gezamenlijk een ondersteuningsplan kunnen opstellen. Vooral bij talige vakken helpen lvs'en om gerichte maatregelen te nemen, volgens een schoolleider.

De opbrengsten van lvs-gebruik voor leerlingen/ouders is dat leerlingen beter begrijpen hoe ze ervoor staan bij de kernvakken. Ze bieden een (methode-)onafhankelijke beoordeling van hun leerprestaties, wat kan bijdragen aan een beter zelfbeeld. Het helpt hen te zien waar ze harder aan moeten werken en maakt hiaten in hun kennis zichtbaar. Voor leerlingen met een duidelijke achterstand kunnen extra lessen, zoals rekenen en begrijpend lezen, gericht worden ingezet. Een respondent stelt dat ouders blij zijn met een lvs omdat ze hiermee een doorlopende leerlijn tussen po en vo realiseren: leerlingen kunnen meteen hulp krijgen in de kernvakken als dat nodig is.

3.3.3 Survey: gebruik van lvs-rapportages

We hebben in de survey gevraagd aan schoolleiders en vakdocenten waarvoor ze lvs-rapportages inzetten. Een overzicht van de antwoorden worden getoond in Tabel 3.1. Van twee respondenten ontbreken hier informatie. Het gaat om een vakdocent van een school zonder lvs, en een schoolleider van een school met lvs.

Tabel 3.1 Vormen van huidig gebruik van lvs-rapportages

Lvs-rapportages worden op onze school ingezet voor ...	Totaal (n=14)
Extra ondersteuning van laag scorende leerlingen	85%
Extra ondersteuning van hoog scorende leerlingen	7%
Aanpassingen van de vormgeving van het onderwijs	36%
Mentorgesprekken met leerlingen	79%
Rapportbesprekingen met ouders	71%
Externe verantwoording	64%

De ruime meerderheid van de respondenten (85%) gebruikt lvs-rapportages om inzicht te krijgen in de extra ondersteuningsbehoeften van laag scorende leerlingen, zodat ze extra kunnen oefenen. Een respondent voegt hier echter aan toe dat aanwijzingen hieromtrent van docenten belangrijker zijn dan de lvs-rapportage. Een ander zegt dat ze hiervoor tevens de voortgangstoetsresultaten gebruiken; die moeten ook een lage score laten zien. Het gebruik van lvs-rapportages om extra ondersteuningsbehoeften van hoog scorende leerlingen vast te stellen wordt door slechts één bevestigd. De overige respondenten stellen dat hiervoor geen geld en/of geen beleid is.

Ruim een derde van de respondenten (36%) gebruikt lvs-rapportages voor aanpassingen in de vormgeving van het onderwijs. De gegeven toelichting hierop is weinig informatief.

Een ruime meerderheid (79%) geeft aan dat lvs-rapportages worden gebruikt voor mentorgesprekken met de leerling of voor rapportbespreking met ouders (71%). De genoemde redenen overlappen sterk. Het gebeurt weinig systematisch: het hangt af van de mentor, het gebeurt alleen als er onverwacht hoge of lage resultaten uitkomen of om eventuele bijlessen te onderbouwen. Als lvs-rapportages niet worden gedeeld met leerlingen of ouders, dan geeft men als reden op dat ze niet te veel druk willen leggen op lvs'en, en dat vaktoetsen of rapportcijfers een beter inzicht geven. Volgens bijna de helft van de respondenten (43%) kan het inzien van de resultaten van lvs'en (soms) motivatieverhogend werken voor leerlingen, maar het hangt af van hoe het gebracht wordt door de mentor. De meeste leerlingen zegt het inzien van de resultaten niet zoveel, naast hun rapportcijfers die leidende zijn, voor de overgang en advies.

Bijna tweede (64%) van de respondenten denkt dat lvs-rapportages op hun school worden gebruikt voor externe verantwoording. Uit de beperkte toelichting die wordt gegeven, blijkt echter dat men hierover onzeker is. Een schoolleider noemt als reden ter verantwoording van NP Onderwijsgelden, en een andere respondent geeft aan dit gebeurt omdat het management het belangrijk vindt.

3.3.4 Vakgebieden

In de vragenlijst vroegen we de respondenten naar de vakgebieden waarvoor het lvs gebruikt wordt. Respondenten konden meerdere antwoorden kiezen. Op nagenoeg alle vestigingen wordt het lvs gebruikt voor de vakken Nederlands en wiskunde of rekenen en op bijna drie vierde van de vestigingen wordt het lvs gebruikt voor Engels (zie Tabel 3.2).

Tabel 3.2 Voor welke vakken maakt uw vestiging gebruik van dit lvs?

	Totaal (n = 404)
Nederlands	97,8%
Wiskunde of rekenen	93,3%
Engels	72,7%

Op de meeste scholen (69%) wordt het lvs alleen gebruikt voor de hierboven genoemde vakken. Op de overige vestigingen (126) wordt het lvs ook gebruikt voor andere ontwikkelingsgebieden. Dan gaat het meestal om de leer-werkhouding en in veel gevallen ook om de sociaal-emotionele ontwikkeling, zie Tabel 3.3.

Tabel 3.3 Voor welke ontwikkelingsgebieden van leerlingen gebruikt uw vestiging het lvs?

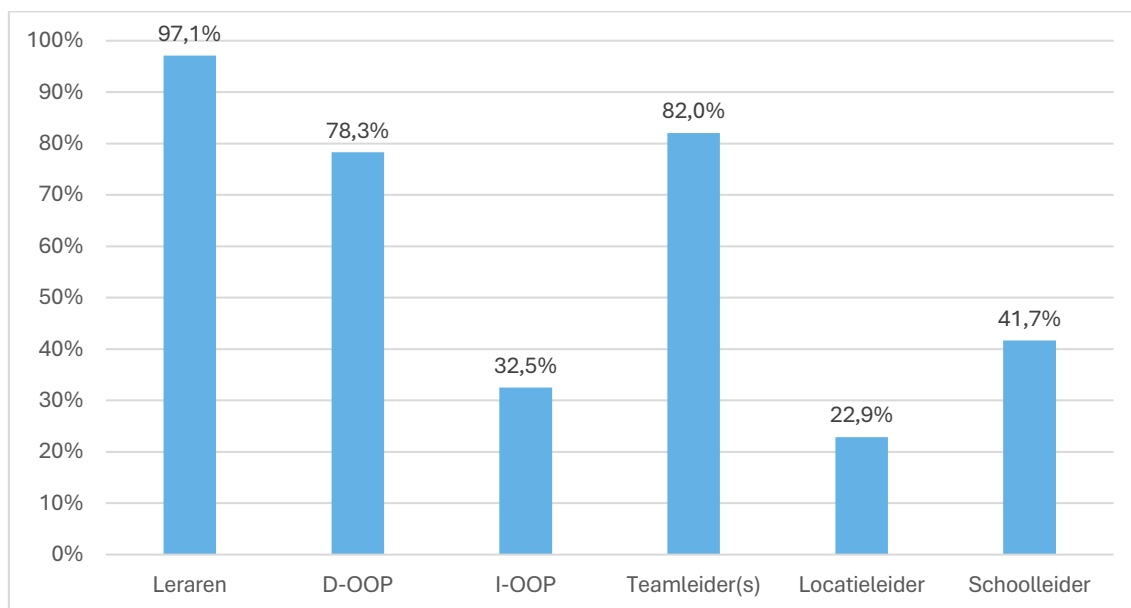
	Totaal (n = 126)
Leer-werkhouding	66,7%
Sociaal-emotionele ontwikkeling	63,5%
Welzijn	37,3%

3.3.5 Gebruikers

De vervolgvraag is wie (de resultaten van) het leerlingvolgsysteem op de vestigingen gebruiken. Ook hier konden meerdere antwoorden gekozen worden. Zoals in Figuur 3.3 te zien, zijn de meest gekozen antwoorden leraren, teamleider(s) en direct onderwijsondersteunend personeel (D-OOP).⁴

⁴ We onderscheiden direct ondersteunend personeel (bijv. orthopedagoog, onderwijscoördinator) en indirect ondersteunend personeel (staffunctionaris, administratieve functies).

Figuur 3.3 Wie maakt gebruik van het lvs? (n = 403)



Noot. D-OOP = direct onderwijsondersteunend personeel; I-OOP = Indirect onderwijsondersteunend personeel.

Waar het lvs voor gebruikt wordt, verschilt per functiegroep. Leraren en D-OOP gebruiken het lvs het meest op leerlingniveau, namelijk voor:

1. het geven van feedback over onderwijsresultaten aan individuele leerlingen;
2. advies/besluit over overgang/doubleren van leerlingen;
3. advies/besluit over doorstroom van leerlingen naar een andere onderwijssoort ('opstroom'/'afstroom').

Indirect OOP (I-OOP), teamleiders, locatieleiders en schoolleiders gebruiken het lvs meer op schoolniveau, zoals het verbeteren van onderwijsprocessen op school en het monitoren van kwaliteitsbeleid.

Uit de gesprekken met scholen halen we op dat iedereen die toegang heeft tot het leerlingadministratiesysteem, over het algemeen ook toegang heeft tot de lvs-resultaten. Meestal gaat het dan om alle docenten, D-OOP en schoolleiding. Afhankelijk van het beleid en de praktijk in de school verschilt het welke medewerker daadwerkelijk de lvs-resultaten gebruikt. Het kan de docent van het betreffende vakgebied zijn, maar ook de mentor. Een school geeft aan dat een taalcoördinator de resultaten van het lvs-Nederlands analyseert en vervolgens naar de docent communiceert over het gevolg (bijvoorbeeld: de leerling heeft extra steunlessen nodig).

Wanneer de resultaten van de lvs-toetsen ingelezen worden in het leerlingadministratiesysteem, hebben leerlingen en ouders over het algemeen ook toegang. Een van de leerlingen gaf aan de resultaten van de lvs-toets in een envelop mee naar huis te krijgen. Een school gaf aan de resultaten altijd te bespreken met ouders in een oudergesprek.

We vroegen de schoolleiders aan te geven of het gebruikers over het algemeen lukt om de gegevens uit het lvs op de juiste manier te interpreteren. Het beeld is dat gebruikers hier over het algemeen goed toe in staat zijn. De schoolleiders geven wel aan dat docenten en I-OOP hier net wat meer moeite mee hebben dan D-OOP en de schoolleiding.

3.4 Aanbod lvs

Om in kaart te brengen hoe de inhoud van de lvs'en zich verhouden tot de huidige wetgeving hebben we deze inhoud vergeleken met de zogenaamde definitieve conceptkerndoelen van de Stichting Leerplanontwikkeling (SLO). Dit betreft een nieuwe generatie kerndoelen die concreter geformuleerd zijn dan de vorige en bestaan uit zowel aanbods- als beheersings- en ervaringsdoelen (SLO, 2023 en 2024). Wat als eerste opvalt is dat er wel kerndoelen met puntsgewijze uitwerkingen van de doelzinnen zijn geformuleerd, maar dat een niveauaanduiding bij de kerndoelen en uitgewerkte doelzinnen ontbreekt. Voor de referentiekaders taal en rekenen is dat wel gedaan, waardoor duidelijker is wanneer een leerling wat zou moeten kennen en kunnen; hetgeen bevorderlijk is voor de standaardisatie van de lvs'en. De lvs-aanbieders hebben aangegeven dat zij zich om deze reden vooral op het referentiekader (zie Bijlage 4 voor een gedetailleerde uitleg over het referentiekader en referentieniveaus) taal en rekenen hebben gericht bij de ontwikkeling van hun lvs Nederlands en rekenen-wiskunde.

Van de vier verschillende lvs-aanbieders voor het voortgezet onderwijs bieden alle aanbieders volgttoetsen aan voor de cognitieve vaardigheden Nederlands en rekenen-wiskunde, en drie van de vier aanbieders hebben ook een lvs voor Engels. Daarnaast is er één aanbieder die ook toetsen aanbiedt voor de andere moderne vreemde talen, namelijk Frans, Duits en Spaans. Naast de toetsen voor cognitieve vaardigheden bieden twee aanbieders instrumenten aan voor het in kaart brengen van de "soft skills" van de leerling. Deze instrumenten richten zich op persoonlijkheidskenmerken van de leerling (zoals zelfvertrouwen, nieuwsgierigheid, vindingrijkheid en flexibiliteit), de leeraanpak van de leerling (motivatie, plannen, werkhouding), en de sociale context van de leerling (beleving van school en de thuissituatie). Deze instrumenten worden ingevuld op basis van zelfevaluatie van de leerling, bij één aanbieder optioneel aangevuld door de evaluatie van de leerkracht en/of ouder/verzorger.

Hieronder volgen de resultaten van de inhoudelijke vergelijking van de lvs'en met de definitieve conceptkerndoelen voor Nederlands (zie paragraaf 3.4.1), Engels (zie 3.4.2) en rekenen-wiskunde (zie 3.4.3) en met het referentiekader taal en rekenen (zie 3.4.5). Bij elk conceptkerndoel heeft de SLO concrete uitwerkingen gegeven in de vorm van doelzinnen, waardoor het doel meer inhoudelijke betekenis krijgt. Voor de kerndoelen zijn we nagegaan in welke mate de puntsgewijze uitwerkingen van doelzinnen voor de onderbouw vo – aangeduid in de brontekst als 'Het gaat hierbij om' – inhoudelijk terugkomt in de beschikbare lvs'en. We hebben ons in eerste instantie gebaseerd op de ontvangen documentatie van de vier toetsaanbieders. Daarmee hebben we conceptschema's geconstrueerd voor iedere aanbieder afzonderlijk. Vervolgens heeft iedere aanbieder ons conceptschema geverifieerd, aangepast en aan ons geretourneerd. In het algemeen bleken onze conceptschema's een geringe onderschatting te zijn van de zelfbeoordeling van dekkingsgraad door aanbieders. Voorbeelden hiervan bij de kerndoelen voor Nederlands zijn luisteren met diep begrip, doelgericht gesprekken voeren en reflecteren op taalactiviteiten (kerndoelen 10 en 11). Mogelijk ligt hieraan ten grondslag dat we niet de beschikking hadden over de feitelijke toetsen van de aanbieders.

De dekkingsgraad van een kerndoel is gecategoriseerd als: 'niet', 'deels', en 'wel' gedekt. De middencategorie 'deels' loopt feitelijk uiteen van bijna niet tot grotendeels wel gedekt. Voor de inhoudelijke vergelijking van de lvs'en met het referentiekader taal en rekenen (zie 3.4.5) hebben we grotendeels dezelfde methode gevolgd; de vergelijking is nog wat verfijnder doordat we onderscheid gemaakt hebben in dekkingsgraad van lvs'en per doelomschrijving voor 1F- en 2F-niveau binnen een subdomein van taal en rekenen.

Bij de bespreking van de resultaten blijven we weg bij het onderscheid tussen ‘wel’ en ‘deels’ gedekt, omdat we willen focussen op de definitieve conceptkerndoelen of referentieniveaus die bij geen enkele lvs-aanbieder terugkomen. Als een kerndoel/referentieniveau geheel of deels wordt gedekt door ten minste één van de vier aanbieders, dan gaan we ervan uit dat het voor de andere aanbieders inhoudelijk haalbaar is om ook een dergelijke toets te ontwikkelen. Als een kerndoel in zijn geheel *niet* gedekt wordt de beschikbare volgroepen, dan is het beleidsmatig belangrijk de reden hiervan te achterhalen. Mogelijk speelt hier dat het toetsen van het betreffende definitieve conceptkerndoel inhoudelijk niet haalbaar of principieel onwenselijk is, commercieel niet interessant genoeg is of dat het kerndoel pas aan de orde is bij het eindexamen en dus niet van toepassing voor leerlingen in de onderbouw vo. De aanbieders hebben hierover een toelichting gegeven die we in de paragrafen 3.4.4 en 3.4.6 hebben samengevat.

3.4.1 Dekking kerndoelen Nederlands

Tabel 3.4 toont een overzicht van de mate waarin de definitieve conceptkerndoelen voor Nederlands (SLO, sept 24) inhoudelijk terugkomen in de lvs'en van de vier aanbieders. Kerndoelen 1 (taal- en leesomgeving) en 2 (taal in de leergebieden) hebben betrekking op scholen in plaats van leerlingen en vallen daarom buiten het bestek van deze ontwerpstudie.

Tabel 3.4 Overzicht van de dekking van de definitieve conceptkerndoelen Nederlands door vier lvs-aanbieders voor de onderbouw van het voortgezet onderwijs

Kerndoelnummer en omschrijving	Doelzin: De leerling ...	Dekking van kerndoel volgens lvs-aanbieders		
		wel	deels	niet
Domein: communicatie				
3. Luisteren en lezen met begrip	... toont begrip van zakelijke en literaire teksten.	2	2	0
4. Luisteren en lezen met diep begrip	... evalueert en reflecteert op zakelijke en literaire teksten.	1	3	0
5. Doelgericht spreken en schrijven	... spreekt en schrijft afgestemd op doel, publiek en context.	1	0	3
6. Creatief taal gebruiken	... gebruikt taal op een creatieve manier.	0	0	4
7. Schrijven om te leren	... schrijft om tot kennisopbouw of begrip te komen.	1	1	2
8. Bronnen verkennen	... verkent de betrouwbaarheid van verschillende bronnen.	0	1	3
9. Doelgericht gesprekken voeren	... voert gesprekken afgestemd op doel, gesprekspartner en context.	1	0	3
10. Gesprekken voeren om te leren	... voert gesprekken om tot kennisopbouw, begrip of een aanpak te komen.	0	0	4
11. Reflecteren op taalactiviteiten	... reflecteert op het proces en evalueert het product van een taalactiviteit.	0	1	3

Kerndoelnummer en omschrijving	Doelzin: De leerling ...	Dekking van kerndoel volgens lvs-aanbieders		
		wel	deels	niet
Domein: taal				
12. Vorm en betekenis beschouwen	... beschouwt de relatie tussen vorm en betekenis van taal.	0	3	1
13. Spelling, formulering en interpunctie	... toont inzicht in regels en procedures voor spelling, formulering en interpunctie.	3	1	0
14. Talige identiteit	... verkent hoe je met taal uiting geeft aan je identiteit.	0	0	4
15. Taalvariatie en taalverandering	... verkent taalvariatie en taalverandering in het Nederlandse taalgebied.	0	0	4
16. Zakelijke tekstgenres (3h/v)	... toont inzicht in zakelijke tekstgenres.	1	2	1
17. Argumentatief taalgebruik (3h/v)	... toont inzicht in de overtuigingskracht van argumentatie.	0	4	0
Domein: literatuur				
18. Leesvoorkeur	... ontwikkelt een eigen leesvoorkeur.	0	0	4
19. Waarde van literatuur	... verkent de waarde van literatuur.	0	0	4
20. Verhalende teksten	... toont inzicht in verhalende teksten (met behulp van kennis over verteltechnieken).	0	0	4
21. Literaire genres	... toon inzicht in genre-kenmerken van literatuur.	0	0	4
22. Literatuur uit verschillende tijden	... verkent literatuur die geschreven is in verschillende periodes en contexten.	0	0	4
23. Literatuur in tijd en context (3h/v)	... legt verbanden tussen de inhoud van literatuur en de periode en context waarin deze is geschreven.	0	0	4

Tabel 3.4 toont dat 10 van de 21 (48%) definitieve conceptkerndoelen Nederlands door geen enkele lvs-aanbieder is verwerkt in een lvs. Het gaat daarbij om creatief taalgebruik (6), gesprekken voeren om te leren (10), het inzetten van taal om uiting te geven aan persoonlijke of groepsidentiteit (taalcultuur) (14 en 15), en alle kerndoelen die vallen onder het domein literatuur (18 tot en met 23). De overige 52% van de definitieve conceptkerndoelen Nederlands komen wel of ten dele terug in een lvs volgens de aanbieders.

Kijken we vervolgens in detail naar de definitieve conceptkerndoelen Nederlands die wél bij ten minste één aanbieder terugkomen, dan valt bij het domein communicatie op dat begrijpend luisteren (3 en 4) beter wordt gedekt dan schrijven en (doelgerichte) gespreksvoering (5, 7, 9). Verder krijgen het reflecteren op taalactiviteiten (11) en bronverkenning (8) weinig aandacht in de huidige lvs'en: slechts één van de vier aanbieders zegt hier ten dele iets mee te doen. Bij het tweede domein taal wordt het kerndoel spelling, formulering en interpunctie (13) het beste gedekt door de vier aanbieders, direct gevolgd door argumentatief taalgebruik voor 3 havo/vwo (17). Minder frequent terugkomend zijn inzicht in zakelijke tekstgenres voor 3 havo/vwo (16) en beschouwen van de relatie tussen vorm en betekenis (12): één van de vier aanbieders laat deze thema's geheel buiten zijn lvs.

3.4.2 Dekking kerndoelen Engels

Tabel 3.5 geeft een beeld van de mate waarin de definitieve conceptkerndoelen Engels (SLO, 2024) inhoudelijk terugkomen in twee lvs'en. De overig twee aanbieders waren door omstandigheden niet tot verificatie in staat of bieden geen lvs Engels aan. Kerndoel 1 (rijke taal en leesomgeving aanbieden) heeft betrekking op scholen en komt daarom niet terug in Tabel 3.5.

Tabel 3.5 Overzicht van de dekking van de definitieve conceptkerndoelen Engels door twee lvs-aanbieders voor de onderbouw van het voortgezet onderwijs

Kerndoelnummer en omschrijving	Doelzin: De leerling ...	Dekking van kerndoel volgens lvs-aanbieders		
		wel	deels	niet
Domein: communicatie				
2. Kijk- en luistervaardigheid	... toont begrip van auditieve en audiovisuele Engelstalige fictie- en non-fictieteksten.	1	0	1
3. Leesvaardigheid	... toont begrip van schriftelijke Engelstalige fictie- en non-fictieteksten.	2	0	0
4. Mondelinge communicatie	... spreekt en voert gesprekken in de Engelse taal, afgestemd op doel, publiek en context.	0	0	2
5. Schriftelijke communicatie	... schrijft in het Engels, afgestemd op doel, publiek en context.	1	0	1
6. Doelgerichte informatie verwerking	... verwerkt Engelstalige teksten mondeling en schriftelijk om tot kennisopbouw en begrip te komen.	0	2	0
7. Creatieve expressie	... experimenteert met de Engelse taal op een creatieve manier.	0	0	2
Domein: taal- en cultuurbewustzijn				
8. Taalleerproces en meertaligheid	... verkent hoe het eigen taalleerproces wordt bevorderd.	0	0	2
9. Cultuurgebonden aspecten en culturele diversiteit	... beschrijft cultuurgebonden aspecten en culturele diversiteit in fictie en non-fictie.	0	0	2
10. Interculturele communicatieve competentie	... stemt de communicatie af op de culturele en talige achtergrond van de ontvanger.	0	0	2

Ruim de helft van de definitieve conceptkerndoelen Engels (5 van de 9 kerndoelen) komt inhoudelijk *niet* terug in de twee lvs'en Engels waarover we informatie hebben ontvangen (Tabel 3.5). Dat is iets meer dan voor lvs-Nederlands (zie 3.4.1). Het gaat dan om de kerndoelen mondelinge communicatie (4) en creatieve expressie (7) binnen het communicatiedomein, en om alle kerndoelen binnen het domein taal- en cultuurbewustzijn (8, 9, 10).

De overige 45% van de definitieve conceptkerndoelen Engels komen wel of ten dele terug bij de twee lvs'en. Meer in detail tredend zien we dat Engels leesvaardigheid (3) inhoudelijk beter wordt gedekt dan doelgericht informatieverwerking (6), gevolgd door kijk- en luistervaardigheid (2) en schriftelijke communicatie. De twee laatste genoemde kerndoelen zijn bij (maar) één van de twee aanbieders aanwezig in hun lvs-Engels.

3.4.3 Dekking kerndoelen rekenen-wiskunde

Een beeld van de mate waarin de geactualiseerde kerndoelen rekenen en wiskunde (SLO, 2024b) terugkomen in de omschrijvingen van de voltooitsen van de vier aanbieders van rekenen en wiskunde voltooitsen wordt getoond in Tabel 3.6. De schoolkerndoelen 13 (wiskundige attitude) en 15 (wiskunde in andere leergebieden) komen niet terug in deze tabel, evenals de specifieke aanvullende kerndoelen voor leerlingen in 3 havo/vwo.

Tabel 3.6 Overzicht van de dekking van de definitieve conceptkerndoelen rekenen-wiskunde door vier lvs-aanbieders voor de onderbouw van het voortgezet onderwijs

Kerndoelnummer en omschrijving	Doelzin: De leerling ...	Dekking van kerndoel volgens lvs-aanbieders		
		wel	deels	niet
Domein A: wiskundige concepten				
1. Getallen en grootheden, en algebra	... redeneert en rekent met getallen en grootheden.	4	0	0
2. Vergelijkingen	... gebruikt wiskundige vergelijkingen.	2	1	1
3. Data	... interpreteert, representeert en analyseert data.	2	2	0
4. Kans	... redeneert en rekent met kansen.	2	1	1
5. Patronen en verbanden	... analyseert en redeneert over patronen en verbanden.	4	0	0
6. Vorm en ruimte	... analyseert en redeneert over twee en driedimensionale ruimte.	4	0	0
Domein B: wiskundige denk-/werkwijzen				
7. Wiskundig probleem oplossen	... lost wiskundige problemen en toepassingsproblemen op.	1	0	3
8. Wiskundig modelleren	... maakt en gebruikt wiskundige modellen.	1	0	3
9. Aantonen	... toont de juistheid van een wiskundige bewering of redenering aan.	0	1	3
10. Gebruiken en beschrijven van algoritmes	... bedenkt en beschrijft algoritmes.	0	0	4
11. Gebruiken van wiskundetaal en wiskundige representaties	... gebruikt wiskundetaal en wiskundige representaties.	2	2	0
12. Gebruiken van wiskundige instrumenten	... gebruikt meetinstrumenten en andere wiskundige instrumenten.	2	1	1
14. Wiskunde in de werkelijkheid	... herkent en gebruikt wiskunde in dagelijkse, maatschappelijke en beroepsmatige situaties.	2	2	0

Voor de lvs'en rekenen-wiskunde hebben we gevonden dat de dekking van de definitieve conceptkerndoelen aanzienlijk beter is dan voor de lvs'en Nederlands en Engels, namelijk 92%. Tabel 3.6 toont dat alleen het kerndoel gebruiken en beschrijven van algoritmen (10) door geen van de aanbieders inhoudelijk is verwerkt in hun lvs.

Als we dan inzoomen in Tabel 3.6 op de mate van dekking per domein, dan zien we dat de kerndoelen van het domein wiskundige concepten (A) beter terugkomen in de lvs'en dan de kerndoelen van het domein wiskundige denk-/werkwijzen (B). Bij het eerste domein wiskundige concepten worden de kerndoelen getallen en grootheden, en algebra (1), patronen en verbanden (5) en vorm en ruimte (6) goed gedekt door alle vier aanbieders, direct gevolgd door data (3). De dekkingsgraad voor vergelijkingen (2) en kans (4) is wat lager: drie van de vier aanbieders hebben aangegeven dit wel of deels te hebben meegenomen in hun lvs-toetsen rekenen-wiskunde. Bij het tweede domein wiskundige denk-/werkwijzen zien we dat het kerndoel gebruiken van wiskundetaal en wiskundige representaties (11) en wiskunde in de werkelijkheid (14) het hoogst scoren, gevolgd door gebruiken van wiskundige instrumenten (bijv. een rekenmachine)(12). Het minder frequent aan de orde zijn wiskundig probleem oplossen (7), wiskundig modelleren (8) en aantonen (9): drie van de vier lvs-aanbieders stellen dat deze kerndoelen niet opgenomen zijn in hun lvs.

3.4.4 Motivatie voor dekking kerndoelen door toetsaanbieders

Bij verificatie-interviews met afzonderlijke lvs-aanbieders hebben we gevraagd naar de achtergronden voor de inhoudelijke aansluiting van hun lvs bij de definitieve conceptkerndoelen, referentiekader en ERK-niveaus. Daaruit volgt dat de definitieve conceptkerndoelen niet volledig worden gedekt, omdat ze voornamelijk op het referentiekader taal en rekenen en de ERK-referentieniveaus voor Engels zijn gebaseerd. Deze vindt men concreter dan de definitieve conceptkerndoelen van de SLO en daardoor beter bruikbaar voor toetsontwikkeling. Bovendien sluit het referentiekader beter aan bij de behoeften vanuit het onderwijsveld.

De lvs-aanbieders merken op dat de (hoofd)domeinen van de definitieve conceptkerndoelen breder geformuleerd zijn dan voorheen; ze combineren meerdere vaardigheden, zoals luisteren en lezen bij Nederlands, terwijl bij de huidige volgtoetsen deze vaardigheden gesplitst zijn – in overeenstemming met de referentiekaders taal en rekenen – en hierover apart gerapporteerd wordt. Een aanbieder stelt dat het rapporteren over de definitieve conceptkerndoelen moeilijker is, omdat de diversiteit en het aantal subdomeinen binnen een hoofddomein zijn toegenomen. Als een leerling uitvalt op bijvoorbeeld 'meten en meetkunde' bij het referentiekader rekenen, dan geeft dat de wiskundedocent veel concretere aanwijzingen waarop hij of zij kan inzetten dan bij 'vorm en ruimte' van het overeenkomstige definitieve conceptkerndoel. Ook de betrouwbaarheid van de toetsing dreigt minder te worden bij meer subdomeinen.

Verder geven de lvs-aanbieders aan dat het noodzakelijk is om voldoende items per subdomein in een toets te hebben om er goed over te kunnen rapporteren, maar een toets mag ook niet te lang worden of te veel tijd innemen (zie ook hoofdstuk 5.2 over Cognitief Diagnostische Modellen). Aangezien de beschikbare toetstijd beperkt is, moeten toetsontwikkelaars keuzes maken. Het toevoegen van extra onderdelen gaat ten koste van de tijd die beschikbaar is voor het toetsen andere vaardigheden. Bij rekenen-wiskunde speelt bijvoorbeeld dat het toetsen van wiskundig probleemoplossen en modelleren te veel tijd per item vraagt. Er kunnen niet veel lange items in een toets zitten, want dan duurt de totale toets te lang. Maar een gering aantal lange items kan de inhoudelijke dekking van de totale toets in gevaar brengen. Een aanbieder voegt hieraan toe dat lvs'en "geen slap aftreksel van het gehele curriculum moeten worden; je wilt niet de illusie geven dat iets gedekt wordt, als dat niet zo is". Dit komt overeen met wat we als risico's hebben gehoord van scholen en organisaties zoals de VO-raad en SLO.

De aanbieders geven aan dat ze bepaalde kerndoelen bewust niet meenemen in hun lvs. Sommige kerndoelen zijn lastig objectief meetbaar. Een aanbieder stelt dat zijn lvs zich beperkt tot beheersingsdoelen: ervaringsdoelen zoals verkennen, reflecteren en evalueren blijven buiten beschouwing. Daarnaast lenen bepaalde definitieve conceptkerndoelen zich beter voor behandeling in de klas, zoals creatief taal gebruiken (6), gesprekken voeren om te leren (10), de ontwikkeling van talige identiteit (14) en leesvoorkeur (18). Toetsing ervan zou leiden tot een vereenvoudiging die geen recht doet aan de oorspronkelijke leerintentie. Bovendien zijn dergelijke doelen slecht te operationaliseren in toetsvragen. Praktische beperkingen spelen ook een rol bij de keuze om een kerndoel mee te nemen bij de ontwikkeling van een lvs. Sommige vaardigheden, zoals spreken en schrijven bij Nederlands en Engels, zijn (nog) moeilijk digitaal te toetsen.

Verder stellen toetsaanbieders dat ze vooral werken vanuit de wensen van de klant. Ze voeren daarvoor eigen onderzoeken naar klantervaringen uit en peilen met enquêtes het draagvlak voor wat ze willen gaan doen. Ze moeten rekening houden met commerciële haalbaarheid en efficiëntie. Voor bepaalde vaardigheden zoals burgerschap was er in het verleden weinig behoefte aan toetsing, al lijkt dit nu te veranderen. Zolang de definitieve conceptkerndoelen geen wettelijke status hebben en een concretere uitwerking door de SLO niet beschikbaar is, worden ze nog niet in de volgtoetsen verwerkt. Volgtoetsen volgen op het curriculum. Pas wanneer de kerndoelen officieel van kracht worden en in de lesmethoden verwerkt worden, zullen ze worden meegenomen bij de lvs-ontwikkeling.

3.4.5 Vergelijking referentiekaders taal en rekenen met inhoud volgtoetsen

De vier lvs-aanbieders hebben aangegeven dat ze de inhoud van hun lvs Nederlands en rekenen-wiskunde voornamelijk op het referentiekader taal en rekenen hebben gebaseerd. We hebben daarom onderzocht in hoeverre de omschrijving van de referentieniveaus 1F en 2F van de (sub)domeinen voor taal en rekenen (Expertgroep doorlopende leerlijnen taal en rekenen, 2008, 2009) inhoudelijk worden gedekt voor de lvs'en Nederlands en rekenen-wiskunde van de vier aanbieders.

De gehanteerde werkwijze komt overeen met die voor de definitieve conceptkerndoelen. Op basis van de ontvangen documentatie van de vier toetsaanbieders hebben we een conceptdekkingsschema voor iedere aanbieder gemaakt. Dit schema is gecontroleerd en aangepast door iedere aanbieder en vervolgens geretourneerd. In tegenstelling tot de conceptschema's voor kerndoelen, werden onze conceptschema's door de aanbieders in vrij sterke mate bijgesteld: de zelfbeoordeling door de aanbieders resulteerde in het algemeen tot een beduidend hogere dekkingsgraad. Voorbeelden hiervan voor rekenen-wiskunde zijn de kerndoelen vergelijkingen, kand en gebruik van wiskundige instrumenten (kerndoelen 2, 4 en 12) en voor Nederlands luisteren met diep begrip, doelgericht gesprekken voeren en reflecteren op taalactiviteiten (kerndoelen 10 en 11).

Hierna volgt eerst de bespreking van de inhoudelijke dekking voor domeinen voor taal, gevolgd door die voor rekenen.

Taal

Tabel 3.7 toont in hoeverre de omschrijving van de referentieniveaus 1F en 2F van de (sub)domeinen voor taal inhoudelijk worden gedekt voor de vier lvs-aanbieders. Daarbij willen we enkele opmerkingen vooraf maken:

- Het niveauverschil tussen 1F en 2F bij luisteren zit voornamelijk in de lengte van de te beluisteren tekst: bij 1F gaat het om kortere teksten, met heldere structuur en lage informatiedichtheid; bij 2F om langere teksten, met heldere structuur en redelijke informatiedichtheid.
- Bij het domein lezen van zakelijke teksten zijn het samenvatten op niveau 1F en vloeiend lezen en woordenschat op niveau 2F niet van toepassing (Expertgroep doorlopende leerlijnen taal en rekenen, 2009).
- De Begrippenlijst als subdomein van het domein taalverzorging en begrippenlijst is in de brontekst (Expertgroep doorlopende leerlijnen taal en rekenen, 2008, 2009) niet nader uitgewerkt en komt daarom niet terug in Tabel 3.7. Het gaat bij de lvs'en vooral om het kunnen toepassen van deze begrippen, hetgeen tot uiting kan komen bij de leestoets.
- De onderwerpen tekstkennis, stilistiek en semantiek, morfologie, opmaak en klanken en ook het onderwerp spelling van het subdomein taalverzorging betreffen een combinatie van onderwerpen waarover de aanbieders niet of nauwelijks specifieke informatie hebben gegeven. Nadere uitsplitsingen zijn daarom achterwege gelaten.

Tabel 3.7 toont in de eerste regel bijvoorbeeld dat bij gesprekken het leerdoel deelnemen aan discussie en overleg op 1F- en 2F-niveau wel terugkomt bij één aanbieder, maar niet bij de overige 3 lvs-aanbieders.

Tabel 3.7 Overzicht van de mate van inhoudelijke dekking van het referentiekader **taal** door de vier lvs-aanbieders voor de onderbouw van het voortgezet onderwijs

Onderwerp	Niveau	Dekking van referentieniveaus door aanbieders		
		wel	deels	niet
Domein: Mondelinge Taalvaardigheid				
Subdomein: Gesprekken				
Taken - Deelnemen aan discussie en overleg	1F en 2F	1	0	3
Taken - Informatie uitwisselen	1F en 2F	1	0	3
Subdomein: Luisteren				
Luisteren naar instructies	1F	2	0	2
	2F	2	0	2
Luisteren als lid van een live publiek	1F	2	0	2
	2F	2	0	2
Luisteren naar radio, televisie en gesproken tekst op internet	1F	2	0	2
	2F	2	0	2
Subdomein: Spreken				
Een monoloog houden	1F	2	0	2
	2F	2	0	2

Onderwerp	Niveau	Dekking van referentieniveaus door aanbieders		
		wel	deels	niet
Domein: Lezen				
Subdomein: Zakelijke Teksten (informatieve teksten, instructies en betogende teksten)				
Vloeiend lezen en woordenschat	1F	4	0	0
	2F	n.v.t.	n.v.t.	n.v.t.
Lezen en begrijpen van zakelijke teksten	1F	4	0	0
	2F	4	0	0
Lezen en interpreteren van zakelijke teksten	1F	4	0	0
	2F	4	0	0
Lezen en evalueren van zakelijke teksten	1F	4	0	0
	2F	4	0	0
Lezen en samenvatten van zakelijke teksten	1F	n.v.t.	n.v.t.	n.v.t.
	2F	3	0	1
Opzoeken en lezen van informatie	1F	3	1	0
	2F	3	0	1
Subdomein: Fictionele, narratieve en literaire teksten				
Lezen, begrijpen en samenvatten van fictionele, narratieve en literaire teksten	1F (en 2F)	1	0	3
Lezen en interpreteren van fictionele, narratieve en literaire teksten	1F (en 2F)	1	0	3
Lezen en evalueren van fictionele, narratieve en literaire teksten	1F (en 2F)	1	0	3
Domein: Schrijven				
Schrijven met samenhang en afstemming op doel en publiek, met aandacht voor (gevarieerd) woordgebruik, correcte spelling, interpunctie, grammatica en leesbaarheid	1F en 2F	2	0	2
Domein: Begrippenlijst en taalverzorging				
Subdomein: Taalverzorging				
Leestekens (interpunctie)	1F	4	0	0
	2F	3	1	0
Woordsoorten	1F	4	0	0
	2F	n.v.t.	n.v.t.	n.v.t.
Grammaticale kennis	1F	4	0	0
	2F	4	0	0
Tekstkennis, stilistiek en semantiek, morfologie, opmaak en klanken (een verzamelcategorie)	1F en 2F	2	2	0
Spelling (een verzamelcategorie)	1F en 2F	4	0	0

In tegenstelling tot de dekking van de definitieve conceptkerndoelen voor Nederlands laat het referentiekader géén inhoudelijke hiaten zien. Tabel 3.7 toont dat alle onderwerpen binnen alle subdomeinen, zowel op 1F als 2F, bij ten minste één lvs-aanbieder aan bod komen. Er zijn wel verschillen in dekkingsgraad tussen de subdomeinen. We bespreken hier de meest in het oogspringende.

Mondelinge Taalvaardigheid

Het subdomein gesprekken wordt minder goed gedekt door lvs'en dan de subdomeinen spreken en luisteren. Er is één aanbieder die aangegeven heeft dat gespreksvoering, in de vorm van deelname aan discussies en overleg en voor informatie-uitwisseling, wel gedekt wordt door zijn volgtoetsen. De overige drie lvs-aanbieders hebben geen onderdeel gespreksvoering. Spreken en luisteren komen terug bij twee van de vier lvs-aanbieders. Bij spreken gaat het inhoudelijk om een monoloog houden door de leerling. Bij luisteren gaat het om luisteren naar instructies, naar radio en televisie en gesproken tekst op het internet, en om luisteren als lid van een live publiek.

Lezen

Bij het hoofddomein Lezen richten de aanbieders zich op vooral op zakelijke teksten: informatieve, betogende teksten en instructies. Slechts één aanbieder heeft aangegeven ook fictionele, narratieve en literaire teksten te gebruiken voor zijn lvs Nederlands. Bij de leestoetsen over zakelijke teksten wordt het begrijpen, interpreteren en evalueren van de tekst getoetst door alle vier de aanbieders, vaak in combinatie met het toetsen van de woordenschat (1F). Daarnaast is ook samenvatten (2F) vaak onderdeel van zakelijke leestoetsen: drie van de vier lvs-aanbieders zeggen dit te doen. De aanbieder die fictionele, narratieve en literaire teksten gebruikt voor zijn lvs, toetst hierbij het begrijpen, interpreteren en evalueren van dergelijke teksten. Verder valt op dat het opzoeken van informatie in naslagwerken of schema's goed wordt gedekt door lvs'en: bij vier (1F) en drie (2F) aanbieders is dit onderdeel van hun lvs Nederlands.

Schrijven

Het hoofddomein schrijven, waarbij het gaat om het schrijven met samenhang en afstemming op doel en publiek, is door twee van de vier aanbieders meegenomen in hun lvs Nederlands.

Begrippenlijst en taalverzorging

Bij het domein begrippenlijst en taalverzorging gaat het feitelijk alleen om het subdomein taalverzorging. Taalverzorging wordt inhoudelijk goed gedekt door de lvs'en: alle onderwerpen die vallen onder taalverzorging worden getoetst door de vier aanbieders. Er is een klein verschil in dekking tussen de onderwerpen binnen taalverzorging: tekstkennis, stilistiek en semantiek, morfologie, opmaak en klanken (1F en 2F) en leestekens (2F) komen iets minder consequent terug dan de overige onderwerpen.

Rekenen

De Expertgroep doorlopende leerlijnen taal en rekenen (2009) heeft voor rekenen vier hoofddomeinen onderscheiden: getallen, verhoudingen, meten en meetkunde en verbanden. Binnen elk hoofddomein zijn drie subdomeinen beschreven: notatie, taal en betekenis, met elkaar in verband brengen en gebruiken. Voor bijna ieder subdomein zijn drie doelomschrijvingen/leerdoelen geformuleerd:

- *paraat hebben*: kennis van feiten en begrippen, reproduceren, routines, technieken;
- *functioneel gebruiken*: kennis van een goede probleemaanpak, het toepassen, het gebruiken binnen en buiten het schoolvak;
- *weten waarom*: begrijpen en verklaren van concepten en methoden, formaliseren, abstraheren en generaliseren, blijk geven van overzicht.

Tabel 3.8 geeft inzicht in hoeverre de referentieniveau 1F en 2F voor rekenen inhoudelijk worden gedekt door de huidige volgtolsten. We hebben de oorspronkelijke tabellen van de Expertgroep doorlopende leerlijnen taal en rekenen (2009) samengenomen en waar mogelijk ingedikt. Negen doelomschrijvingen voor rekenen zijn niet nader uitgewerkt door de Expertgroep; bij acht van de negen gaat het om *weten waarom*. Ze zijn in Tabel 3.8 aangegeven als niet van toepassing (n.v.t.) en niet meegenomen in het berekenen van de dekkingspercentages.

Tabel 3.8 Overzicht van de mate van inhoudelijke dekking van het referentieniveaus **rekenen** door de vier aanbieders van leerlingvolgtolsten in de onderbouw van het voortgezet onderwijs

Niveau	Doelomschrijving: Leerlingen ...	Dekking van referentieniveaus door aanbieders		
		wel	deels	niet
Domein: Getallen				
Subdomein: Notatie, taal en betekenis				
1F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
	... weten waarom	4	0	0
2F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
	... weten waarom	4	0	0
Subdomein: Met elkaar in verband brengen				
1F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
	.. weten waarom	4	0	0
2F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
	... weten waarom	2	1	1
Subdomein: Gebruiken				
1F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
	... weten waarom	2	0	2
2F	... hebben paraat	4	0	0
	... gebruiken functioneel	2	1	1
	... weten waarom	3	1	0

		Dekking van referentieniveaus door aanbieders		
Niveau	Doelomschrijving: Leerlingen ...	wel	deels	niet
Domein: Verhoudingen				
Subdomein: Notatie, taal en betekenis				
1F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
	... weten waarom	n.v.t.	n.v.t.	n.v.t.
2F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
Subdomein: Met elkaar in verband brengen				
1F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
	... weten waarom >>n.v.t.	n.v.t.	n.v.t.	n.v.t.
2F	... hebben paraat	4	0	0
	... gebruiken functioneel	3	0	1
	... weten waarom >>n.v.t.	n.v.t.	n.v.t.	n.v.t.
Subdomein: Gebruiken				
1F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
	... weten waarom	4	0	0
2F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
	... weten waarom	2	1	1
Domein: Meten en meetkunde				
Subdomein: Notatie, taal en betekenis				
1F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
	... weten waarom	3	1	0
2F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
	... weten waarom	n.v.t.	n.v.t.	n.v.t.
Subdomein: Met elkaar in verband brengen				
1F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
	... weten waarom	4	0	0
2F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
	... weten waarom	4	0	0
Subdomein: Gebruiken				
1F	... hebben paraat	4	0	0
	... gebruiken functioneel	3	0	1
	... weten waarom	n.v.t.	n.v.t.	n.v.t.
2F	... hebben paraat	4	0	0
	... gebruiken functioneel	3	0	1
	... weten waarom	3	0	1

		Dekking van referentieniveaus door aanbieders		
Niveau	Doelomschrijving: Leerlingen ...	wel	deels	niet
Domein: Verbanden				
Subdomein: Notatie en betekenis				
1F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
	... weten waarom	3	0	1
2F	... hebben paraat	4	0	0
	... gebruiken functioneel	n.v.t.	n.v.t.	n.v.t.
	... weten waarom	n.v.t.	n.v.t.	n.v.t.
Subdomein: Met elkaar in verband brengen				
1F	... hebben paraat	3	1	0
	... gebruiken functioneel	3	1	0
	... weten waarom	3	0	1
2F	... hebben paraat	3	1	0
	... gebruiken functioneel	4	0	0
	... weten waarom	4	0	0
Subdomein: Gebruiken				
1F	... hebben paraat	3	0	1
	... gebruiken functioneel	4	0	0
	... weten waarom	n.v.t.	n.v.t.	n.v.t.
2F	... hebben paraat	4	0	0
	... gebruiken functioneel	4	0	0
	... weten waarom	2	1	1

Tabel 3.8 toont dat in het algemeen de overall dekkingsgraad van de vier rekendomeinen goed is: 73% tot 88% van de doelomschrijvingen per domein worden wel of deels gedekt door de lvs-aanbieders. Verder zien we dat de dekking iets beter is op de leerdoelen op 1F- dan op 2F-niveau.

Als we ons focussen op de ongedekte doelomschrijvingen (de kolom 'niet' in Tabel 3.8) per domein, dan krijgen we een beeld van de meest opvallende inhoudelijke hiaten voor de lvs'en rekenen-wiskunde van de vier aanbieders:

- Voor het eerste domein **getallen** zijn in totaal 18 doelomschrijvingen op 1F- en 2F-niveau beschreven en drie van deze subdoelen (17%) worden *niet* getoetst door ten minste een aanbieder. Het gaat dan om leerlingen weten waarom van de subdomeinen met elkaar in verband brengen (2F, bij 1:4 aanbieders niet) en gebruiken (1F, bij 2:4 niet), en leerlingen gebruiken functioneel van het subdomein gebruiken (2F, bij 1:4 niet).
- Voor het domein **verhoudingen** is de dekkingsgraad van de leerdoelen het hoogst. Er zijn 14 doelomschrijvingen voor verhoudingen geformuleerd op 1F- en 2F-niveau. Slechts twee (12%) daarvan komen *niet* terug bij in ieder geval één lvs-aanbieder. Het betreft hier het functioneel gebruiken bij het leerdoel met elkaar in verband brengen (2F, bij 1:4 aanbieders niet), en het weten waarom bij het leerdoel gebruiken (2F, bij 1:4 niet).

- Voor het domein **meten en meetkunde** is de dekkinggraad ook goed te noemen. Slechts drie van de 16 doelomschrijvingen (19%) op 1F- en 2F-niveau worden niet getoetst door ten minste één lvs-aanbieder. Deze drie subdoelen behoren tot het subdomein gebruiken en betreffen functioneel gebruiken (1F en 2F) en weten waarom (2F).
- Voor het laatste domein **verbanden** is de dekkinggraad van de doelomschrijvingen het laagst. Vier van de 15 leerdoelen (27%) op 1F- en 2F-niveau worden niet getoetst door ten minste één lvs-aanbieder. Het gaat drie keer om het leerdoel weten waarom bij de subdomeinen notatie en betekenis (1F), met elkaar in verband brengen (1F) en gebruiken (2F). Het vierde leerdoel dat niet getoetst wordt door één aanbieder is paraat hebben (1F) van het subdomein gebruiken.

3.4.6 Motivatie voor dekking van referentiekader door toetsaanbieders

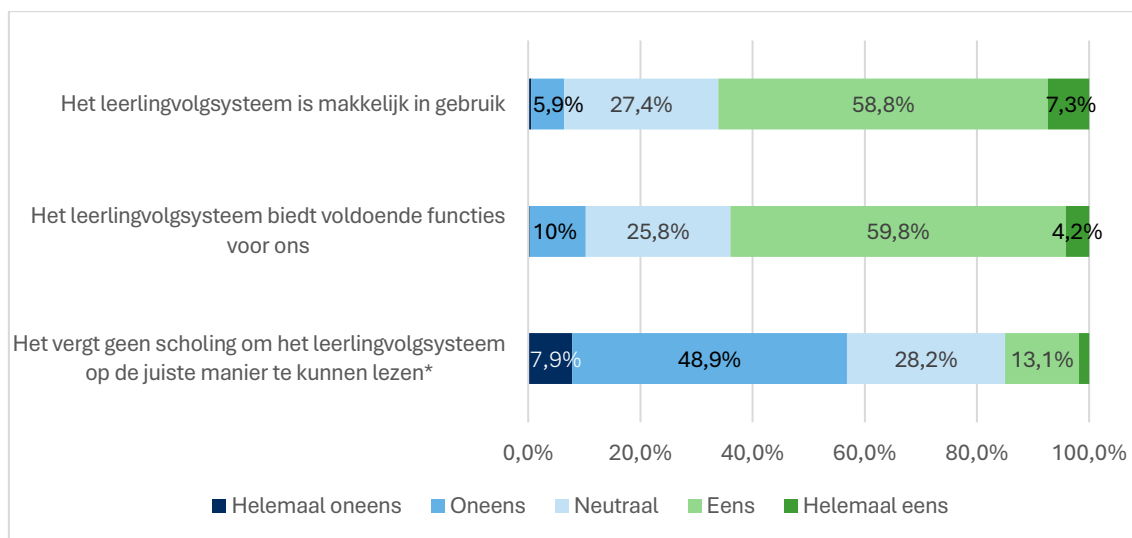
De referentiekaders voor taal en rekenen vormen voor de toetsaanbieders de basis voor de leerlingendoelen van volgtoetsen, omdat ze momenteel concreter en beter toepasbaar zijn en een duidelijker structuur bieden voor de rapportage van de resultaten per domein dan de brede en abstracte definitieve conceptkerndoelen (SLO, 2024). De huidige volgtoetsen rapporteren over het referentieniveau of functioneringsniveau van leerlingen en sluiten daarmee aan bij de wens van scholen. Scholen sturen op deze informatie, mede omdat de Inspectie van het Onderwijs gebruikmaakt van de referentieniveaus bij de beoordeling van scholen.

De toetsaanbieders hebben wel de intentie om met de definitieve conceptkerndoelen aan de slag te gaan door te onderzoeken wat er momenteel ontbreekt en hoe de structuur van de toetsing aangepast kan worden. Dit proces bevindt zich bij de meeste toetsaanbieders nog in een beginfase. De focus blijft voorlopig op de basisvaardigheden Nederlands, Engels en rekenen-wiskunde, terwijl de ontwikkelingen rondom de nieuwe basisvaardigheden nauwlettend worden gevolgd. Totdat er meer duidelijkheid is over de implementatie en nadere uitwerking van de definitieve conceptkerndoelen door SLO, blijven de referentiekaders leidend voor de ontwikkeling van lvs'en.

3.5 Knelpunten huidige gebruik

Met betrekking tot het gebruiksgemak komt in de vragenlijst naar voren dat het merendeel van de respondenten van mening is dat het scholing vergt om het lvs op de juiste manier te kunnen lezen, zie Figuur 3.4.

Figuur 3.4 In hoeverre zijn respondenten het eens met de volgende stellingen met betrekking tot het gebruiksgemak van het lvs? (n = 371)



*Oorspronkelijk stond het woord 'geen' niet in deze stelling. Omwille van de vergelijkbaarheid met de andere stellingen, hebben we deze stelling gehercodeerd (helemaal mee eens is nu het 'positieve' antwoord).

Uit de gesprekken komt een aantal knelpunten naar voren bij het gebruik van een lvs. Allereerst gaat dit praktisch om de organisatie van de inzet van een lvs op school, bijvoorbeeld vanwege onvoldoende capaciteit. Zo geven docenten van verschillende scholen aan dat de data uit een lvs zinvol kunnen zijn, bijvoorbeeld om het onderwijs aan te passen op basis van de resultaten. Er is niet altijd voldoende tijd en capaciteit om de resultaten uit te werken en aanpassingen door te voeren. Op verschillende scholen worden de resultaten uit een lvs wel ingezet op klasniveau of schoolniveau, maar niet op leerlingniveau, ook weer vanwege onvoldoende tijd en/of capaciteit. Er zijn te veel klassen en leerlingen om van elke leerling de progressie bij te houden en een persoonlijk traject uit te zetten, hoewel dat volgens respondenten wel waardevol zou zijn.

Daarnaast wordt een aantal knelpunten genoemd die betrekking hebben op het praktische gebruik van een lvs. Op een van de scholen wordt bijvoorbeeld aangegeven dat er geen richtlijnen zijn op de school over hoe de lvs-resultaten gebruikt moeten worden. Hierdoor gebruiken mentoren de resultaten op een verschillende manier, wat kansenongelijkheid in de begeleiding tot gevolg kan hebben. Ook wordt er tegen de AVG-regels aan gelopen. Een wiskundedocent geeft bijvoorbeeld aan te willen weten hoe leerlingen op leesvaardigheid presteren. In verband met AVG heeft de wiskundedocent geen toegang tot deze resultaten, hoewel dat wel waardevol zou zijn omdat leesvaardigheid in veel vakken van belang is. Namens de vakvereniging wiskunde⁵ geven respondenten aan dat thema's en vaardigheden bij wiskunde vaak los van elkaar staan, waardoor het moment van afname effect zal hebben op de toetsresultaten. Het kan voorkomen dat onderwerpen in een lvs-toets nog niet in de les aan bod zijn gekomen. Daarbij ligt de nadruk binnen het vmbo meer op rekenen, terwijl op havo/vwo algemene wiskundevaardigheden meer centraal staan. Hierdoor is het gebruik van een lvs toegankelijker voor het vmbo dan voor havo/vwo.

⁵ Nederlandse Vereniging van Wiskundeleraren

Tot slot is er een aantal knelpunten dat betrekking heeft op de leerlingen. Het kan lastig zijn om helder over het doel van de lvs-toetsen te communiceren naar leerlingen. Enerzijds kan onduidelijke communicatie toetsstress opleveren bij leerlingen. Een praktijkschool benadrukt dat de leerlingen op deze school zeker toetsstress krijgen wanneer ze het woord ‘toets’ horen. Leraren geven aan dat zij merken dat leerlingen de resultaten onderling vergelijken. De druk van deze toetsen wordt volgens hen niet zozeer veroorzaakt door een angst om te falen, maar om aanzien: ben ik beter of slechter dan klasgenoten? Anderzijds kan onduidelijke communicatie leiden tot niet-interpreteerbare resultaten omdat leerlingen de toetsen niet serieus genoeg maken. De toetsen ‘tellen niet mee’, waardoor leerlingen ze minder gefocust maken.

3.5.1 Survey: knelpunten bij interpretatie en implementatie van lvs-resultaten

In de survey met een vignet over problemen die vakdocenten kunnen hebben met het gevolg geven aan de lvs-resultaten hebben we aan 14 schoolleiders en vakdocenten gevraagd of deze problemen ook van toepassing zijn op hun school. De resultaten zijn gepresenteerd in Tabel 3.9.

Tabel 3.9 *Vormen van problemen met interpretatie en implementatie van lvs-resultaten door vakdocenten*

	Totaal (n=14)
Het beeld van veel leerlingen in de rapportages wijkt af van eigen waarneming	64%
De resultaten geven een wisselend beeld op deelaspecten van leerdomeinen	36%
Analyse en interpretatie van de resultaten in de rapportages zijn moeilijk	29%
Voorgestelde onderwijsaanpassingen zijn onvoldoende handelingsgericht	29%
De docenthandleiding bij de toetsen is lastig te begrijpen	0%

Alle respondenten stellen dat op hun school de docenten geen moeite hebben om de lvs-handleidingen goed te begrijpen; ze zijn heel duidelijk. Desondanks geeft bijna een derde van de respondenten (29%) aan dat het analyseren en interpreteren van de resultaten moeilijk is voor vakdocenten: het is vrij specialistisch en vraagt wel enige kennis van scoreverdelingen (A-E of I-V) en percentielen. Respondenten lichten toe dat de opgaven geheim zijn en het omslachtig is om bij de antwoorden van individuele leerlingen te komen, hetgeen de analyse en interpretatie van de lvs-resultaten door vakdocenten belemmert. Verder worden problemen met analyseren en interpreteren van lvs-resultaten ondervangen door docenten een opleiding te laten volgen.

Ruim een derde van de respondenten (36%) beaamt dat deelaspecten van leerdomeinen een wisselend beeld kunnen tonen. Gelet op gegeven toelichtingen is een wisselend beeld over de deelaspecten geen prangend probleem. Een zorgcoördinator bijvoorbeeld kan hierover uitleg geven aan de vakdocent. Het ligt anders voor het beeld dat voortkomt uit de rapportages in vergelijking met de eigen waarneming. Bijna twee derde van de respondenten (64%) stelt dat dit vaak verschillend is. Ze vertellen dat afwijkingen zich vooral voordoen bij leerlingen die lvs-toetsen niet serieus maken, maar soms zijn er hoge scores bij leerlingen die niet passen bij het beeld in de les. Een afwijking is aanleiding om te onderzoeken waar het vandaan komt.

Naast de in Tabel 3.9 genoemde beperkingen bij het analyseren en interpreteren van de lvs-resultaten brengt een respondent naar voren dat toegang tot het lvs-systeem op hun school beperkt is, mentoren kunnen er geen gebruik van maken.

Tot slot vindt bijna een derde van de respondenten (29%) dat de voorgestelde onderwijsaanpassingen uit de docenthandleiding onvoldoende handelingsgericht zijn. Het zijn globale adviezen, dat mag wel wat gerichter volgens een respondent. Opvallend is echter dat drie respondenten die de tegenovergestelde mening zijn toegedaan, als opmerking geven dat ze niets doen of weinig aandacht besteden aan de voorgestelde aanpassingen. Al met al ontstaat het beeld dat schoolleiders en vakdocenten in onze survey weinig doen met voorgestelde onderwijsaanpassingen uit docenthandleidingen.

In aanvulling op in het vignet genoemde beperkingen om met lvs-resultaten aan de slag te gaan (zie Tabel 3.9) stellen de respondenten dat lvs-resultaten op hun school weinig aandacht krijgen, omdat het tijdrovend is om er dieper op in te gaan. Het zijn heel veel gegevens waar je wat mee kunt, maar het volledig benutten vraagt een dusdanig grote tijdsinvestering die niet passend is binnen de taken die uitgevoerd moeten worden.

Toekomstig gebruik van lvs in het vo

4.1 Inleiding

Om het gebruik van een lvs optimaal in te kunnen zetten, moet het gedragen worden door de betrokken stakeholders. Dit geldt voor zowel het gebruiken van een lvs in het algemeen, als voor de vakken en vaardigheden waarvoor het lvs ingezet wordt. Om dit in kaart te brengen, is gesproken met verschillende stakeholders. Hierbij maken we een onderscheid tussen ‘externe’ stakeholders (beleidsmakers, inspecteurs, toetsontwikkelaars en belangenorganisaties; vanaf nu: stakeholders) en scholen. De ondervraagden hebben allemaal gesproken op persoonlijke titel, dus genoemde standpunten worden niet per definitie gedragen door de gehele organisatie. Of de vaardigheden geschikt zijn om op te nemen in een leerlingvolgsysteem hangt ook af van de psychometrische mogelijkheden en de haalbaarheid van het meten aan de hand van een lvs. Dit is in kaart gebracht met behulp van vakexperts en literatuuronderzoek. De informatie die hieronder uitgewerkt staat, geeft inzicht in welke inhoud geschikt is om te monitoren aan de hand van een leerlingvolgsysteem in het voortgezet onderwijs en kan dienen als input voor inhoudelijk criteria en voorwaarden die gesteld worden aan de landelijk erkende leerlingvolgsystemen.

4.2 Wenselijkheid verplichting lvs

Om de kansen en uitdagingen die gepaard gaan met een verplichtstelling van een lvs in de onderbouw van het vo vanuit verschillende kanten te belichten, is een aantal belangrijke stakeholders in deze context geïnterviewd. Daarnaast hebben we gesproken met een aantal scholen om ook vanuit hun perspectief de kansen en uitdagingen te belichten. Er werd wisselend gekeken naar de invoering van een verplicht lvs in de onderbouw van het vo door de verschillende stakeholders en scholen, zo bleek uit de interviews en de vignettenstudie. Waar de ene stakeholder het wenselijk acht deze verplichting door te voeren, zijn andere stakeholders kritisch op deze invoering. Wat in ieder geval duidelijk naar voren kwam, is dat er met een aantal belangrijke kansen en uitdagingen rekening gehouden moet worden alvorens een eventuele verplichtstelling door te voeren. In deze besprekingen kwamen verschillende thema's naar voren. Deze worden hieronder verder toegelicht.

4.2.1 Ontwikkeling volgen

Het doel van het lvs zou volgens alle stakeholders, scholen en leerlingen moeten zijn om objectief de ontwikkeling van leerlingen te volgen en daarop het onderwijs aan te passen. De meesten benoemen daarbij expliciet dat dat het enige doel mag zijn. Stakeholders gaven aan dat een lvs docenten kan voorzien van een objectieve tweede meting naast het eigen oordeel. Dit kan nuttig zijn voor alle docenten, maar specifiek voor beginnende docenten of docenten die minder ervaring hebben met het ontwikkelen van kwalitatief sterke toetsen. Juist in het vo, waar het zicht op de ontwikkeling van leerlingen door de vakkenstructuur meer versnipperd is dan in het po, kan het lvs van meerwaarde zijn om een consistent beeld te krijgen. Deze inzichten zouden dan vakoverstijgend bekeken moeten worden, zodat er binnen de verschillende vakken een samenhangende aanpak is om te werken aan de basisvaardigheden.

Het lvs biedt betrouwbare en gestandaardiseerde toetsen die de kwaliteit en consistentie van beoordelingen kunnen versterken. Hierbij wordt wel benadrukt dat de gegevens van het lvs puur formatief moeten worden ingezet. Het biedt dus diagnostische ondersteuning die inzicht geeft in waar leerlingen staan en wat ze nodig hebben om zich verder te ontwikkelen. Het feit dat gestandaardiseerde toetsen in een lvs objectief van aard zijn, zou ook kunnen bijdragen aan het verkleinen van onbewuste bias in de beeldvorming van voortgang van leerlingen. De meningen zijn verdeeld over het inzetten van een lvs als ondersteuning bij plaatsing, kwaliteitsbewaking, landelijke inzichten en oudercommunicatie. De algemene tendens bij de meerderheid van de respondenten luidt echter dat de gegevens uit een lvs daarvoor niet gebruikt moeten worden en puur een formatief karakter dienen te hebben. Een risico dat het inzetten van lvs-gegevens in bijvoorbeeld plaatsing met zich meebrengt, is dat leerlingen onderschat kunnen worden door eerdere lage prestaties. Wanneer leerlingen op basis van hun prestaties bijvoorbeeld alleen in homogene groepen geplaatst worden, kan dit het risico met zich meebrengen dat leerlingen met lagere scores minder kans hebben om te leren van hun beter presterende klasgenoten. Aan de andere kant noemt één stakeholder dat objectieve lvs-resultaten juist gebruikt kunnen worden voor het bevorderen van kansengelijkheid, als de toetsen bijvoorbeeld gebruikt worden om leerlingen goed te kunnen plaatsen in een vervolgstroom. Andere stakeholders geven echter aan dat een lvs niet gebruikt zou moeten worden voor plaatsingsdoeleinden, en alleen formatief ingezet zou moeten worden.

Daarnaast wordt de beperkte voorspellende waarde van lvs-resultaten benadrukt, zoals door het CPB onderzocht is in het primair onderwijs. Lvs-toetsen zijn ontworpen als momentopnamen die slechts een indicatie geven van het huidige prestatieniveau van de leerling. Hoewel de lvs-resultaten inzicht bieden in vaardigheden, zegt de score weinig over de toekomstige ontwikkeling en potentie van een leerling en zijn ze niet bedoeld om langdurige beslissingen op te baseren. Een te vroeg bepaalde positie van een leerling kan door docenten, ouders, en zelfs leerlingen zelf, onbedoeld als definitief worden opgevat. Dit kan leiden tot een selffulfilling prophecy. Daarnaast geven lvs-resultaten slechts inzicht in een beperkt aantal vaardigheden. De ontwikkeling van leerlingen is zoveel breder dan dat. Wanneer er te veel focus komt op de kennis en vaardigheden die het lvs toetst, kan het zijn dat leerlingen die, vanwege bijvoorbeeld een taalachterstand, minder goed presteren op de basisvaardigheden, minder de kans krijgen om laten zien waar hun kwaliteiten liggen. Dit benadrukt het belang van een puur formatief doel en het goed informeren van docenten over het juiste gebruik van de data.

In één van de gesprekken werd benadrukt dat de normering in een lvs vo zorgvuldig moet worden ingericht en dat de manier waarop leerlingprestaties in het lvs worden gepresenteerd aandacht vereist. In beide gevallen kan gekeken worden naar hoe hier in het lvs in het po rekening mee gehouden is. Wanneer de prestaties van een leerling in een lvs op een relatieve manier worden afgezet tegen de prestaties van (bijvoorbeeld) hun klasgenoten kan dit meerdere negatieve bijeffecten in de hand werken, zoals demotivatie en een toename in ervaren toetsdruk bij leerlingen, of het scheppen van een voortijdige verwachting bij docenten van het uiteindelijke prestatieniveau van een leerling. Ook kan de relatieve presentatie van leerlingprestaties de indruk wekken dat de beoordeling van hun prestatieniveau wordt beïnvloed door de relatieve prestaties van klasgenoten, terwijl dat niet het geval is.

Scholen geven aan dat de ontwikkeling van leerlingen uit lvs-resultaten soms niet overeenkomt met andere resultaten van de leerlingen, zoals methodetoetsen. Dat is verwarrend, en maakt het moeilijker om de lvs-resultaten te interpreteren en geschikte onderwijsaanpassingen te maken. Een verklaring voor deze discrepantie zou de timing van de lvs-toetsen kunnen zijn: soms komen onderwerpen in de lvs-toets aan bod die leerlingen nog niet in de les gehad hebben. Anderzijds geven docenten aan dat het wel goed is dat de lvs-toetsen niet methodegericht zijn, waardoor er niet alleen specifieke vraagtypen in zitten. Bij de inrichting van een verplicht systeem zou volgens de docenten dus rekening gehouden moeten worden met het moment van aanbod van bepaalde onderwerpen in de gebruikte methodes.

Respondenten zien ook een risico voor de professionaliteit en het beroepsbeeld van de docent. Het verplichten van een lvs kan een gebrek aan vertrouwen in de docent uitstralen. Namelijk dat docenten zelf niet (voldoende) in staat zijn om de ontwikkeling van de leerling te volgen. Ook de extra administratie die een lvs met zich meebrengt kan demotiverend werken voor de docent en het beroepsbeeld schaden. Verder is de angst van enkele respondenten dat ouders de resultaten van de lvs-toetsen zwaarder gaan wegen dan het professionele oordeel van de docent. Ze kunnen rechten gaan ontlenen aan de uitkomsten van de lvs-toetsen. Bijvoorbeeld als het gaat om het wel of niet overgaan van hun kind naar het volgende leerjaar. Scholen benoemen dat dit onwenselijk is.

4.2.2 Curriculum

Er zijn bij verschillende stakeholders zorgen dat een verplicht lvs zou kunnen leiden tot sturing van het curriculum. Hierbij zou de focus in het curriculum voornamelijk komen te liggen op datgene wat getoetst wordt in de gestandaardiseerde toetsen, ten koste van het andere minder makkelijk meetbare deel van het curriculum (teaching to the test). Het is zowel praktisch (qua hoeveelheid en lengte van toetsen) als inhoudelijk (qua meetbaarheid van onderwerpen) niet mogelijk om het gehele curriculum te toetsen in een leerlingvolgsysteem. Het risico dat de lvs-toetsen sturend werken voor het curriculum wordt versterkt als de lvs-toetsen als higher stakes gezien worden dan het andere onderwijsaanbod. Dit risico wordt groter als de lvs-resultaten op klas-, docent- of schoolniveau gebruikt gaan worden voor verantwoording. Bijvoorbeeld als de lvs-resultaten van een school vergeleken worden binnen een bestuur of als de inspectie ze gebruikt voor het bepalen van de onderwijskwaliteit. De druk die er daardoor op scholen wordt gelegd, heeft naar verwachting van de respondenten effect op de docent die die druk ook weer kan doorgeven aan de leerling.

Ook de ruimte voor scholen om hun eigen curriculum vorm te geven kan lijden onder het verplichte lvs. Het toetsen wordt 'een moetje', en niet meer gedaan vanuit intrinsieke motivatie van de school of docent. Door enkele respondenten wordt daarnaast nog genoemd dat door de verhoging van de toetsdruk er ook minder onderwijstijd overblijft, wat ook voor verschraving van het curriculum kan zorgen.

Een voordeel van een verplicht lvs voor het curriculum zou kunnen zijn dat het lvs vo-scholen meer bewust maakt van de kerndoelen en daaraan verbonden referentiekaders voor taal en rekenen-wiskunde. Volgens één stakeholder zijn scholen in het vo zich nog weinig bewust van de wettelijke doelen die op het gebied van taal en rekenen-wiskunde gelden. Daarnaast geven geïnterviewden van één school aan dat zij de informatie uit het lvs-systeem zeer nuttig vinden voor vakoverstijgende vaardigheden, bijvoorbeeld voor het inzetten van vakoverstijgend rekenen. Op deze manier zouden lvs-gegevens een hulpmiddel zijn voor een meer vakoverstijgend curriculum.

4.2.3 Toetsdruk

Een veel gehoorde zorg bij het verplichten van het lvs in de onderbouw van het vo is een toename van toetsdruk. Als scholen te veel toetsen afnemen om aan lvs-verplichtingen te voldoen, kan dit leiden tot een gevoel van druk bij zowel leerlingen als docenten. Het beeld van respondenten is dat lvs-toetsen niet in plaats van de al bestaande toetsen in het vo komen, maar er bovenop. Dat maakt dat er voor de leerlingen extra toetsen op het rooster komen. Hoewel er vaak niet voor deze toetsen geleerd hoeft te worden, zorgt het (naar verwachting) wel voor een hogere toetsdruk en daarbij stress voor de leerlingen. Voor docenten zorgen de extra toetsen ook voor extra druk, door zowel het inplannen ervan, tijd vrij maken in het lesrooster en de administratie die erbij komt kijken. Bovendien zouden scholen zich genooddaakt kunnen voelen om toetsresultaten centraal te stellen in hun onderwijsaanpak, waardoor andere doelen mogelijk minder aandacht krijgen. Wanneer dit doorslaat naar een eenzijdige nadruk op toetsen, gaat dit ten koste van het integrale beeld van de leerling in zijn geheel. Scholen zouden gebaat zijn bij duidelijke richtlijnen over de frequentie, de hoeveelheid en de doelen van toetsen, om te voorkomen dat er onnodig veel toetsen worden afgenomen. Geïnterviewden op scholen gaven aan dat het organiseren van de afname van lvs-toetsen ook organisatorisch veel druk op hen legt: het is soms lastig om de toetsen voor een hele jaarlaag op hetzelfde moment af te kunnen nemen, zeker als er voor veel verschillende vakgebieden toetsen afgenomen moeten worden. Om de toetsresultaten goed te kunnen interpreteren en vergelijken, willen scholen vaak graag dat de toetsen kort op elkaar afgenomen zijn (dus informatie geven over hetzelfde moment), en dat er tijd is voor een reflectiemoment met leerlingen over de toetsresultaten.

Aan de andere kant kan een goed lvs de noodzaak van aanvullende toetsen wellicht ook verminderen, volgens een van de stakeholders. Daarnaast hoeven leerlingen niet te leren voor de lvs-toetsen. Als dat ook duidelijk wordt gemaakt aan de leerlingen en scholen, zou het wellicht niet een grote impact op ervaren toetsdruk hebben. Scholen geven aan dat deze uitleg tijd kost: er zijn veel verschillen in het ervaren belang van de lvs-toetsen in het po. Omdat in het po het schooladvies deels gebaseerd wordt op lvs-toetsen (en de doorstroomtoets, die vaak van dezelfde aanbieder is), ervaren de leerlingen soms stress op basis van de vorige ervaringen. Aan de andere kant zien docenten en toetsaanbieders dat de motivatie van leerlingen voor de lvs-toetsen afneemt zodra ze merken dat de resultaten niet meetellen voor hun rapport. Hierdoor nemen leerlingen de lvs-toetsen minder serieus en zijn de uitkomsten dus ook minder betrouwbaar. Daardoor is het voor scholen lastig om een balans te vinden tussen de lvs-toetsen genoeg gewicht te geven dat ze serieus gemaakt worden en te veel toetsstress bij leerlingen. Uitleg over wat de school doet met de lvs-resultaten is daarin helpend. Daarnaast geeft een docent aan dat het moment van afname mogelijk ook invloed heeft op hoeveel motivatie en/of toetsstress leerlingen ervaren én op het maken van onderwijsloopbaanbeslissingen op basis van toetsresultaten. Aan het eind van het jaar zal bijvoorbeeld een lage score kunnen resulteren in een afstroomadvies, terwijl aan het begin van het jaar de vraag gesteld wordt: wat kunnen we hieraan doen?

4.2.4 Werkdruk docenten

Een verplicht lvs zou de administratieve last van docenten kunnen verhogen, voornamelijk als ze gevraagd worden om gegevens vast te leggen die ze op dit moment nog niet vastleggen. Het is daarom bij de inrichting van een verplicht lvs belangrijk om te bekijken welke huidige administratie door het invoeren van het lvs vervangen kan worden, zodat de totale administratieve last niet stijgt. Hierbij is de gebruiksvriendelijkheid van het systeem een belangrijke randvoorwaarde: het systeem moet zonder extra inspanning voor de docent te koppelen zijn aan andere systemen die op school gebruikt worden, en goed aansluiten bij de onderwijspraktijk. De vertaling van de resultaten in het lvs naar het handelen in de praktijk moet dus goed te maken kunnen zijn. Door de al bestaande werkdruk lukt het volgens docenten nu niet om de resultaten van de lvs-toetsen goed op leerlingniveau te analyseren en het onderwijs erop aan te passen.

Daarnaast hangt de hoeveelheid werkdruk van de docenten ook af van hoe gestandaardiseerd het nakijkproces is. Zo wordt leesvaardigheid over het algemeen automatisch nagekeken bij de digitale toetsen, maar bij de productieve vaardigheden, zoals spreken en schrijven, wordt meer van de docent gevraagd in het nakijkproces. Dit vraagt dus om een zorgvuldige afweging van het belang van het meten van bepaalde vaardigheden en de werkdruk die daarbij komt kijken.

4.2.5 Eigenaarschap en visie

Er wordt een aantal belangrijke kanttekeningen gemaakt bij een invoering van een verplicht lvs rondom eigenaarschap en de visie van de school. Zo zou een lvs recht moeten doen aan de visie en het profiel van een school en daarbij horende competenties. Een suggestie daarvoor van een stakeholder is dat er bijvoorbeeld eigen instrumenten opgenomen moeten kunnen worden in het lvs. Volgens scholen zou het gebruik van het lvs ook minder effectief worden als het verplichte lvs niet aansluit bij hun onderwijsvisie, omdat ze dan geen voor hen relevante conclusies uit de gegevens kunnen trekken. Een van de toetsaanbieders geeft ook aan dat er een flink aantal scholen is dat bewust geen lvs inzet in hun onderwijs. Zij zetten dan ook hun vraagtekens bij hoe het lvs wordt ingezet op deze scholen zodra het verplicht gesteld wordt, aangezien deze scholen zelf er de toegevoegde waarde niet van in zien. De vraag heerst dus of het relevant is een lvs in de onderbouw verplicht te stellen als vooral deze scholen een verandering moeten doorvoeren, ongeacht dat ze er geen behoefte aan hebben.

Vanwege de vakkenstructuur in het vo bestaat er een risico dat informatie uit een lvs bij één iemand terecht komt en dat er verder weinig mee gebeurt. Een verplichting zorgt nog niet automatisch voor het gevoel van eigenaarschap in de gehele school. Er moet nog wel actief iets gedaan worden met de informatie. Hierbij wordt benadrukt dat het lvs niet de rol of het inzicht van de leraar moet overnemen, want die ziet in de dagelijkse praktijk ook heel veel andere en aanvullende informatie. Het lvs zou niet als verantwoordingsinstrument voor de docent moeten worden gebruikt, maar als een van de vele informatiebronnen die een docent gebruikt. Geïnterviewden op scholen geven aan dat lvs-resultaten in de praktijk vaak verschillend gebruikt worden door docenten, omdat er geen duidelijke richtlijnen zijn over hoe de resultaten gebruikt zouden moeten worden. Een richtlijn voor de interpretatie en het gebruik van de resultaten zou de kansengelijkheid in de begeleiding volgens docenten kunnen bevorderen.

Daarnaast kan het volgens enkele stakeholders wenselijk zijn om leerlingen zelf inzicht te geven in hun resultaten, als deze besproken worden met de leerling. Ook docenten geven aan het als nuttig te ervaren om met leerlingen over de lvs-resultaten in gesprek te gaan, omdat zij dan beter begrijpen met welke reden sommige onderwijsaanpassingen gemaakt worden.

Hoewel een lvs in theorie kan bijdragen aan beter inzicht in de basisvaardigheden van leerlingen, merken sommige stakeholders op dat veel vo-scholen al werken met genormeerde toetsen voor taal en rekenen. Een wettelijke verplichting kan daarom als overbodig worden ervaren en mogelijk weerstand oproepen in het onderwijsveld, omdat er weinig aanvullende informatie zou worden opgehaald ten opzichte van de huidige situatie. Daarnaast geven veel stakeholders aan dat scholen al actief bezig zijn met basisvaardigheden, en het succes van een lvs hangt vooral af van hoe het wordt gebruikt, niet van het feit dat het verplicht is. Een verplicht lvs is alleen wenselijk als het daadwerkelijk voordelen voor de leerling oplevert en niet voor andere doeleinden wordt ingezet.

4.2.6 Marktwerving, kwaliteit en databeheer

Meerdere stakeholders gaven aan dat er een risico bestaat van marktwerving en prijsfactoren binnen het lvs-aanbod. Doordat veel lvs-pakketten bestaan uit meerdere vakken of vaardigheden, kiezen scholen soms voor een compleet pakket vanwege de kostenefficiëntie. Dit betekent echter niet altijd dat de inhoud van elk onderdeel optimaal aansluit op de behoeften of populatie van de school. Deze afhankelijkheid van prijs en marktwerving is volgens stakeholders onwenselijk, omdat scholen daarmee soms voor suboptimale opties kiezen. De voorkeur zou moeten uitgaan naar het beste instrument voor elk vakgebied, gebaseerd op onderwijskundige overwegingen in plaats van kostenbesparingen. Om de kwaliteit van de toetsen van verschillende lvs-aanbieders te waarborgen, benadrukken meerdere partijen het belang van een kwaliteitscontrole zoals in het po gedaan wordt door het College voor Toetsen en Examens (CvTE). Hierbij is het belangrijk deze partij vroegtijdig te betrekken in het proces.

Een ander genoemd risico dit betreffende gaat over de commerciële aanbieders. De marktwerving zorgt ervoor dat de aanbieders telkens willen blijven ontwikkelen en daarbij ook luisteren naar de wensen van de scholen. Hoewel dit zeker positieve kanten heeft, kan het ook zijn dat daardoor ontwikkelingen plaatsvinden die niet aansluiten bij het originele doel van het lvs. Bijvoorbeeld het beschikbaar maken van bepaalde overzichten op school- of bestuursniveau die vergelijking tussen scholen mogelijk maakt, terwijl dat niet aansluit bij het doel dat het lvs de focus moet leggen op de leerling.

Daarnaast wordt door meerdere stakeholders benoemd dat er bij een verplichting genoeg rekening gehouden moet worden met privacy en databeheer. In oplossingen voor databeheer schuilt een paradox. Grote aanbieders zullen heel grote hoeveelheden data opslaan, waardoor het risico op een groot datalek aanwezig is. Door het databeheer echter bij scholen te leggen, komt er veel verantwoordelijkheid bij hen te liggen, terwijl ze hier niet altijd de juiste expertise voor in huis hebben. Dit vergroot ook het risico op datalekken, hoewel de datalek dan kleiner zal zijn.

Tevens wordt door meerdere stakeholders aangegeven dat er goed nagedacht moet worden over wie toegang heeft tot welke informatie uit het lvs. Het is voor ouders bijvoorbeeld niet nodig om te zien hoe hun kind presteert ten opzichte van klasgenoten. Daarentegen geven docenten aan soms juist wel baat te hebben bij inzicht in gegevens buiten hun vakgebied (bijvoorbeeld een docent rekenen-wiskunde die inzicht wil in de leesvaardigheid van leerlingen i.v.m. rekenopgaven met veel tekst), maar dit vanwege privacyregels niet mogen krijgen. Dit beperkt het vakoverstijgend gebruik van de lvs-resultaten.

4.2.7 Survey: wenselijkheid verplichting lvs'en voor kernvakken

In de survey hebben we gevraagd aan 14 schoolleiders en vakdocenten of lvs'en Nederlands, Engels en rekenen-wiskunde verplicht moeten worden voor de onderbouw vo. Ruim een derde (36%) vindt van wel, de overige twee derde (64%) vindt van niet. Gezien het lage aantal respondenten, moeten we deze resultaten met voorzichtigheid interpreteren.

Diverse argumenten worden door de voorstanders genoemd. Ze kunnen een goede vergelijking met het landelijke beeld van het reguliere vo bieden, hetgeen speciaal van belang is in de Randstad waar taalachterstanden zo hard groeien. Ook past het bij de nieuwe standaard van de inspectie (OP0) dat als doel heeft de basisvaardigheden van leerlingen in het funderend onderwijs te verbeteren. Verder geven lvs'en inzicht in de resultaten van de school en bieden ze handvatten om met individuele leerlingen aan de slag te gaan. Een respondent is voorwaardelijk voorstander van lvs'en: het voortgezet onderwijs heeft al een taak om leerlingen voor te bereiden op het examenprogramma, maar de extra verantwoordelijkheid voor een passend aanbod voor taal en rekenen kan een uitdaging zijn.

De argumenten van de tegenstanders van een verplicht lvs zijn ook divers. Scholen moeten de vrijheid houden om naar eigen inzicht mee te doen. Als alle scholen mee moeten doen, dan gaan scholen trainen op de toets. Anderen vinden de kosten heel hoog, en de opbrengsten (te) gering. Scholen kunnen beter hun eigen leerlingvolgsysteem ontwikkelen of een andere manier dan met volgtoetsen aan de slag gaan om achterstanden van leerlingen weg te werken. Verder stellen tegenstanders dat lvs'en geen compleet beeld van leerlingen (kunnen) geven en niet valide zijn; ze meten vooral kennis en vaardigheden, nauwelijks inzicht.

4.2.8 Survey: nadelen en voordelen van gebruik lvs'en

Door middel van een vignet met een concrete probleemsituatie in de survey hebben we aan 16 schoolleiders en vakdocenten gevraagd of diverse vormen van weerstand tegen het gebruik van lvs'en ook gelden op hun school. De resultaten zijn gepresenteerd in Tabel 4.1.

Tabel 4.1 *Vormen van weerstand tegen het gebruik van lvs'en voor Nederlands, Engels en rekenen-wiskunde die van toepassing op hun school*

Lvs'en ...	Totaal (n =16)
geven een te beperkt en wisselend beeld van de leervorderingen	56%
worden door veel docenten gezien als afvinklijstjes	50%
worden door veel van onze leerlingen niet serieus ingevuld	50%
leveren te weinig op voor de inrichting van ons onderwijs	44%
zorgen voor een te hoge toetsdruk voor onze leerlingen en docenten	25%

Een kwart van de respondenten is van mening dat lvs'en voor een te hoge toetsdruk zorgen (Tabel 4.1). Als toelichting geven ze dat de afname te veel tijd kost, de toetsen duur zijn en ten koste gaan van de reguliere lessen en sommige groepen leerlingen in zijn algemeenheid niet gewend zijn aan toetsen. Driekwart is een andere mening toegedaan: lvs'en zorgen niet voor een te hoge toetsdruk. Omdat de toetsafname maar een à twee keer per schooljaar is, er geen leerwerk aan vooraf gaat, het niet meetelt voor een cijfer waardoor leerlingen het niet als toets ervaren, er sowieso weinig getoetst wordt op hun school.

De helft van de respondenten geeft van dat hun leerlingen een lvs-toets niet serieus maken omdat het niet voor een cijfer is. Dat geldt niet voor de brugklas, maar voor hogere leerjaren wel. Meerdere respondenten voegen daaraan toe dat het maken van de toets beter is geworden nadat duidelijk is uitgelegd wat het doel en belang voor de leerlingen is, en er extra ondersteuningslessen aan de toetsresultaten gekoppeld zijn.

Iets meer dan de helft van schoolleiders/vakdocenten (56%) stelt dat een lvs een te beperkt beeld geeft van de leerontwikkeling van hun leerlingen. Ze vinden dat een lvs-toets slechts een momentopname betreft die niet direct aansluit bij de leerstof en waarvan ze de vragen niet direct zelf kunnen inzien. Een respondent voegt daaraan toe dat de getoetste kennis en context niet past bij hun specifieke groep internationale leerlingen. De overige respondenten herkennen zich niet in de stelling en vinden dat het beeld van een leerling dat voortkomt uit een lvs vrijwel altijd wel overeenkomt met het beeld dat het team heeft, ondanks dat het een momentopname is en slechts één keer per jaar wordt afgenomen.

De meerderheid (56%) van de respondenten vindt dat lvs'en voldoende opleveren als het gaat om onderwijsontwikkeling. Eventuele leerachterstanden komen tijdig in beeld, waardoor bijsturing nog mogelijk is. Lvs'en geven een methode-neutrale momentopname, hetgeen handig is voor als leermateriaal zelf ontwikkeld wordt. Respondenten die vinden dat lvs'en onvoldoende de ontwikkeling van het onderwijs bevorderen stellen: dat alleen hun directie iets met de resultaten doet, dat een lvs niet goed aansluit bij hun onderwijsconcept of, te algemeen van aard is, en dat met name het lvs rekenen-wiskunde niet goed aansluit bij hun wiskunde-curriculum.

De helft van de respondenten beaamt dat veel docenten lvs'en vooral zien als afvinklijstjes omdat er toch niets mee wordt gedaan op hun school. Respondenten die het tegenovergestelde vinden stellen dat dit zeker niet geldt voor docenten in opleiding en als aan de lvs-resultaten concrete stappen worden verbonden.

4.3 Oneigenlijk gebruik

We vroegen respondenten naar wat zij zien als oneigenlijk gebruik van het lvs. Dat wil zeggen: de inzet van het lvs voor een doel waarvoor het lvs niet bedoeld is.

Onze respondenten zijn unaniem dat het lvs niet gebruikt moet gaan worden als verantwoordingsinstrument voor de kwaliteit van onderwijs door een toezichthouder, zowel intern (schoolbestuur) als extern (inspectie). Het lvs wordt gezien als een momentopname en maar één van de meetinstrumenten die scholen gebruiken om de ontwikkeling van een leerling of klas te volgen. Wanneer het lvs het zware gewicht krijgt dat scholen "erop afgerekend" worden, is de angst dat scholen een grote druk ervaren goed te presteren op de lvs-toetsen en – zoals hierboven beschreven – de kans is dat het onderwijs verschaalt.

Het belang van duidelijke verwachtingen en doelen van het lvs onderstrepen veel respondenten: waartoe dient het lvs, maar waartoe ook *niet*. Het risico dat door een aantal respondenten wordt gezien is dat data die er zijn, nou eenmaal gebruikt gaan worden. Ook voor doelen waar het lvs (initieel) niet voor bedoeld is. Mogelijk niet op de korte termijn, maar wel in de toekomst.

Ditzelfde kan gelden op leerlingniveau. De kans bestaat dat de uitkomsten van het lvs voor leerlingen summatief worden ingezet, bijvoorbeeld als het gaat om determinatie. We zien dat scholen nu al aangeven het lvs te gebruiken in de niveaubepaling van de leerling, vermoedelijk in combinatie met andere gegevens die een school verzamelt in de loop van het jaar. Afgezien van de vraag of het wenselijk is dat het lvs wordt ingezet als determinatie-instrument, is daarnaast een veel gestelde vraag of de toetsen zich daarvoor lenen. Scholen geven aan dat de toetsen vaak niet serieus worden gemaakt door de leerlingen. De toets 'telt niet mee voor een cijfer', sluit niet altijd goed aan bij het curriculum van de school (toetsstof is bijvoorbeeld nog niet of op een andere manier aangeboden) en blijft een momentopname waardoor de uitkomst de ontwikkeling van de leerling niet goed weergeeft. Verschillende respondenten vragen zich af of een al te zware weging van de lvs-toetsen – voor de leerling, maar ook voor de school – wel terecht is.

Voor de docenten zien respondenten het vergelijken van resultaten tussen docenten als oneigenlijk gebruik. Het risico is dat er frictie ontstaat omdat de ene docent hogere resultaten heeft behaald dan de andere docent. In de praktijk zien scholen dat nu nog niet ontstaan. Maar als het lvs verplicht is en er (mogelijk) meer gewicht aan de toetsen hangt, is dat wel een risico dat men ziet.

4.3.1 Survey: voorbeelden van oneigenlijk gebruik

In de survey hebben we gevraagd aan schoolleiders en vakdocenten of ze voorbeelden kennen van oneigenlijk gebruik. Negen van de vijftien respondenten (60%) kennen geen voorbeelden, de overige wel. Op overheid/stelsel-niveau worden voorbeelden gegeven die betrekking hebben op de Inspectie van het Onderwijs. Enkele respondenten noemen dat de inspectie een te algemeen beeld van de schoolpopulatie hanteert en dat lvs-resultaten op negatieve manier bij de beoordeling worden meegenomen terwijl de examenresultaten goed zijn.

Op schoolleiding-niveau geeft een respondent het voorbeeld dat vaksecties onder druk worden gezet, terwijl de examenresultaten uitstekend zijn en beter inzicht bieden. Een ander stelt dat lvs'en niet betrouwbaar het prestatieniveau kunnen meten en daarom niet zinvol zijn om als input te gebruiken voor het wegwerken van achterstanden.

Als voorbeeld van oneigenlijk gebruik door leerlingen of ouders geeft een respondent dat als een kind hoog scoort op een bepaald vakgebied, ouders wel eens willen vragen naar het prestatieniveau dat volgt uit een lvs.

4.4 Onderwijsinhouden toekomstig lvs

4.4.1 Wenselijkheid van het opnemen van specifieke vakgebieden

Wanneer een lvs in de onderbouw van het vo verplicht gesteld wordt, roept dit de vraag op hoe dit er dan uit zou moeten komen te zien. Een belangrijk vraagstuk daarbij is *wat* er gemeten zou moeten worden door het lvs. Een duidelijke tendens die naar voren kwam bij de stakeholders is dat een dergelijk lvs een zorgvuldige en doelgerichte opzet moet hebben waarbij vaardigheden worden gemeten die in het verdere leven van belang zijn, en dat schoolvak-overkoepelend is. Daarnaast wordt de noodzaak voor een beperkte inhoud benadrukt. Een suggestie hierbij is om de inhoud van het verplichte deel te beperken tot het minimum en eventuele andere vakken aan te bieden als optionele delen.

Een van de stakeholders suggereert om vaardigheden bij te houden die nog niet worden bijgehouden of waar moeilijk inzicht in te krijgen is (bijvoorbeeld door al afgenomen toetsen). Hierbij werden voorbeelden genoemd zoals ervaringsdoelen en inzicht in vakoverstijgende vaardigheden zoals kritisch/analytisch denken, digitale geletterdheid en burgerschap. De andere stakeholders waren het voornamelijk met elkaar eens over het includeren van Nederlands en rekenen-wiskunde. Deze vakken worden als kernvaardigheden gezien die voor alle andere schoolvakken ook van belang zijn. Tevens werd expliciet benoemd dat deze vakken volgens kleinschalig onderzoek het meest voorspellend zijn voor succes in het vo. Ook de mensen die werkzaam zijn op scholen beaamden dat de basisvaardigheden Nederlands en rekenen-wiskunde een waardevolle inhoud bieden voor een lvs. Daarbij rijst ook de vraag welke vaardigheden er vanuit deze vakgebieden moeten worden opgenomen. Bij Nederlands bijvoorbeeld wordt het momenteel vaak beperkt tot leesvaardigheid, maar daarbij dek je niet het volledige vakgebied van Nederlands. Hier moet een zorgvuldige afweging in gemaakt worden. Engels werd door een paar stakeholders ook als kernvak gezien.

Een deel van de stakeholders ziet voornamelijk risico's bij het opnemen van burgerschap in een verplicht lvs. In de survey kwam naar voren dat slechts 36% van de schoolleiders en docenten die de survey hebben ingevuld, vond dat burgerschap in het lvs zou moeten zitten; 64% was hierop tegen. Dit heeft te maken met de schooleigen invulling van het domein burgerschap en het risico op 'teaching to the test' en sociaal-wenselijke antwoorden. Een zorg die bij meerdere stakeholders ook leeft is dat burgerschap moeilijk te meten is. Ook de medewerkers op scholen gaven aan het niet wenselijk te vinden om burgerschap te meten, gezien de aard van het domein, sociaal-wenselijke antwoorden, het moeilijk meetbare karakter en omdat ze de informatie niet relevant genoeg vonden.

Digitale geletterdheid wordt door een minderheid van de stakeholders als relevant gezien, maar een van deze stakeholders benoemt ook dat er een gevaar schuilt dat het belangrijker gemaakt wordt dan het is. Op de scholen waren de meningen over het opnemen van digitale geletterdheid verdeeld, bleek uit de interviews. De meeste scholen gaven enerzijds aan dat het wenselijk is om digitale geletterdheid te meten, aangezien het een vakoverstijgende vaardigheid is. Anderzijds vermoedde een docent dat de digitale geletterdheid ook goed zal zijn als de andere vaardigheden goed zijn, omdat er volgens deze docent veel overlap is met andere vaardigheden. Weer een andere docent uitte zorgen over hoe meetbaar digitale geletterdheid is. Uit de survey bleek dat slechts 36% van de schoolleiders en docenten vonden dat digitale geletterdheid in een lvs zou moeten komen; 64% vond dit geen goed idee.

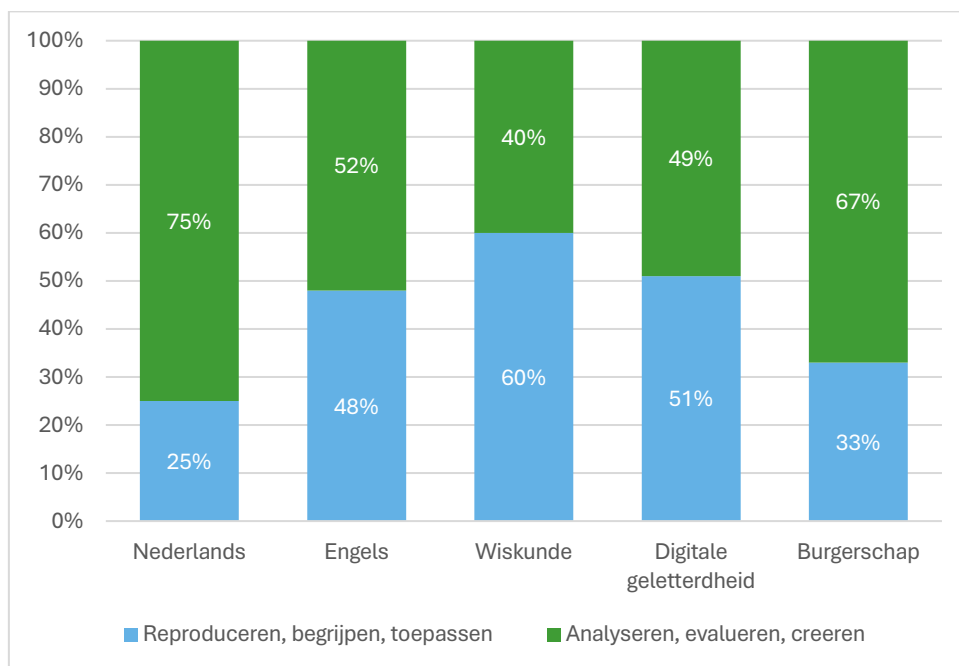
De stakeholders benoemden geen andere vakken of vaardigheden waarvan zij vonden dat die opgenomen moeten worden in een verplicht lvs. Meerdere docenten op de scholen waarmee we gesproken hebben, gaven wel aan dat sociaal-emotioneel functioneren en welzijn zinvol zijn om mee te nemen, omdat je alleen kunt leren als het goed met je gaat. Op een andere school had men interesse in het meten van executieve vaardigheden, ofschoon men ook aangaf dat dat niet voor iedere leerling relevant is en het voor extra toetsdruk kan zorgen.

4.4.2 Meetbaarheid van de onderwijsdoelen in een lvs

Naast een beeld van de vaardigheden die wenselijk zouden zijn om op te nemen in een lvs, is het voor een verplicht lvs ook van belang om te bekijken welke vaardigheden betrouwbaar en valide te meten zijn in volgtoetsen. Hiervoor is het van belang dat opgaven geconstrueerd kunnen worden die representatief zijn voor het curriculum en die in overeenstemming zijn met de beoogde leerdoelen. Bovendien moeten opgaven helder geformuleerd zijn en onderscheid kunnen maken tussen hoog- en laagvaardige leerlingen.

De basisvaardigheden taal en rekenen worden al lange tijd getoetst in volgsystemen, zowel in het po als het vo. Hiervoor geeft het referentiekader een beschrijving van te meten doelstellingen, waardoor de afstemming tussen toets en curriculum gerealiseerd kan worden. De validiteit en betrouwbaarheid van deze toetsen is door verschillende aanbieders van leerlingvolgsystemen uitgebreid onderzocht (zie bijv. Van Til & Van Boxtel, 2015) en deze vaardigheden zijn grotendeels goed meetbaar. Daarnaast is voor de peiling digitale geletterdheid in het primair onderwijs een valide instrumentarium ontwikkeld om digitale geletterdheid te meten (Heitink et al., 2024). Dit instrument is gebaseerd op directe meting van vaardigheden, waarbij leerlingen hun vaardigheden moeten toepassen in authentieke taken in verschillende toetsomgevingen. Deze toetsomgevingen waren gericht op handelen in een digitale omgeving en op het creëren van een digitaal product. Hiermee werden gegevens verzameld over vier domeinen van digitale geletterdheid: digitale informatievaardigheden, mediawijsheid en de aanverwante domeinen ICT-basisvaardigheden en computational thinking. Een uitgebreide analyse van de meetbaarheid van de door SLO opgestelde conceptkerndoelen voor Nederlands, Engels, wiskunde, burgerschap en digitale geletterdheid is te vinden in Bijlage 3.

Figuur 4.1 Cognitief proces onderliggende aan de leerdoelen per vakgebied



Slechts een kleine minderheid van alle leerdoelen uit de hier behandelde vakgebieden lijkt zich te lenen voor een vertaling naar makkelijk scorebare items, op basis van de indeling in cognitieve processen volgens de taxonomie van Bloom. Dit geldt voor slechts 93 van de 316 leerdoelen. Dit komt neer op 29%. Het aantal leerdoelen waarvoor dit zeker niet haalbaar lijkt, ligt een fractie hoger (100 leerdoelen, ofwel 32%). Voor de grootste groep (123, 39%) is dit wellicht niet bij voorbaat uitgesloten, maar lijkt het toch bijzonder moeilijk om dit te bewerkstelligen. Voor leerdoelen die betrekking hebben op reproducen, begrijpen en toepassen geldt veel vaker dat ze zich lenen voor operationalisering naar makkelijk te scoren toets-items. Dit geldt voor een meerderheid van de leerdoelen in deze categorie (76, 55%). Dit geldt voor slechts 17 leerdoelen (10%) die een beroep doen op analyseren, evalueren en creëren. Het totaal aantal leerdoelen in deze categorie (177) ligt ook iets hoger dan het aantal doelen uit de categorie reproducen, begrijpen en toepassen.

4.4.3 Burgerschap en digitale geletterdheid

Voor Engels, Nederlands en wiskunde is er sprake van een lange traditie van toetsing en examinering. Decennialang zijn voor deze vakgebieden gestandaardiseerde toetsen en examens ontwikkeld en afgenomen. Naast high-stakes tests gaat het daarbij ook om surveys, peilingen en leerlingvolgsystemen. In deze vakgebieden kan men bouwen op jarenlange ervaring wat betreft monitoring, toetsing en examinering. Dat ligt anders voor de betrekkelijk nieuwe vakgebieden burgerschap en digitale geletterdheid.

Uit de wetenschappelijke literatuur over de meetbaarheid van burgerschap blijkt dat instrumenten die burgerschap in kaart beogen te brengen veelal gebruik maken van attitudemetingen en zelfrapportages in aanvulling op traditionele meerkeuze-items.⁶ Daarnaast worden de mogelijkheden verkend van meer arbeidsintensieve en tijdrovende opties zoals portfolio's, vignetten en rubrics.⁷ Deze opties vragen niet alleen meer tijd en moeite van de beoordelaar, maar ook van de leerlingen. Vergeleken met gestandaardiseerde toetsitems en enquêtevragen leveren deze instrumenten een rijker en meer genuanceerd beeld op. Enige subjectieve interpretatie is bij de beoordeling met deze instrumenten echter onvermijdelijk.⁸

In leerlingvolgsystemen voor andere vakgebieden wordt weinig gebruikgemaakt van zelfrapportages of attitudemetingen. De door SLO geformuleerde kerndoelen voor burgerschap vermelden niets over (correcte) attitudes. Zelfrapportages zijn voor elk vakgebied voorstelbaar, maar hierbij ligt binnen burgerschap het gevaar van sociaal-wenselijke antwoorden en zelfonderschatting op de loer.

Uit wetenschappelijk onderzoek naar het meten van digitale geletterdheid blijkt dat veel toetsen voornamelijk bestaan uit meerkeuzevragen.⁹ Daarbij ligt de nadruk vooral op het meten van vaardigheden, terwijl kennis en attitudes minder aandacht krijgen.¹⁰ Dit verschilt van het meten van burgerschap, waar attitudemetingen en zelfrapportages veel worden gebruikt. Hoewel digitale geletterdheid kan worden opgevat als een verzameling van uiteenlopende vaardigheden en competenties, laten verschillende studies zien dat prestaties in verschillende digitale vaardigheidsgebieden sterk met elkaar samenhangen.¹¹ Daarnaast weten we vanuit peilingsonderzoek dat er een grote samenhang is tussen digitale geletterdheid en de vaardigheid in begrijpend lezen van leerlingen.¹²

4.4.4 Samenhang tussen vakgebieden

Uit onderzoeksliteratuur komt naar voren dat de resultaten op verschillende vakgebieden in zekere mate met elkaar samenhangen. Het meeste recente PISA-onderzoek laat bijvoorbeeld zien dat de scores voor leesvaardigheid en wiskunde bij de Nederlandse leerlingen op 15-jarige leeftijd zeer sterk samenhangen.¹³ De peiling digitale geletterdheid in het basisonderwijs laat een sterke samenhang zien tussen begrijpend lezen en digitale geletterdheid op 12-jarige leeftijd.¹⁴ Deze relaties achten we echter minder relevant voor het ontwerpen van een lvs. Ook al is er sprake van een sterke statistische samenhang tussen vaardigheden uit diverse vakgebieden, dat neemt de inhoudelijke verschillen niet weg. Bovendien blijven op individueel niveau grote verschillen altijd mogelijk. Een uitbinker in wiskunde presteert niet automatisch even goed bij andere vakken. In hoeverre bepaalde cognitieve processen onderliggend zijn aan de kerndoelen per vakgebied geeft een bruikbaar beeld van de overeenkomsten. De cognitieve processen waarop een beroep wordt gedaan bepalen in sterke mate of een leerdoel zich leent voor operationaliseerbaarheid naar makkelijk objectief scorebare toets-items.

⁶ Ten Dam et al., 2010

⁷ Dijkstra et al., 2023

⁸ Daas et al., 2016

⁹ Siddiq et al., 2016

¹⁰ Godaert et al., 2022

¹¹ Reichert et al., 2023

¹² Heitink et al., 2023

¹³ OECD, 2024

¹⁴ Heitink et al., 2023

Tabel 4.2 laat per vakgebied het verband zien tussen het cognitief proces van de kerndoelen en de operationaliseerbaarheid. Voor alle vakgebieden geldt dat doelen die een beroep doen op lagere orde vaardigheden zich beter lenen voor operationalisering naar lvs-items. De resultaten kunnen als volgt worden samengevat:

- Voor de vakgebieden wiskunde, digitale geletterdheid en burgerschap is deze samenhang sterker dan voor Nederlands en Engels.
- Over het algemeen genomen (Nederlands, rekenen-wiskunde, Engels, burgerschap en digitale geletterdheid) geldt dat 55% van de leerdoelen die een beroep doen op reproduceren, begrijpen of toepassen (RBT) waarschijnlijk zonder grote problemen geoperationaliseerd kan worden.
- Wanneer we dit per vak uitsplitsen (zie Tabel 4.2) zien we dat dit varieert van 25,9% voor Engels tot 75% voor burgerschap.
- Voor 27% van de RBT-leerdoelen lijkt het nog wel mogelijk om ze te operationaliseren, maar is het wel erg moeilijk. Dit varieert van 0% bij burgerschap tot 40,7% bij Engels.
- Voor 18% van de RBT-leerdoelen lijkt het niet realistisch om ze te operationaliseren, variërend van 3,8% bij digitale geletterdheid tot 45,8% bij Nederlands.
- Voor de leerdoelen die een beroep doen op hogere orde vaardigheden (analyseren, evalueren en creëren) zijn deze percentages respectievelijk 10% (variërend van 3,5% voor Engels tot 16% voor digitale geletterdheid), 48% (variërend van 32% voor digitale geletterdheid tot 66,7% voor wiskunde) en 42% (variërend van 3,7% voor wiskunde tot 50% voor burgerschap).

Tabel 4.2 Cognitief proces leerdoelen en operationaliseerbaarheid per vak

Vakgebied	Cognitief proces	Mogelijk zonder grote problemen	Mogelijk, maar erg moeilijk	Niet realistisch
Nederlands (95 leerdoelen)	Reproducen, Begrijpen, Toepassen	33,3%	20,8%	45,8%
	Analyseren, Evalueren, Creëren	9,9%	43,7%	46,5%
Engels (56 leerdoelen)	Reproducen, Begrijpen, Toepassen	25,9%	40,7%	33,3%
	Analyseren, Evalueren, Creëren	3,4%	51,7%	44,8%
Wiskunde (90 leerdoelen)	Reproducen, Begrijpen, Toepassen	66,7%	29,6%	3,7%
	Analyseren, Evalueren, Creëren	11,1%	66,7%	22,2%
Digitale geletterdheid (51 leerdoelen)	Reproducen, Begrijpen, Toepassen	73,1%	23,1%	3,8%
	Analyseren, Evalueren, Creëren	16,0%	32,0%	52,0%
Burgerschap (24 leerdoelen)	Reproducen, Begrijpen, Toepassen	75,0%	0,0%	25,0%
	Analyseren, Evalueren, Creëren	6,3%	43,8%	50,0%

Ter illustratie wordt in Tabel 4.3 voor tien kerndoelen (twee per vakgebied) weergegeven op welk cognitief proces ze betrekking hebben en hoe hun operationaliseerbaarheid is beoordeeld.

Uit een inventarisatie van de huidige lvs'en volgt dat de dekkinggraad van de definitieve conceptkerndoelen voor Nederlands 48%, voor Engels 55% en voor rekenen-wiskunde 92% is (zie 3.4.1, 3.4.2, 3.4.3). Een kerndoel wordt als gedekt beschouwd als ten minste één aanbieder heeft aangegeven dat een bepaald kerndoel deel uitmaakt van hun toets. Over de kerndoelen die door geen enkele aanbieder gedekt worden, zoals 'de leerling verkent hoe je met taal uiting geeft aan je identiteit' (kerndoel 14) of 'de leerling verkent de waarde van literatuur' (kerndoel 18) bij het lvs-Nederlands, stellen de toetsaanbieders dat deze niet of moeilijk digitaal toetsbaar zijn, niet te bevragen zijn met gesloten vragen, of dat zij geen literatuurtoetsen aanbieden. Sommige aanbieders zeggen wel de mogelijkheden te verkennen: misschien zijn vormen van 'de leerling verkent' in de toekomst toetsbaar met nieuw te ontwikkelen beoordelingsformulieren voor gespreksvoering.

Tabel 4.3 Voorbeelden van leerdoelen inclusief cognitief proces en operationaliseerbaarheid

Vakgebied	Leerdoel	Cognitief Proces	Operationa- liseerbaar- heid
Nederlands	De leerling evalueert en reflecteert op zakelijke en literaire teksten. Bijvoorbeeld benoemen van inhoudelijke relaties binnen en tussen verschillende teksten	Begrijpen	Mogelijk, zonder grote problemen
	De leerling gebruikt taal op een creatieve manier. Bijvoorbeeld verwoorden van eigen gedachten, ervaringen, gevoelens en fantasieën.	Creëren	Niet realistisch
Engels	De leerling spreekt en voert gesprekken in de Engelse taal, afgestemd op doel, publiek en context. Bijvoorbeeld gebruiken van passende spreek- en gespreksbevorderende strategieën.	Toepassen	Niet realistisch
	De leerling stemt de communicatie af op de culturele en talige achtergrond van de ontvanger. Bijvoorbeeld interpreteren van culturele interacties die in het Engels plaatsvinden	Analyseren	Mogelijk, maar erg moeilijk
Wiskunde	De leerling analyseert en redeneert over patronen en verbanden. Bijvoorbeeld herkennen, beschrijven en voortzetten van patronen in rijen getallen en figuren	Begrijpen	Mogelijk, zonder grote problemen
	De leerling bedenkt en beschrijft algoritmes. Bijvoorbeeld beoordelen van het resultaat van een doorlopen algoritme	Evalueren	Mogelijk, maar erg moeilijk

Vakgebied	Leerdoel	Cognitief Proces	Operationa- liseerbaar- heid
Digitale geletterdheid	De leerling navigeert doelgericht in het digitale media- en informatielandschap voor het verwerven en verwerken van informatie. Bijvoorbeeld beoordelen van aangeboden en gevonden informatie op betrouwbaarheid en bruikbaarheid, rekening houdend met eigenschappen van bronnen, zoekhulpmiddelen en gebruikte media	Evalueren	Mogelijk, zonder grote problemen
	De leerling analyseert hoe digitale technologie, digitale media en de samenleving elkaar wederzijds beïnvloeden en verkent toekomstscenario's. Bijvoorbeeld analyseren hoe samenlevingen afhankelijk zijn van digitale technologie en van grote technologiebedrijven	Analyseren	Niet realistisch
Burgerschap	De leerling geeft aan hoe diens handelen verbonden is met democratische waarden. Bijvoorbeeld kennis over het belang van de grondwet, kinder- en mensenrechten en internationale verdragen ter bescherming van vrijheid, gelijkheid en solidariteit	Reproducen	Mogelijk, zonder grote problemen
	De leerling weegt af welke mogelijkheden die heeft om ten aanzien van maatschappelijke vraagstukken te handelen. Bijvoorbeeld analyseren welke belangen en perspectieven een rol spelen binnen actuele, maatschappelijke en planetaire vraagstukken	Creëren	Niet realistisch

4.5 Koppeling met het lvs po

In deze paragraaf bespreken we de mogelijkheden en wenselijkheid om resultaten van de lvs-toetsen in het vo te koppelen aan de resultaten die leerlingen eerder behaald hebben bij lvs-toetsen in het po. Met een koppeling kan er een completer beeld van leerlingen tot stand komen en kunnen bepaalde patronen mogelijk sneller inzichtelijk worden. De vraag is echter of en hoe dit gerealiseerd kan worden, en of dit als wenselijk wordt gezien door verschillende stakeholders.

Op dit moment is er geen directe koppeling tussen leerlingvolgsystemen in het po en het vo, althans niet zodanig dat gegevens uit het po direct vergelijkbaar zijn met gegevens uit het vo. Er worden in het overstapdossier wel gegevens van lvs-toetsen uit het po overgedragen aan de nieuwe vo-school van een leerling. Voor het overstapdossier zijn afspraken ontwikkeld in het kader van de Overstap Service Onderwijs (OSO). OSO is het geheel van afspraken, standaarden en beveiligde verbindingen voor veilige digitale overdracht van studie- en begeleidingsgegevens van een leerling vanuit de leerlingadministratie van de huidige school naar de nieuwe school.¹⁵ Bij de overgang van po naar vo is in het overstapdossier ook ruimte voor overdracht van gegevens van lvs-toetsen. Welke toetsen dit voor OSO zijn, wordt gepubliceerd in de OSO-toetscodelijst. Voor de toetsen in deze lijst is vastgesteld welke afnamescores (ruwe scores) verplicht en optioneel zijn en welke referentiescores verplicht en optioneel zijn. Verplichte scores moeten in het resultaat worden meegegeven en er mogen geen andere scores dan de verplichte en optionele scores worden meegegeven. Welke scores precies in het dossier staan, verschilt voor verschillende aanbieders van leerlingvolgsystemen in het po (bijv. percentielscores of didactische leeftijdsequivalent). Omdat leerlingen in het vo van verschillende po-scholen komen, is hierdoor niet voor alle leerlingen binnen een klas of school dezelfde informatie beschikbaar. Bovendien ontbreekt er soms informatie, omdat scholen niet verplicht zijn om alles in te vullen. Naast de verschillen tussen aanbieders, zijn er ook verschillen in rapportagemethoden in het po en vo binnen aanbieders. Zo zijn bepaalde gegevens die in het po gerapporteerd worden lang niet altijd bekend en begrijpelijk voor stakeholders in het vo (bijv. de niveaus A t/m E van het po-volgsysteem Leerling in beeld die aangeven hoe leerlingen scoren t.o.v. het landelijk gemiddelde komen niet voor in het vo-volgsysteem Leerling in beeld). Deze verschillen ervoor dat gegevens uit het po vaak moeilijk te interpreteren zijn door stakeholders in het vo.

In dit hoofdstuk gaan we in op de mogelijkheden voor een koppeling tussen lvs-toetsen in het po en vo. Thema's die in dit hoofdstuk behandeld worden, gaan achtereenvolgens over de inhoudelijke aansluiting van lvs-toetsen in het po en vo (paragraaf 4.5.1), de mogelijkheden om toetsresultaten (psychometrisch) te koppelen tussen het po en vo (4.5.2) en de wenselijkheid van een koppeling (4.5.3). We sluiten af met een beschrijving van belangrijke aandachtspunten (4.5.4).

4.5.1 Aansluiting onderwijsinhoud volgtoetsen po en vo

Om de resultaten van lvs-toetsen in het po en vo aan elkaar te koppelen, is het van belang dat de toetsen inhoudelijk vergelijkbaar zijn. Er is overlap in de vakgebieden die getoetst worden in po en vo: Nederlands leesvaardigheid, Nederlands taalverzorging en rekenen-wiskunde worden in zowel het po als het vo gemeten. Deze vakgebieden zijn onderdeel van het 'Referentiekader doorlopende leerlijnen taal en rekenen' van de SLO (Expertgroep doorlopende leerlijnen taal en rekenen, 2009). Hierin is voor het hele onderwijs (van de basisschool tot het hoger onderwijs) vastgelegd wat leerlingen moeten kennen en kunnen als het gaat om Nederlandse taal en rekenen-wiskunde. In die zin omvatten lvs-toetsen in het po en vo dus overkoepelend (deels) dezelfde onderwijsinhouden en zouden leerlingresultaten op deze onderdelen uit het vo vergeleken kunnen worden met de resultaten uit het po.¹⁶

¹⁵ https://www.edustandaard.nl/standaard_afspraken/oso-gegevensset-en-profielen-oso/oso-gegevensset-en-profielen-2024-1/

¹⁶ Dit geldt niet voor alle vakgebieden; Engels wordt bijvoorbeeld niet altijd getoetst in het po.

Echter is het niet zo dat de constructen die gemeten worden precies hetzelfde blijven over tijd. (Deel)constructen kunnen veranderen, bijvoorbeeld wanneer er nieuwe onderwerpen aan bod komen die getoetst worden. Bovendien geven verschillende lvs-aanbieders op een verschillende manier vorm aan het meten van de constructen. Bij het maken van een koppeling tussen po en vo is het belangrijk dat er rekening gehouden wordt met deze (mogelijke) verschillen in onderwijsinhoud en dat resultaten op een legitieme manier met elkaar vergeleken worden.

4.5.2 (Psychometrische) mogelijkheden voor een koppeling

Om een psychometrische koppeling te realiseren, kunnen ofwel vaardigheidsschalen worden doorgetrokken van po naar vo, ofwel referentieniveaus gebruikt worden. In deze sectie lichten we beide manieren toe en beschrijven we de voor- en nadelen hiervan. Vervolgens beschrijven we enkele alternatieve ideeën om een koppeling te maken.

Koppeling via vaardigheidsschalen

Waarom vaardigheidsschalen?

Het gebruik van een vaardigheidsschaal is zeker in het po wijdverbreid gebruik. Een vaardigheidsschaal is vergelijkbaar met een meetlat, die aangeeft hoe goed een leerling een bepaalde vaardigheid beheerst. De vaardigheidsschaal maakt het mogelijk om “28 goed” op de ene toets voor een bepaalde vaardigheid te vergelijken met “31 goed” op een andere toets voor diezelfde vaardigheid. Daardoor kun je resultaten van een leerling op verschillende toetsmomenten met elkaar vergelijken, plus de resultaten van leerlingen in dezelfde groep die verschillende toetsen hebben gemaakt.¹⁷ Zeker binnen volgsystemen biedt het gebruik van een vaardigheidsschaal veel voordelen.

Bij het gebruik van een vaardigheidsschaal kunnen drie verschillende vormen van normeringen¹⁸ gebruikt worden. Dat zijn achtereenvolgens het vergelijken van de leerling met een eerdere prestatie van zichzelf (groeigerichte normering), de vergelijking van de prestatie van een leerling met een inhoudelijk criterium (absolute normering) en de vergelijking van de prestatie van een leerling met andere (groepen) leerlingen (relatieve normering).

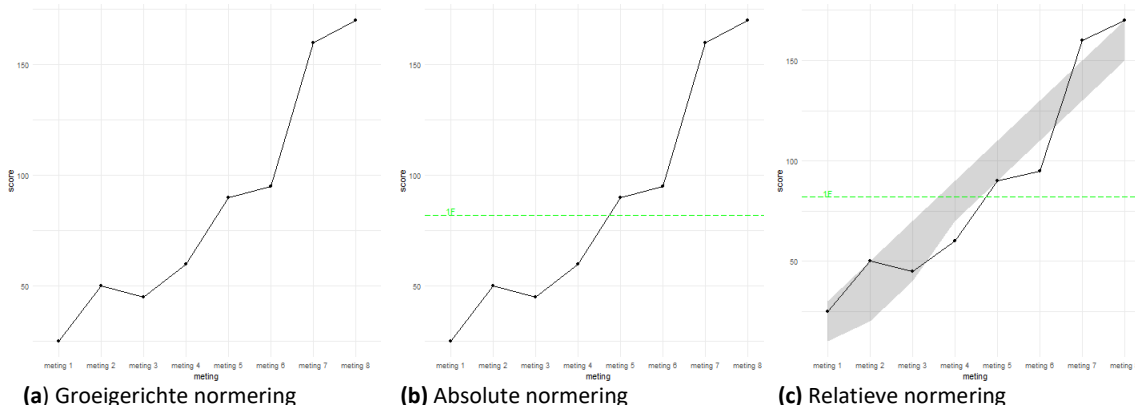
De groei van leerlingen kan met behulp van een vaardigheidsschaal over de tijd heen gevolgd worden. Een voorbeeld is weergegeven in Figuur 4.2a. Daarnaast kan een bepaalde inhoudelijke standaard overgebracht worden op de vaardigheidsschaal. Een voorbeeld hiervan is het gebruik van referentieniveaus binnen een volgsysteem. Voor het behalen van een bepaald referentieniveau zijn prestatiestandaarden vastgesteld. Deze prestatiestandaarden kunnen op hun beurt weer omgezet worden naar de gebruikte vaardigheidsschaal in een volgsysteem. Hiermee kan ook binnen het volgsysteem gerapporteerd worden of een leerling een bepaald inhoudelijke standaard of criterium heeft behaald. Een voorbeeld hiervan is weergegeven in Figuur 4.2b. Deze figuur laat met behulp van de groene horizontale lijn zien welke vaardigheidsscore (of welk prestatieniveau) correspondeert met het aantonen van het beheersen van minimaal niveau 1F (fictief voorbeeld). Leerkrachten en leerlingen kunnen dus ook zien dat hier bijvoorbeeld geldt dat de leerling bij meting 4 nog behoorlijk verwijderd is van het bereiken van het niveau 1F, maar dat dit niveau bij meting 5 wel bereikt is.

Ten slotte kan met behulp van vaardigheidsschaal inzichtelijk worden gemaakt hoe een leerling presteert ten opzichte van andere leerlingen of een groep andere leerlingen (een normpopulatie). Een voorbeeld hiervan is te zien in Figuur 4.2c.

¹⁷ <https://cito.nl/media/sz4junrc/cito-flyer-toetsscore-vaardigheidsscore-en-dan.pdf>

¹⁸ Normering heeft betrekking op het betekenis geven aan toetsscores.

Figuur 4.2 Drie verschillende vormen van normeringen



Een bijkomend voordeel bij een vaardigheidsschaal (en in het verlengde van het kunnen gebruiken van absolute criteria), is dat de opgaven in de toetsen op dezelfde schaal liggen als de vaardigheid. Dit geeft de leerkrachten en leerlingen informatie over welke opgaven zij waarschijnlijk wel beheersen en welke opgaven nog niet. Deze vorm van rapportages wordt onder andere vaak gebruikt bij (internationaal) peilingsonderzoek. Een voorbeeld hiervan is te zien in Figuur 4.3a. In deze figuur geeft een voorbeeld uit PISA OECD (2023) weer hoe de inhoudelijke kenmerken van de taken via de moeilijkheid van de opgaven aan de vaardigheid van leerlingen gekoppeld kunnen worden.

Een ander voorbeeld waarin de inhoud van de opgaven gebruikt wordt om betekenis te geven aan de vaardigheidsschaal is te vinden in Figuur 4.3b. De schaal krijgt inhoudelijk waarde door diverse opgaven op deze schaal af te beelden.¹⁹ Het niveau van de opgave geeft aan bij welke vaardigheid de helft van de leerlingen deze vraag goed hebben en de helft van de leerlingen een fout antwoord geven. Bij een lagere vaardigheid hebben meer leerlingen de opgave fout, en bij een hogere vaardigheid hebben meer leerlingen de opgave goed beantwoord. Er is gekozen voor dit punt om een opgave af te beelden, omdat dit het punt is waar de opgave het best onderscheid maakt tussen de leerlingen rond dat niveau. Inhoudelijk is dit het meest interessante punt omdat dit het beste aangeeft waar de opgaven uitdagend zijn: ze zijn niet te moeilijk en ze zijn niet te makkelijk.

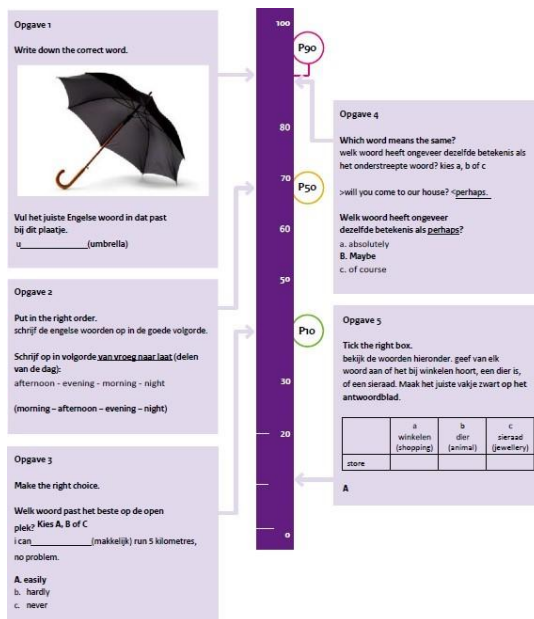
¹⁹ De schaal woordenschat is hier gegeven als het percentage goede antwoorden op de toets gegeven de vaardigheid van de leerlingen. Meer over deze schaal in de technische rapportage van de peiling Engels 2017-2018: Ritzema, L., Hemker, B., Naayer, H., Lowie, W., Corda, A., Van Aken, M. & Rekers-Mombarg, L. (2018). Peiling Engels einde basisonderwijs, 2018, Technische Rapportage. <https://www.onderwijsinspectie.nl/onderwerpen/peil-onderwijs/documenten/rapporten/2019/11/08/peiling-engels-einde-basisonderwijs-2018-technische-rapportage>.

Figuur 4.3 Opgaven en vaardigheid op dezelfde schaal

Table I.3.1. Description of the eight levels of mathematics proficiency in PISA 2022

Level	Lower score limit	Percentage of students able to perform tasks at each level or above (OECD average)	Characteristics of tasks
6	669	2.0%	At Level 6, students can work through abstract problems and demonstrate creativity and flexible thinking to develop solutions. For example, they can recognise when a procedure that is not specified in a task can be applied in a non-standard context or when demonstrating a deeper understanding of a mathematical concept is necessary as part of a justification. They can link different information sources and representations, including effectively using simulations or spreadsheets as part of their solution. Students at this level are capable of critical thinking and have a mastery of symbolic and formal mathematical operations and relationships that they use to clearly communicate their reasoning. They can reflect on the appropriateness of their actions with respect to their solution and the original situation.
5	607	8.7%	At Level 5, students can develop and work with models for complex situations, identifying or imposing constraints, and specifying assumptions. They can apply systematic, well-planned problem-solving strategies for dealing with more challenging tasks, such as deciding how to develop an experiment, designing an optimal procedure, or working with more complex visualisations that are not given in the task. Students demonstrate an increased ability to solve problems whose solutions often require incorporating mathematical knowledge that is not explicitly stated in the task. Students at this level reflect on their work and consider mathematical results with respect to the real-world context.
4	545	23.6%	At Level 4, students can work with explicit models for complex concrete situations, sometimes involving two variables, as well as demonstrate an ability to work with undefined models that they derive using a more sophisticated computational-thinking approach. Students at this level begin to engage with aspects of critical thinking, such as evaluating the reasonableness of a result by making qualitative judgements when computations are not possible from the given information. They can select and integrate different representations of information, including symbolic or graphical, linking them directly to aspects of real-world situations. At this level, students can also construct and communicate explanations and arguments based on their interpretations, reasoning, and methodology.
3	482	45.6%	At Level 3, students can devise solution strategies, including strategies that require sequential decision-making or flexibility in understanding of familiar concepts. At this level, students begin using computational-thinking skills to develop their solution strategy. They are able to solve tasks that require performing several different but routine calculations that are not all clearly defined in the problem statement. They can use spatial visualisation as part of a solution strategy or determine how to use a simulation to gather data appropriate for the task. Students at this level can interpret and use representations based on different information sources and reason directly from them, including conditional decision-making using a two-way table. They typically show some ability to handle percentages, fractions and decimal numbers, and to work with proportional relationships.
2	420	68.9%	At Level 2, students can recognise situations where they need to design simple strategies to solve problems, including running straightforward simulations involving one variable as part of their solution strategy. They can extract relevant information from one or more sources that use slightly more complex modes of representation, such as two-way tables, charts, or two-dimensional representations of three-dimensional objects. Students at this level demonstrate a basic understanding of functional relationships and can solve problems involving simple ratios. They are capable of making these interpretations of results.
1a	358	87.6%	At Level 1a, students can answer questions involving simple contexts where all information needed is present, and the questions are clearly defined. Information may be presented in a variety of simple formats and students may need to work with two sources simultaneously to extract relevant information. They are able to carry out simple routine procedures according to direct instructions in explicit situations, which may sometimes require multiple iterations of a routine procedure to solve a problem. They can perform actions that are obvious or that require very minimal synthesis of information, but in all instances the actions follow clearly from the given stimuli. Students at this level can employ basic algorithms, formulae, procedures, or conventions to solve problems that most often involve whole numbers.
1b	295	97.4%	At Level 1b, students can respond to questions involving easy to understand contexts where all information needed is clearly given in a simple representation (i.e. tabular or graphic) and as necessary, recognise when some information is extraneous and can be ignored with respect to the specific question being asked. They are able to perform simple calculations with whole numbers, which follow from clearly prescribed instructions, defined in short, syntactically simple text.
1c	233	99.7%	At Level 1c, students can respond to questions involving easy to understand contexts where all relevant information is clearly given in a simple, familiar format (for example, a small table or picture) and defined in a very short, syntactically simple text. They are able to follow a clear instruction describing a single step or operation.

Source: OECD, PISA 2022 Database, Table I.B1.3.1.



(a) Taakbeschrijving opgaven gerelateerd aan de vaardigheid van leerlingen

(b) Voorbeeld van representatie van opgaven op een vaardigheidsschaal (Engels woordenschat, groep 8)

Vaardigheidsschalen doortrekken van po naar vo

Vaardigheidsschalen lenen zich goed voor het volgen van leerlingen met een lvs, omdat met vaardigheidsscores de groei van leerlingen op de vaardigheidsschaal bepaald kan worden. Voor lvs-toetsen geldt dat de resultaten van verschillende toetsen binnen een vakgebied, aanbieder en schoolsoort (po/vo) met elkaar te vergelijken zijn met behulp van de vaardigheidsscores. Alle opgaven van dat vakgebied staan namelijk op dezelfde onderliggende vaardigheidsschaal. De vaardigheidsscores van verschillende vakgebieden, zoals rekenen en leesvaardigheid, zijn niet te vergelijken. De verschillende toetsen meten immers andere vaardigheden, waarvoor andere vaardigheidsschalen worden ontwikkeld. Bovendien zijn de vaardigheidsschalen van verschillende aanbieders niet vergelijkbaar. Of de vaardigheden in het po en vo van één aanbieder op dezelfde schaal liggen, verschilt per aanbieder. Aanbieders kunnen de vaardigheidsschalen van eenzelfde vakgebied in het po en vo op één schaal zetten door zogenoemde ankeropgaven op te nemen die de aanbieder zowel in het po als in het vo afneemt (mits de gemeten constructen in po en vo overeenkomen).

Om een doorlopende vaardigheidsschaal van het po naar het vo te realiseren, zouden er landelijke vaardigheidsschalen moeten komen die voor alle aanbieders hetzelfde zijn. Hiermee zouden de vaardigheidsscores van een leerling uit het po direct vergeleken kunnen worden met de vaardigheidsscores uit het vo. Op die manier kunnen leerlingen doorlopend gevolgd worden en is groei t.o.v. het po in kaart te brengen. Een voorwaarde hiervoor is dat de gemeten vaardigheden in het po en het vo hetzelfde zijn, d.w.z. dat dezelfde constructen gemeten worden (zie paragraaf 4.4.2).

Het tot stand brengen van landelijke vaardigheidsschalen is een complex proces dat goed gecoördineerd zou moeten worden. Om te zorgen dat de lvs-toetsen van een vakgebied in het po en in het vo én van verschillende aanbieders op een vaardigheidsschaal komen te liggen, zouden de toetsen verbonden moeten worden door opgaven die door alle aanbieders zowel in het po als in het vo afgenomen worden. Deze opgaven worden ankeropgaven genoemd, die in overeenstemming geconstrueerd moeten worden en vervolgens geproeftoetst. Met een proeftoets worden opgaven in een pilot afgenomen om de kwaliteit van de opgaven te onderzoeken. Omdat aanbieders mogelijk verschillende eisen aan toetsopgaven stellen, kan het moeilijk zijn om tot overeenstemming te komen wat betreft deze set opgaven. Het is belangrijk om overeenstemming te hebben over de inhoud van de opgaven voordat de items geproeftoetst worden; na de proeftoets kunnen de opgaven namelijk niet meer gewijzigd worden. Als een set van ankeropgaven gerealiseerd is, zouden aanbieders deze moeten opnemen in hun toetsen in het po en vo. Hierbij speelt de vraag in welke po-toetsen de opgaven opgenomen kunnen worden. Ankering in de doorstroomtoets brengt uitdagingen met zich mee, omdat de stakes daarbij veel hoger zijn dan bij de volgtoetsen in het vo. Ankering in de reguliere volgtoetsen in het po zou daarom geschikter zijn, hoewel ook daar de stakes hoger zijn dan in het vo, omdat deze toetsen als voorbereiding op de doorstroomtoets worden beschouwd. Al met al lijkt het doortrekken van vaardigheidsschalen te ambitieus, gegeven de verschillen in onderwijsinhoud en de motivatieverschillen (van leerlingen) vanwege andere stakes.

Koppeling via referentieniveaus

Referentieniveaus als koppeling tussen lvs po en vo

Referentieniveaus (zie Bijlage 4 voor een uitgebreide uitleg over wat referentieniveaus precies inhouden) zijn ontwikkeld om prestaties tussen schooltypes met elkaar te vergelijken, en daarom in de basis bij uitstek geschikte instrumenten om gebruikt te worden als informatiebron in de overgang van po naar vo. Tegelijkertijd is de informatie op leerlingniveau door het beperkt aantal referentieniveaus relatief grofmazig. Het verschil tussen het al dan niet behalen van niveau 1F of 2F is soms het verschil tussen een handvol items meer of minder correct maken. Bij toetsing heb je altijd met meetfouten te maken. Echter speelt dit bij referentieniveaus een extra grote rol, omdat we te maken hebben met classificatie waarbij veel leerlingen rond de classificatiegrens scoren.²⁰ Als referentieniveaus op leerlingniveau gebruikt worden, is het daarom raadzaam om hierbij de meetnauwkeurigheid te rapporteren en de resultaten bij voorkeur alleen formatief te gebruiken (hoewel de referentieniveaus niet direct een handelingsperspectief voor leerkrachten bieden). Tot slot willen we meegeven dat de referentieniveaus, gegeven de actualisatie van de kerndoelen en “leeftijd” van de huidige referentieset (inmiddels bijna 20 jaar oud), aan vervanging toe zijn. Na de wijzigingen van de kerndoelen moeten ook de referentieniveaus aangepast worden en moeten nieuwe referentiesets worden gemaakt. De referentiesets kunnen vervolgens worden gebruikt in de normering van de toetsen in termen van referentieniveaus. Hiervoor moeten de aanbieders hun toetsen aanpassen.

Andere mogelijkheden

Hoewel referentieniveaus in de basis geschikt zijn om prestaties in po en vo met elkaar te vergelijken, geeft de grofmazige informatie slechts beperkt handelingsgerichte handvatten aan docenten in het vo. Een andere optie om volgtoetsen uit het po op een zinvolle manier in het vo te kunnen gebruiken, is door fijnmazige informatie uit het po terug te koppelen naar het vo met absolute normen, zoals rapportage in termen van beheersing van deelvvaardigheden of leerdoelen (zie ook hoofdstuk 5.2). Hierbij is uniformiteit in de rapportages van verschillende aanbieders in het po gewenst, zodat stakeholders in het vo dezelfde informatie krijgen van po-scholen met die verschillende volgsystemen gebruiken.

²⁰ Doordat op groepsniveau gerapporteerd wordt op *percentage* behaald referentieniveau speelt deze grofmazigheid op geaggregeerd niveau veel minder een rol. Echter zouden uitspraken op populatieniveau betrouwbaarder zijn als *plausible values* gebruikt worden in plaats van puntschattingen. Dit zou een belangrijk nadeel van het werken met referentieniveaus kunnen wegnemen.

4.5.3 Wenselijkheid

Naast de vraag of het mogelijk is om een koppeling te maken tussen de voltooitsresultaten in het primair en voortgezet onderwijs, is een belangrijk vraagstuk of dit überhaupt wenselijk is. Om dit te achterhalen zijn verschillende scholen en andere stakeholders benaderd. De meningen van zowel scholen als stakeholders zijn enigszins verdeeld.

Sommige stakeholders en docenten zien voordelen in het koppelen van de voltooitsresultaten van het po en vo. Zo wordt genoemd dat zo'n koppeling kan bijdragen aan het waarborgen van continuïteit in de leerloopbanen van leerlingen. Docenten in het voortgezet onderwijs zouden dan vanaf het begin beter kunnen inspelen op de behoeften van de leerling op basis van de gegevens in het primair onderwijs. Ook op de scholen ziet men voordelen van een koppeling tussen het lvs in het po en in het vo. Het zou hen helpen om sneller een inschatting te maken van de groei van de leerling en om een nulmeting in het vo beter te kunnen interpreteren. Daarnaast kan het hen helpen in het beoordelen of een leerling in de juiste onderwijssoort is ingedeeld, in het bepalen waar extra op ingezet wordt in de klas en om zorgleerlingen gericht te identificeren en gericht te ondersteunen. Hierbij lijkt een doorlopende vaardigheidsschaal niet per definitie nodig te zijn, maar zou het kunnen inzien en het juist kunnen interpreteren van de resultaten uit het po al kunnen volstaan.

De meeste stakeholders benadrukken echter voornamelijk de risico's en moeilijkheden bij het opzetten van een dergelijke koppeling. Zo werd door meerdere stakeholders benadrukt dat leerlingen een schone lei verdienen aan het begin van hun loopbaan in het vo en dat het koppelen van po-gegevens voor vo- scholen daarbij niet helpend is. Een van de stakeholders benadrukt specifiek dat de voorspellende waarde en houdbaarheid van lvs-toetsgegevens beperkt lijken te zijn en hiernaar blijven handelen kan de groei en motivatie van leerlingen belemmeren. Daarnaast geven zij aan dat een koppeling van de gegevens zeer onwenselijk is met het oog op kanselijkheid, toetsdruk en de mentale gezondheid van pubers. Schoolleiders en docenten erkennen dat een schone lei ook goed kan zijn voor leerlingen en vinden dat leerlingen de kans moeten krijgen om zich te ontwikkelen.

Daarnaast werden enkele praktische bezwaren genoemd door zowel de stakeholders als de scholen. Er werden zorgen geuit over de technische mogelijkheden en inhoudelijke interpretatiemogelijkheden doordat er verschillende aanbieders in het po en vo aanwezig zijn die op een andere manier de ontwikkeling van de leerlingen in kaart brengen. Om een koppeling mogelijk te maken, moeten dezelfde constructen gemeten worden bij alle aanbieders in zowel po als vo. Een hoge mate van standaardisatie is dan noodzakelijk. Daarnaast maakten zowel stakeholders als scholen zich zorgen om de privacy van leerlingen. Scholen verwachten dit ook terug te horen bij ouders wanneer een koppeling tussen po en vo doorgevoerd wordt. Privacy-technische vraagstukken zouden volgens hen dus eerst goed uitgezocht en ondervangen moeten worden.

4.5.4 Aandachtspunten

Tot slot gaan we in op twee aandachtspunten die van belang zijn bij een koppeling tussen lvs in het po en vo: gebruiksvriendelijkheid en gegevensoverdracht. Als vo-scholen de gegevens van lvs-toetsen uit zowel het po als het vo willen inzien, moeten deze op een gebruiksvriendelijke manier toegankelijk zijn en met elkaar vergeleken kunnen worden. Als dit niet makkelijk gaat, doordat ze bijvoorbeeld meerdere keren in verschillende systemen moeten inloggen en op zoek moeten gaan naar de juiste informatie, kan dit tot frustratie (en verloren tijd) bij de gebruiker leiden.

Dit zou voorkomen kunnen worden door een directe koppeling tussen de gegevens uit lvs-toetsen in het po en vo te maken, die via het overstapdossier overgedragen kunnen worden. Hierbij spelen overwegingen met betrekking tot privacy en databeheer een rol.²¹ Bovendien is – zoals beschreven – uniformiteit in de rapportages van de verschillende lvs-aanbieders het po hierbij van belang, zodat deze eenduidig kunnen worden geïmporteerd en gerapporteerd in het lvs-vo. Omdat dit gaat over de rapportages in het po, valt dit buiten de scope van deze ontwerpstudie voor het vo.

4.6 Voorwaarden voor succesvolle implementatie

Het simpelweg verplichtstellen van een lvs in de onderbouw van het vo lijkt niet eenduidig te leiden tot verbetering van onderwijskwaliteit. De implementatie hiervan vraagt dus om een aantal belangrijke voorwaarden om succesvol te kunnen zijn. De voorwaarden hebben te maken met de implementatie van de lvs-verplichting, het eigenaarschap en de vaardigheden die gevraagd worden van het onderwijsveld en de manier van rapporteren over de lvs-resultaten.

4.6.1. Voorwaarden in de implementatie van de verplichting

Een les die door meerdere stakeholders geleerd werd vanuit de invoering van een verplicht lvs en de doorstroomtoets in het po is dat het belang en doel van het lvs vanaf het begin duidelijk moet zijn. Door duidelijkheid te verschaffen over het belang van het gebruik van een lvs kan bijgedragen worden aan de intrinsieke motivatie van de docenten, aldus een van de stakeholders. Het is daarnaast belangrijk dat voor alle partijen duidelijk is waarvoor het lvs gebruikt dient te worden en dat het geen dubbele functies kan bekleden. Voor scholen kan het rust geven duidelijk te maken dat het gaat om intern gebruik van de resultaten om de onderwijskwaliteit te verbeteren; het gaat niet om afrekenen op de resultaten. Anderzijds zorgt deze duidelijkheid ervoor dat de resultaten meer valide geïnterpreteerd kunnen worden, doordat het lvs daadwerkelijk gebruikt wordt waarvoor het bedoeld is en de kans op oneigenlijk gebruik wordt verkleind.

Daarnaast moet duidelijk gecommuniceerd worden hoe de verplichting eruitziet: wordt bijvoorbeeld gevraagd om een lvs-toets minimaal tweemaal in de schoolloopbaan van de leerling af te nemen of moet de lvs-toets elk schooljaar afgenomen worden? Moeten de lvs-toetsen bij alle leerlingen afgenomen worden? En voor welke vakgebieden zou dit moeten en voor welke vakgebieden wordt het slechts aangeraden? In het po blijkt dit voor scholen niet altijd duidelijk te zijn en daardoor wordt onterecht gedacht dat zij de lvs-toets tweemaal per jaar moeten afnemen. Door duidelijk te zijn in de richtlijnen die gegeven worden en door scholen enige vrijheid hierin te geven, kan toetsdruk en werkdruk van docenten beperkt blijven. Daarnaast maken scholen vaak bewuste keuzes in hoe vaak zij een lvs-toets afnemen en bij welke leerlingen zij een lvs-toets afnemen, zoals beschreven in hoofdstuk 3.2 en 3.3. Deze overwegingen zouden als startpunt moeten dienen van waaruit beslissingen over de invulling van de verplichting opgesteld kunnen worden, aldus een aantal stakeholders.

²¹ <https://www.overstapserviceonderwijs.nl/privacy-en-beveiliging/>

4.6.2 Voorwaarden in het eigenaarschap en de vaardigheden van het onderwijsveld

Wanneer het onderwijsveld geen eigenaarschap voelt over de lvs-resultaten, zullen de resultaten hoogstwaarschijnlijk ook minder bijdragen aan de onderwijskwaliteit. Het is daarom van belang dat het onderwijsveld inziet waarom het belangrijk is het lvs te gebruiken. Daarnaast zouden de resultaten niet bij een centraal persoon in de school moeten blijven steken, maar duidelijk gecommuniceerd worden naar belangrijke schakels in het onderwijsproces van de leerlingen. Hierbij gaat het om de vakdocenten van de desbetreffende vakgebieden, maar de resultaten kunnen ook nuttig zijn voor de docenten op andere vakgebieden die gerelateerd zijn aan de gemeten vakgebieden. Het gevoelde eigenaarschap van docenten en onderwijsondersteunend personeel is een voorwaarde om ook daadwerkelijk iets te doen met de lvs-resultaten.

4.6.3 Voorwaarden in de rapportage van de lvs-resultaten

Om de resultaten uit het lvs direct toepasbaar en betekenisvol te maken voor zowel docenten als leerlingen en hun ouders, is het van belang dat er heldere rapportages worden gegeven. Daarbij geeft een stakeholder aan dat de rapportages het mogelijk moeten maken om leerlingen met zichzelf te vergelijken. Daarnaast zou het concrete handvatten moeten geven, op leerdoelen of deelvaardigheden. Wanneer slechts een cijfer, een score of een referentieniveau wordt gegeven, is dit voor docenten, leerlingen en hun ouders moeilijk te interpreteren. Concrete handvatten zijn dus, volgens meerdere stakeholders, cruciaal voor het vertalen van lvs-resultaten naar handelingen om leerlingen te ondersteunen in het leerproces. Om naast de kwaliteit van de lvs-toetsen ook de kwaliteit van de rapportages te waarborgen, wordt aangeraden een externe partij in te schakelen om de lvs-toetsen hierop te evalueren.

Vervolgonderzoek

5.1 Praktische vaardigheden

Het ministerie van Onderwijs, Cultuur en Wetenschap benadrukt in zijn Kamerbrief van 7 juni 2024 het belang van waardering voor praktijkgerichte prestaties.²² Een van de manieren om praktische vaardigheden en praktijkgerichte prestaties meer waardering te geven, zou kunnen zijn door ze op te nemen in een leerlingvolgsysteem. Voorwaarde hiervoor is dat de praktische vaardigheden wel een inhoudelijke toevoeging moeten zijn op wat er gemeten wordt in het lvs. Dit zou tot uiting kunnen komen in een samenhang van schoolse of naschoolse successen. Daarnaast moet er draagkracht voor zijn in het onderwijsveld om het lvs zinvol te kunnen toepassen. In het huidige onderzoek hebben we gekeken naar de draagkracht in de onderwijspraktijk en bij andere stakeholders in het onderwijs rondom het opnemen van praktische vaardigheden in een lvs. Daarnaast hebben we onderzocht wat er al bekend is over het meten van praktische vaardigheden en hun eventuele samenhang of voorspellende waarde voor de schoolloopbaan van leerlingen in het voortgezet onderwijs.

5.1.1 Conceptualisering van praktische vaardigheden

Om te achterhalen of praktische vaardigheden opgenomen zouden moeten worden in het lvs in de onderbouw van het voortgezet onderwijs, is het eerst van belang om een duidelijke definitie van praktische vaardigheden te formuleren. Over een definitie van (praktische) vaardigheden is echter nog niet zomaar consensus bereikt – zo wordt er bijvoorbeeld binnen Europa, vanwege historische en theoretische ontwikkelingen, al verschillend gedacht over het begrip ‘skill’.²³ In de notitie Vaardigheden in het landelijke curriculum (2023)²⁴ geeft de SLO een meer algemene omschrijving van het begrip ‘vaardigheid’.²⁵ De concrete, door SLO geboden, definitie is als volgt: “Een vaardigheid is een vermogen om door middel van denken en handelen op basis van bepaalde kennis voorliggende problemen op te lossen of taken adequaat uit te voeren.”²⁶ Historisch gezien is er ook kritiek op een generieke, contextonafhankelijke interpretatie van vaardigheden.²⁷ Het werk van Adriaan de Groot spreekt hier tot de verbeelding: hij maakte aannemelijk dat “professionele schakers in de context van dat spel wel analytisch en strategisch kunnen denken, maar die vaardigheid niet noodzakelijk kunnen transfereren naar de context van het dammen, en ook niet per se goed zijn in wiskunde.”²⁸ Met andere woorden: hoewel sommige vaardigheden redelijk algemeen van aard lijken, blijkt dat er nog niet altijd een duidelijke transfer naar een andere context plaatsvindt. Wanneer een leerling analytisch naar teksten kan kijken, betekent dit nog niet automatisch dat de leerling ook sterk is in het analyseren binnen wiskunde.

In de notitie van de SLO worden praktische vaardigheden kort genoemd onder de categorie ‘vakspecifieke vaardigheden’ en worden beschreven als vaardigheden “[...] zoals het doelmatig en verantwoord gebruik van specifieke middelen en materialen, gereedschappen en machines, enzovoort, in de context van een vak of een aantal verwante vakken.” Deze beschrijving van praktische vaardigheden lijkt zich te beperken tot vakspecifieke psychomotoriek, en laat andere veelgenoemde gebieden als organisatorische, sociale en communicatieve vaardigheden buiten beschouwing. Een preciezere uitwerking van relevante praktische vaardigheden (voor in een leerlingvolgsysteem voor voortgezet onderwijs) blijft nodig.

²² Paul, 2024

²³ Clarke & Winch, 2006

²⁴ SLO, 2023

²⁵ Sol & Visser, 2023

²⁶ Sol & Visser, 2023

²⁷ Tricot & Sweller, 2014

²⁸ De Groot (1946)

Dit abstraherende, conceptuele vraagstuk lijkt in veel onderwijscontexten niet direct urgent. De invulling en beheersing van praktische vaardigheden is hier gerelateerd aan criteriumsituaties en -taken – in dit geval gaat het dan om beroepen of beroepsrichtingen.²⁹ Aan de hand hiervan worden, terug redenerend vanuit beroepscontext naar onderwijscontext, praktische vaardigheden geconceptualiseerd. Het betreft in deze gevallen de vaardigheden die onderscheiden kunnen worden van de meer cognitieve, theoretische onderdelen van het aangeboden onderwijs. Te denken valt hierbij aan het verschil tussen onderwijs over de werking van verbrandingsmotoren en onderwijs over het uitvoeren van reparaties aan een wielophanging binnen automonteursopleidingen, en aan het verschil tussen onderwijs over menselijke anatomie en onderwijs over het uitvoeren van kleine chirurgische ingrepen binnen geneeskundige opleidingen.

5.1.2 Praktische vaardigheden in beroepscontext

Zoals gezegd herkennen wij de aandacht voor praktische vaardigheden in het Nederlandse onderwijs eigenlijk primair binnen beroepsgerichte opleidingen – in brede zin zijn dit opleidingen die opleiden voor de uitoefening van een specifiek beroep. Dit soort opleidingen zijn terug te vinden over de gehele waaier van onderwijstypen, te denken valt aan fysiotherapeuten, automonteurs en (tand)artsen. Het is de beroepsgerichte context die hier van belang is voor de conceptualisering van (relevante) praktische vaardigheden. Er is een voorgenomen beroepscontext die weergeeft welke vaardigheden van belang zijn voor de beroepscontext waarvoor wordt opgeleid. De beroepscontext geeft dus richting aan wat de leerling of student bijgebracht moet worden.³⁰ Wat betekent dit voor onderwijscontexten waarin deze (expliciete) beroepscontext ontbreekt?

In de onderbouw van het voortgezet onderwijs, de doelcontext van het in deze studie onderzochte leerlingvolgsysteem, ontbreekt een specifieke beroeps(gerichte)context waaruit praktische vaardigheden geabstraheerd kunnen worden. Dit roept de vraag op hoe praktische vaardigheden in deze context gedefinieerd zouden moeten worden en welke praktische vaardigheden vervolgens relevant zijn om te meten in de onderbouw van het voortgezet onderwijs. Ook in gesprekken met experts op het gebied van Centraal Schriftelijke Praktische Examens (CSPE's) komt naar voren dat praktische vaardigheden vooral meetbaar blijken in beroepscontext. Daarnaast vraagt het beoordelen van deze vaardigheden ook inhoudelijke expertise van de beoordelaar, wat er voor zorgt voor organisatorische uitdagingen wanneer deze vaardigheden op grote schaal op dezelfde manier gemeten zouden moeten worden.

5.1.3 Draagkracht voor het opnemen van praktische vaardigheden in een lvs

Over het algemeen heerst er een tendens op zowel scholen als bij andere stakeholders dat praktische vaardigheden inderdaad meer gewaardeerd zouden moeten worden en dat er meer aandacht voor zou moeten zijn. Echter, slechts 14% van de respondenten op onze survey geeft aan dat praktische vaardigheden onderdeel moeten zijn van het lvs. Om praktische vaardigheden op te nemen in het lvs in de onderbouw van het vo moet aan twee belangrijke voorwaarden zijn voldaan: 1) er moeten algemene praktische vaardigheden te identificeren zijn die zinvol zijn om te meten over de jaren heen, en 2) het moet mogelijk zijn om de praktische vaardigheden gestandaardiseerd en objectief te meten.

²⁹ OCW Skills Platform, 2016; Bjork & Romyn, 1999

³⁰ Bjork & Romyn, 1999

Om praktische vaardigheden te kunnen identificeren die zinvol zijn om over de jaren heen te meten, moet eerst vastgesteld worden wat er verstaan wordt onder praktische vaardigheden. Zoals eerder beschreven, is daar vanuit de literatuur geen eenduidigheid over. Ook in de interviews zien we dat er antwoorden gegeven worden die met verschillende conceptualisering van praktische vaardigheden geassocieerd kunnen worden. Zo werden door stakeholders sociaal-emotionele ontwikkeling en executieve functies genoemd om op te nemen, en door een docent werd motoriek genoemd. Dit zijn allemaal vaardigheden of kenmerken die bij het beoordelen vaak gepaard gaan met een gevaar voor subjectiviteit; een objectieve meting zou de praktijk kunnen helpen. Uit de variatie van genoemde praktische vaardigheden in de interviews blijkt ook dat er verschillen zijn tussen onderwijssoorten in wat er onder praktische vaardigheid verstaan wordt. Het is onduidelijk of er generieke praktische vaardigheden bestaan die voor alle onderwijssoorten gemeten kunnen worden. Naast het eenduidig conceptualiseren van praktische vaardigheden moet er ook draagkracht zijn voor de vaardigheden die gemeten worden. In een interview met een docent van een praktijkvak (techniek) gaf deze docent aan dat het objectief meten van praktische vaardigheden lastig is, omdat dit gebeurt door observaties in de klas. Een hele objectieve en gestandaardiseerde manier van meten is daarom niet wenselijk, omdat daarbij het risico groot is dat leerlingen in hokjes geplaatst worden op basis van de meetbare vaardigheden. Als er een manier zou zijn om subjectieve beoordelingen van professionals in kaart te brengen in een leerlingvolgsysteem, zou dat volgens deze docent wel waardevol kunnen zijn. Daarnaast is het voor een docent van een praktijkvak nuttig om inzicht te hebben in cognitieve vaardigheden die gemeten kunnen worden in een lvs, die voorwaardelijk zijn voor een praktische vaardigheid. Dat is nu op scholen vaak niet het geval, omdat de inzichten in het lvs vakgebonden zijn.

5.1.4 Concluderend

In de omvang van deze ontwerpstudie is het te voorbarig om een uitspraak te doen over de wenselijkheid en mogelijkheid van het opnemen praktische vaardigheden in een leerlingvolgsysteem voor voortgezet onderwijs. Er is vervolgonderzoek nodig: wat zijn precies praktische vaardigheden? Welke zijn in deze context relevant? En hoe zijn deze vaardigheden gestandaardiseerd en uitvoerbaar meetbaar in een leerlingvolgsysteem? De oproep tot vervolgonderzoek onderstreept hier de relevantie van aandacht voor praktische vaardigheden en verdient opvolging, bijvoorbeeld door het onderzoeken van de voorgaande onderzoeksvragen of de verkenning van andere mogelijkheden tot het betekenisvol opnemen van praktische vaardigheden in het voortgezet onderwijs.

5.2 Cognitief Diagnostische Modellen

5.2.1 Achtergrond en vraagstelling

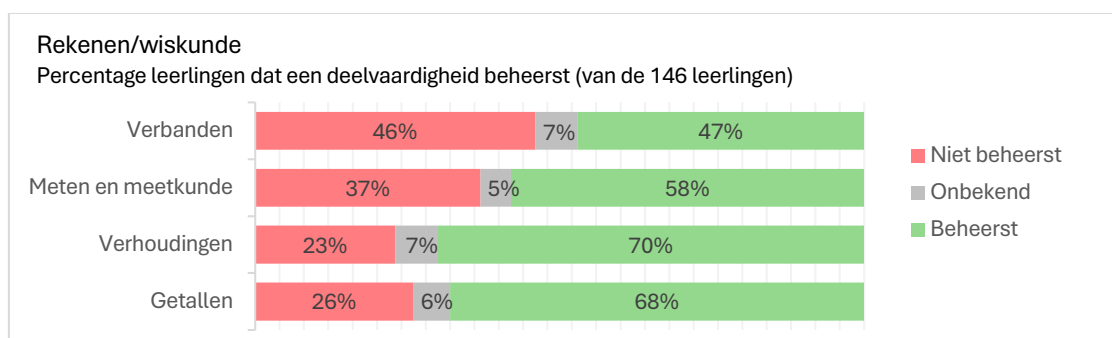
Leerlingvolgsystemen brengen de groei van leerlingen in kaart voor verschillende vakgebieden. Hiermee wordt gemonitord of de groei volgens verwachting verloopt. De volgtoetsen hebben een formatief karakter, waarbij het doel is om bij te sturen als de groei niet passend is. Anders dan bij methodetoetsen en andere vormen van formatieve evaluatie gaat het bij lvs-toetsen echter om een meer algemene en bredere meting van vaardigheden. Een signalering van een achterblijvende groei vertelt niet waar de leerling moeite mee heeft of wat moet worden geredimeerd. Om de resultaten op een formatieve manier in te zetten en het onderwijsaanbod zo goed mogelijk af te kunnen stemmen op de onderwijsbehoeften, kan het voor docenten en leerlingen relevant zijn om informatie te hebben over beheersing van deelvaardigheden, zodat er gericht gehandeld kan worden. Op dit moment zijn lvs-toetsen vaak niet gericht op het meten van deelvaardigheden en het geven van feedback hierover. Een veelbelovende aanpak om gedetailleerdere, diagnostische feedback te geven is cognitief diagnostisch toetsen met behulp van cognitief diagnostische modellen (CDM's). Deze modellen kunnen feedback geven op deelvaardigheden en zijn mogelijk beter geschikt in een formatieve context, waarin feedback op leerprestaties essentieel is. In het kader van het ontwikkelen van ontwerpprincipes voor lvs is de toepassing van CDM's onderzocht.

Cognitief diagnostisch toetsen

Bij cognitief diagnostisch toetsen wordt een set deelvaardigheden gespecificeerd en vervolgens worden items afgenomen die deze vaardigheden meten³¹. Van tevoren wordt door experts vastgesteld welke items welke deelvaardigheden meten (in een zogenoemde Q-matrix)³². Op basis van de antwoorden op deze items worden leerlingen met behulp van CDM's geassocieerd in termen van beheersing van elke deelvaardigheid. Zo wordt een vaardighedenprofiel geschat, waarin per deelvaardigheid is aangegeven of een leerling deze al dan niet beheerst. Deze informatie kan vervolgens worden gerapporteerd op groepsniveau of op individueel niveau.³³ In Figuur 5.1 is een (fictieve) voorbeeldrapportage weergegeven van rekenen-wiskunde op groepsniveau. Hierin is te zien welk percentage van de leerlingen elke deelvaardigheid al dan niet beheerst. In Figuur 5.2 is een (fictieve) voorbeeldrapportage weergegeven van Nederlands leesvaardigheid op individueel niveau. Hierin is per leerling aangegeven of ze de verschillende deelvaardigheden wel of niet beheersen.

De gepresenteerde voorbeelden geven slechts beknopt weer welke informatie gepresenteerd kan worden. In de literatuur worden uitgebreide richtlijnen gegeven voor het construeren van rapportages van cognitief diagnostische toetsen.³⁴ Hierin wordt beschreven hoe resultaten gepresenteerd kunnen worden voor verschillende stakeholders, o.a. aan de hand van visualisaties, personalisatie en beknopte beschrijvingen van de gepresenteerde informatie.

Figuur 5.1 Voorbeeldrapportage op groepsniveau



Figuur 5.2 Voorbeeldrapportage op individueel niveau

Nederlands leesvaardigheid				
Leerling	Begrijpen	Interpreteren	Evalueren	Samenvatten
0001	✗	✗	✗	✗
0002	✓	✓	✓	✗
0003	✓	✗	?	✗
...	...			

³¹ Norris, Macnab & Phillips, 2007

³² Tatsuoka, 1983

³³ Omdat CDM's probabilistische classificatiemodellen zijn (zie *Technische achtergrond CDM's* in bijlage), kan het zijn dat er voor sommige leerlingen onvoldoende zekerheid is om ze te classificeren. Hiervoor kan een categorie Onbekend toegevoegd worden, zoals in de voorbeelden in Figuur 5.1 en 5.2.

³⁴ Zie bijv. Roduta Roberts en Gierl (2010).

Een belangrijke, niet-triviale vraag met betrekking tot de CDM-classificaties is: wat betekent beheersing van een deelvaardigheid? Het is hierbij cruciaal dat deelvaardigheden zodanig gedefinieerd zijn dat classificaties geïnterpreteerd kunnen worden in termen van wat een leerling weet, kan of beheerst. Om deze interpretatievraag te beantwoorden, beschrijven we hoe de classificaties tot stand komen.³⁵ CDM's zijn latente klasse modellen die model-gebaseerde classificaties geven op basis van de modelparameters en itemresponsdata. Om de classificaties te verkrijgen, is geen expertoordeel nodig (behalve bij het specificeren van de Q-matrix). Het model onderscheidt twee groepen leerlingen: een lager presterende en een hoger presterende groep. Hierbij zal de tweede groep beter presteren dan de eerste, maar er kunnen niet direct uitspraken gedaan worden over het behalen van bepaalde standaarden zonder de definities van deelvaardigheden en de items te evalueren. Als een deelvaardigheid heel fijnmazig gedefinieerd is, zal de betekenis van beheersing eenduidig zijn. Denk bijvoorbeeld aan het optellen van breuken, waarbij deelvaardigheden zoals “een geheel getal omzetten naar een breuk” en “een gemeenschappelijke noemer vinden” een rol spelen. Bij zulke fijnmazige deelvaardigheden is een dichotoom onderscheid eenvoudig te interpreteren: een leerling kan het wel of niet. In leerlingvolgsystemen zullen deelvaardigheden echter niet zo fijnmazig gedefinieerd kunnen worden, omdat het aantal te meten deelvaardigheden dan sterk oploopt en daarmee ook het benodigde aantal items om deze deelvaardigheden te meten, en dus de toetsdruk. Als deelvaardigheden breder gedefinieerd worden, zijn geen of weinig extra items nodig ten opzichte van de huidige volgoetsen en zal de toetsdruk niet verhogen. Het kan dan echter lastiger zijn om tot een interpretatie van de classificaties te komen die wordt ondersteund door het psychometrisch model en het ontwerp van de toets. In dat geval kan een standaardbepaling met experts helpen.³⁶

Toepassing van CDM's in een LVS

Op basis van de vaardighedenprofielen die CDM's schatten, kunnen we gedetailleerder feedback geven dan met traditionele psychometrische modellen. Bovendien leiden de uitkomsten van deze modellen mogelijk minder snel tot oneigenlijk gebruik van de toetsresultaten. Deze methode is eerder succesvol ingezet in het hoger onderwijs.³⁷ In deze studie brengen we in kaart wat de mogelijkheden zijn om gebruik te maken van cognitief diagnostische modellen in een leerlingvolgsysteem. De onderzoeksvragen luiden:

- Zijn de onderwijshouden in een lvs geschikt om te modelleren met CDM's?
- Over welke deelvaardigheden is het wenselijk om te rapporteren?
- Helpt de gedetailleerdere rapportage stakeholders bij het vormgeven van hun onderwijs?

Deze vragen worden beantwoord door achtereenvolgens de mogelijkheid, wenselijkheid en bruikbaarheid van de toepassing van CDM's in een LVS te onderzoeken.

5.2.2 Methode

Zoals gezegd bestaat deze studie uit drie onderdelen: mogelijkheid, wenselijkheid en bruikbaarheid. Voordat we hierop ingaan, beschrijven we het gekozen toepassingsdomein. Voor het gebruik van CDM's is het van belang dat de deelvaardigheden gespecificeerd zijn voordat het model toegepast wordt. Het construeren van zo'n cognitief model vereist dat een leergebied goed gedefinieerd is. Daarom beperken we ons in deze deelstudie tot de leergebieden rekenen-wiskunde, Nederlands leesvaardigheid en Engels leesvaardigheid. Naar deze basisvaardigheden is veel onderzoek gedaan, waardoor deze helder gedefinieerd zijn, i.t.t. de nieuwe vaardigheden digitale geletterdheid en burgerschap. Desalniettemin zien we ook bij de nieuwe vaardigheden kansen om CDM's toe te passen.

³⁵ We beperken ons in deze tekst tot een beknopte, conceptuele uitleg; in de bijlage wordt een gedetailleerdere beschrijving gegeven.

³⁶ Henson & Templin, 2008; Min & He, 2022

³⁷ Maas, Brinkhuis, Kester & Wijngaards-de Meij, 2022a

Om de mogelijkheid van het toepassen van CDM's te onderzoeken, wordt eerst aan hand van de kenmerken van (bestaande) lvs-toetsing onderzocht in welke mate deze modellen toepasbaar zijn in het voortgezet onderwijs. Documenten van bestaande lvs-toetsen worden geanalyseerd om na te gaan welke deelvaardigheden daarin aan bod komen. Vervolgens worden psychometrische analyses uitgevoerd op bestaande LVS-toetsdata om te onderzoeken of het toepassen van CDM's psychometrisch haalbaar is. Bij de psychometrische analyses wordt gebruikgemaakt van de R packages CDM38, GDINA39, TDCM40 en mirt41.

Om de wenselijkheid te onderzoeken, worden interviews afgenomen bij stakeholders om te inventariseren of zij het wenselijk vinden om te rapporteren over deelvaardigheden. We interviewen vakdocenten Nederlands, Engels en wiskunde. Dit zijn relevante stakeholders, omdat zij degenen zijn die informatie uit het lvs kunnen gebruiken om direct aanpassingen te doen aan hun onderwijs of om te bepalen welke leerlingen in aanmerking komen voor remediëring. Bij vakdocenten wordt uitgevraagd hoe fijnmazig zij rapportage over deelvaardigheden wenselijk vinden.

Als de psychometrische analyses en interviews resultaten opleveren die leiden tot bruikbare input voor het construeren van (fictieve) diagnostische rapportages, gaan we vervolgens de bruikbaarheid onderzoeken om de meerwaarde van de toepassing van CDM's te bepalen. Door in een vignettenstudie diagnostische rapportages voor te leggen aan docenten kunnen we onderzoeken hoe zij deze informatie interpreteren en zouden (kunnen) gebruiken om hun onderwijs vorm te geven.

5.2.3 Resultaten

Mogelijkheid

Deelvaardigheden in huidige lvs-toetsen

In Tabel 5.1 is weergegeven welke deeldomeinen in de huidige lvs-toetsen in het vo worden getoetst volgens de documentatie van de verschillende lvs-aanbieders. Voor rekenen-wiskunde zijn deze deelvaardigheden voor alle aanbieders gelijk; deze volgen het referentiekader.⁴² Hoewel er voor leesvaardigheid enkele verschillen te zien zijn, komen de typen deeldomeinen wel met elkaar overeen. Bovendien worden deze typen deeldomeinen (begrijpen, interpreteren, etc.) ook gevonden in de literatuur over CDM-toepassingen bij leesvaardigheid.⁴³ Om te bepalen wat de beste manier is om de deelvaardigheden voor een CDM in het vo te specificeren is uitgebreider onderzoek nodig, maar de voorbeelden uit de literatuur laten zien dat toepassing van CDM's in deze leergebieden mogelijk is.

³⁸ George, Robitzsch, Kiefer, Groß & Ünlü, 2016

³⁹ Ma & de la Torre, 2020

⁴⁰ Madison, Haab, Jeon & Cotterell, 2024

⁴¹ Chalmers, 2012

⁴² Expertgroep doorlopende leerlijnen taal en rekenen, 2009

⁴³ Zie bijv. Fan en Yan (2020) of H. Chen en Chen (2016)

Tabel 5.1 Getoetste deeldomeinen zoals beschreven in de documentatie van lvs-aanbieders

Leergebied	CVVO (Cito)	Dia LVS (Diattoetsen)	Jij! (Bureau ICE)
Rekenen-wiskunde	Getallen Verhoudingen Meten en meetkunde Verbanden	Getallen Verhoudingen Meten en meetkunde Verbanden	Getallen Verhoudingen Meten en meetkunde Verbanden
Nederlands leesvaardigheid	Begrijpen en interpreteren Reflecteren	Begrijpen Interpreteren Reflecteren Samenvatten	Techniek en woordenschat Begrijpen Interpreteren Evalueren Opzoeken Samenvatten
Engels leesvaardigheid	Begrijpen en interpreteren Reflecteren	Informatieverwerking op: Microniveau Mesoniveau Macroniveau	Wordt niet getoetst

Psychometrische haalbaarheid

Beschrijving van de data

Om de psychometrische haalbaarheid van de toepassing van CDM's in lvs in het voortgezet onderwijs te onderzoeken, zijn explorerende analyses uitgevoerd op data van het Cito volgsysteem voortgezet onderwijs (CVVO⁴⁴). Er is (willekeurig) één toets gekozen om te analyseren, namelijk de toets rekenen-wiskunde aan het begin van leerjaar 1 van de havo in het schooljaar 2022/2023. Deze toets wordt Toets 0 (T0) genoemd, omdat deze wordt gezien als nulmeting aan de start van het vo. In navolging van het referentiekader⁴⁵ worden in deze toets vier domeinen gemeten, te weten getallen en variabelen (D1), verhoudingen (D2), meten en meetkunde (D3) en verbanden en formules (D4). Deze domeinen definiëren we als deelvaardigheden voor het CDM. De toets bestaat uit 52 unidimensionele items die gelijk verdeeld zijn over de deelvaardigheden, d.w.z. 13 items per deelvaardigheid. De itemresponsdata van 4111 leerlingen waren beschikbaar.

In Tabel 5.2 zijn de gemiddelden en de standaarddeviaties van de ruwe totaalscores en domeinscores weergegeven. Er is een toets- en itemanalyse (TIA) uitgevoerd binnen het raamwerk van de klassieke testtheorie (KTT). Hieruit volgde een Cronbach's alpha van 0,82. Dit is voldoende voor belangrijke beslissingen op individueel niveau volgens de COTAN-richtlijnen.⁴⁶ In Tabel 5.3 zijn gegevens over de p-waarden en rit- en rir-waarden van de 52 items weergegeven. Hierin werden geen opvallende resultaten gevonden.

Tabel 5.2 Gemiddelde en standaarddeviatie (SD) van de ruwe totaalscores en domeinscores van T0

	Totaal	D1	D2	D3	D4
Gemiddelde	34,6	8,6	9,4	8,1	8,5
SD	7,1	2,2	2,3	2,2	2,2

⁴⁴ Van Til & Van Boxtel, 2015

⁴⁵ Expertgroep doorlopende leerlijnen taal en rekenen, 2009

⁴⁶ Evers, Lucassen, Meijer & Sijtsma, 2010

Tabel 5.3 TIA-resultaten van de 52 items van T0

	Gemiddelde	SD	Min	Max
P-waarde	0,66	0,15	0,26	0,91
Rit-waarde	0,31	0,06	0,13	0,41
Rir-waarde	0,25	0,06	0,07	0,35

CDM model fit en betrouwbaarheid

We hebben eerst de toepassing van het log-lineaire cognitief diagnostisch model (LCDM⁴⁷) onderzocht. Hiermee worden leerlingen dichotoom geclassificeerd: ze beheersen een deelvaardigheid wel of niet. De model fit-resultaten zijn weergegeven in Tabel 5.4. De weergegeven fit indices zijn M2⁴⁸, RMSEA2 en SRMSR⁴⁹, gemiddelde item RMSEA, MADcor (de gemiddelde absolute afwijking tussen geobserveerde en voorspelde itemcorrelaties), maximum χ^2/χ^2 , Fischer getransformeerde correlaties en log-odd ratios⁵⁰. Hoewel de significantietoetsen over het algemeen significante resultaten geven, wat duidt op misfit, laten de maten van effectgrootte adequate fit zien.

Als we voor de classificatie uitgaan van een drempelwaarde van 0.5 – d.w.z. bij een geschatte beheersingskans groter dan 0,5 zeggen we dat een leerling de deelvaardigheid beheerst en anders niet – dan vinden we dat de vier deelvaardigheden door 59-62% van de leerlingen beheerst worden. In Tabel 5.5 zijn de classificatie accuraatheid en betrouwbaarheid gegeven. De accuraatheid is een maat van overeenstemming tussen de geobserveerde en verwachte proportie leerlingen die de deelvaardigheden beheerst.⁵¹ De gerapporteerde betrouwbaarheid is de gekwadrateerde punt-seriële correlatie tussen de ware vaardigheid en de posterior mean van de vaardigheid, waarvoor een waarde onder de 0,75 als onvoldoende wordt beschouwd.⁵² Deze resultaten zijn veelbelovend. De geanalyseerde toets is niet ontwikkeld vanuit cognitief diagnostisch perspectief; in de huidige praktijk is een unidimensioneel model gebruikt, namelijk het één-parameter logistisch model (One Parameter Logistic Model⁵³). De latente correlaties tussen de vier deelvaardigheden zijn dan ook hoog: deze liggen tussen 0,89 en 0,90. De verwachting is dat het LCDM tot betere resultaten leidt wanneer de toets vanuit diagnostisch oogpunt wordt geconstrueerd.

Tabel 5.4 LCDM model fit-resultaten

Maat	Waarde	Maat	Waarde
M ₂	2752,9 ($p < .001$)	MADcor	0,025
RMSEA ₂	0,017	Max(χ^2)	105,3 ($p_{holm} < .001$)
SRMSR	0,032	Max. log-odds ratio	0,886 ($p_{holm} < .001$)
Mean item RMSEA	0,021	Max. transf. cor.	0,171 ($p_{holm} < .001$)

Tabel 5.5 LCDM-classificatie accuraatheid en betrouwbaarheid

	D1	D2	D3	D4
Accuraatheid	0,91	0,92	0,91	0,88
Betrouwbaarheid	0,74	0,76	0,74	0,65

⁴⁷ Rupp, Templin & Henson, 2010

⁴⁸ Maydeu-Olivares & Joe, 2006

⁴⁹ Maydeu-Olivares & Joe, 2014

⁵⁰ Chen, de la Torre & Zhang, 2013

⁵¹ Wang, Song, Chen, Meng & Ding, 2015

⁵² Johnson & Sinharay, 2020

⁵³ zie Van Til & Van Boxtel, 2015

Mogelijk zijn stakeholders geïnteresseerd in gedetailleerdere informatie over het niveau van beheersing van deelvaardigheden dan volgt uit de dichotome classificaties. Daarom hebben we gekeken naar de mogelijkheid om polytome deelvaardigheden te modelleren met het polytome diagnostische classificatiemodel (PDCM⁵⁴). Hiermee kunnen leerlingen per deelvaardigheid op meerdere beheersingsniveaus geclassificeerd worden. We hebben gekeken naar een PDCM met drie beheersingsniveaus. Dit model is nog niet geïmplementeerd in de gangbare CDM-packages in R. Om deze reden hebben we het *mirt* package gebruikt. Omdat het multidimensionele PDCM hiermee niet gespecificeerd kon worden, is voor elke deelvaardigheid een apart model geschat. De model fit-resultaten van deze modellen zijn weergegeven in Tabel 5.6. Omdat er een ander package is gebruikt, zijn deze resultaten minder uitgebreid dan bij het LCDM. Wel zien we een vergelijkbare trend: de significantietoetsen geven significante resultaten, wat duidt op misfit, maar de maten van effectgrootte laten adequate fit zien.

Tabel 5.6 *PDCM model fit-resultaten*

	D1	D2	D3	D4
M₂	88,8 ($p = .001$)	119,5 ($p < .001$)	115,2 ($p < .001$)	79,6 ($p < .001$)
RMSEA₂	0,014	0,018	0,018	0,012

We hebben de polytome classificaties van het PDCM vergeleken met de dichotome classificaties van het LCDM. Met het PDCM zijn de leerlingen per deelvaardigheid geclassificeerd in één van drie categorieën, die we voor nu beginner, gevorderde en expert noemen.⁵⁵ Van de leerlingen en deelvaardigheden die werden geclassificeerd als beginner, werd 99,7% door het LCDM geclassificeerd als niet beheerst. Van de leerlingen en deelvaardigheden die werden geclassificeerd als gevorderde, de middelste categorie, werd 56,2% door het LCDM geclassificeerd als niet beheerst en 43,8% als wel beheerst. Van de leerlingen en deelvaardigheden die werden geclassificeerd als expert, werd 95,9% door het LCDM geclassificeerd als wel beheerst. De classificaties in de twee uiterste categorieën van het PDCM zijn dus sterk in overeenstemming met de dichotome classificaties van het LCDM.

In een lvs wordt de vaardigheid van leerlingen gevolgd over tijd door meerdere toetsen af te nemen. Binnen het CDM-raamwerk bestaat hiervoor het transitie diagnostisch classificatiemodel (TDCM⁵⁶). Dit is een longitudinaal CDM waarmee groei wordt gemodelleerd als verandering in beheersingsstatus over tijd. Op groepsniveau worden met dit model zogenoemde transitiekansen geschat, die aangeven wat de kans is dat leerlingen van geen beheersing naar wel beheersing gaan (en vice versa). Het TDCM biedt kansen in een lvs-context, maar er zijn wel een aantal voorwaarden: het is van belang dat de toetsen dezelfde constructen meten en er zijn ankeritems nodig. Dit zijn items die onderdeel zijn van de toetsen op verschillende meetmomenten. Hoewel CDM resultaten in theorie item-invariant zijn⁵⁷, wat betekent dat de classificaties onafhankelijk zijn van welke items afgenomen zijn, is het zonder ankeritems heel moeilijk om af te dwingen dat de betekenis van beheersing hetzelfde is bij verschillende toetsen. Dit is een cruciaal punt in lvs-toetsen, waar mogelijk verschillende leeruitkomsten worden gemeten in verschillende toetsen/leerjaren, omdat eindtermen kunnen oplopen in niveau. In dat geval verandert de betekenis van beheersing over tijd en is het TDCM niet geschikt om toe te passen.⁵⁸ Hiervoor is aanvullend onderzoek nodig naar de gemeten constructen.

⁵⁴ Bao, 2019

⁵⁵ In de praktijk moet nagedacht worden over de keuze van deze labels. De gekozen labels kunnen namelijk een rol spelen bij de interpretatie van classificaties, zoals we verderop zullen toelichten.

⁵⁶ Madison & Bradshaw, 2018

⁵⁷ Bradshaw & Madison, 2016

⁵⁸ Mogelijk is het *polytomous generalized deterministic inputs, noisy, "and" gate model* (pG-DINA; J. Chen & de la Torre, 2013) dan toepasbaar. Bij dit model worden voorafgaand aan de toetsafname per deelvaardigheid verschillende beheersingsniveaus gedefinieerd.

Ter exploratie hebben we enkele longitudinale analyses uitgevoerd. Hiervoor is in aanvulling op de data van T0 gekeken naar de afname van de toets rekenen-wiskunde aan het eind van leerjaar 1 van de havo in het schooljaar 2022/2023. Deze toets wordt Toets 1 (T1) genoemd. In deze toets werden dezelfde vier deelvaardigheden gemeten als in T0, ook elk door 13 unidimensionele items. Er was geen overlap in de items van de twee toetsen. Er waren data van 3268 leerlingen beschikbaar die zowel T0 als T1 gemaakt hadden. De resultaten van de psychometrische analyses waren niet zoals verwacht. Er was geen groei te zien voor D3 en D4; voor geen enkele leerling werden deze deelvaardigheden door het TDCM als beheerst aangemerkt bij T1, terwijl het schatten van het LCDM op T1 liet zien dat bijna de helft van de leerlingen deze deelvaardigheden beheerst. We vermoeden dat dit te maken heeft met het feit dat de toetsen niet precies dezelfde constructen meten, maar we hebben geen inzicht gehad in de items om dit te verifiëren, en bovendien waren er geen ankeritems aanwezig. Er is uitgebreider onderzoek nodig naar de toepassing van het TDCM in een lvs.

Psychometrische richtlijnen

Om leerlingen direct te kunnen classificeren op basis van hun itemresponsdata is een gekalibreerde itemset nodig. Voor de kalibratie is het van belang dat er voldoende data aanwezig zijn. Maas en collega's⁵⁹ geven richtlijnen voor benodigde leerling- en itemaantallen, die ruimschoots haalbaar zijn binnen de context van een lvs in het vo; er kunnen bijvoorbeeld met gegevens van honderd leerlingen en dertig items al vier deelvaardigheden accuraat gemeten worden.

Na de classificatie volgt een interpretatie hiervan. De betekenis van de beheersingsniveaus komt voort uit de interpretatie van toetsontwikkelaars, zoals we eerder al beschreven. Deze interpretatie moet ondersteund worden door het psychometrisch model en het ontwerp van de toets. Dit is een niet-triviale, maar belangrijke stap.⁶⁰ Daarnaast spelen de gebruikte labels een rol bij de interpretatie van classificaties. Bij een dichotome classificatie kan bijvoorbeeld beheerst vs. niet beheerst of op schema vs. aandachtspunt gebruikt worden. Het eerste kan een bijklink van een afgerond leerproces hebben, terwijl het tweede op een lopend leerproces duidt.

In het empirische deel van de huidige studie zijn we uitgegaan van de beschikbare gegevens. Zo hebben we deelvaardigheden gedefinieerd op basis van de aanwezige metadata voor de items (die aangaf welke domeinen elk item meet), in plaats van eerst een cognitief model te specificeren en daar items bij te verzamelen of construeren. Bovendien bieden CDM's mogelijkheden om complexere processen te modelleren dan we in deze studie gedaan hebben. Er kunnen bijvoorbeeld multidimensionele items in een toets opgenomen worden (d.w.z. items die meerdere deelvaardigheden meten), waarbij de bijdrage van verschillende deelvaardigheden aan de kans op een correct antwoord op verschillende manieren gemodelleerd kan worden. Ook kan een hiërarchische structuur in deelvaardigheden worden gemodelleerd, waarbij een bepaalde deelvaardigheid als voorkennis voor een andere deelvaardigheid wordt beschouwd.⁶¹ Zo zijn er nog meer uitbreidingen mogelijk. Het vakgebied CDM is relatief nieuw in vergelijking met IRT, waardoor er nog veel aspecten te onderzoeken zijn, maar de methodologische ontwikkelingen zijn op dit moment al ver genoeg om de basis CDM-modellen in de praktijk toe te passen.

⁵⁹ Maas, Brinkhuis, Kester en Wijngaards-de Meij, 2022b

⁶⁰ zie Bradshaw & Levy, 2019

⁶¹ Templin & Bradshaw, 2014

Wenselijkheid en bruikbaarheid

We hebben middels interviews met vakdocenten Nederlands, Engels en wiskunde geïnventariseerd of zij diagnostische rapportages in een lvs wenselijk vinden. Dit zijn stakeholders die deze informatie zouden kunnen gebruiken om vakspecifieke aanpassingen te maken aan het onderwijs. Dit gaat verder dan de vraag welke leerling remediëring nodig heeft voor een bepaald vak: het kan richting geven aan de inhoud van deze remediëring.

De gesprekken laten gemengde resultaten zien. Docenten Nederlands en Engels geven enerzijds aan dat remediëring binnen leesvaardigheid niet gericht zou moeten zijn op specifieke deeldomeinen, maar anderzijds zoeken ze wel naar specifiekere, handelingsgerichte handvatten dan enkel 'leesvaardigheid verbeteren'. Hoewel sommige lvs'en hier ondersteuning voor bieden (bijvoorbeeld door te rapporteren over leesvaardigheid op micro-, meso- en macroniveau), lukt het docenten vaak niet om met de resultaten uit een lvs tot een overzicht van benodigde acties te komen. Wiskundeleraars gaven aan dat de deeldomeinen die op dit moment in de lvs'en gemeten en gerapporteerd worden (zie Tabel 5.1) relevant zijn, maar slechts een klein onderdeel van wiskunde beslaan.⁶²

Hoewel docenten aangeven op zoek te zijn naar handelingsgerichte feedback, waar CDM's een bijdrage aan kunnen leveren, konden zij over het algemeen niet goed aangeven welke definitie van deelvaardigheden in de praktijk het meest zinvol zou zijn om over te rapporteren. We gaven eerder al aan dat uitgebreider onderzoek nodig is voor specificatie van deze deelvaardigheden en de interviews met docenten bevestigen dat nogmaals. Alvorens we de bruikbaarheid van CDM-rapportages goed kunnen onderzoeken, is het nodig deelvaardigheden te definiëren. Daarmee kunnen vervolgens realistische vignetten met CDM-rapportages geconstrueerd worden. Die kunnen dan worden voorgelegd aan docenten om na te gaan hoe zij deze informatie interpreteren en zouden (kunnen) gebruiken om hun onderwijs vorm te geven.

5.2.4 Conclusie

Al met al zouden CDM's een belangrijke rol zouden kunnen spelen in de vraag uit de praktijk naar handelingsgerichte aanwijzingen in lvs-rapportages. De eerste analyses suggereren dat het toepassen van CDM's in een lvs psychometrisch haalbaar is. Er is echter nog wel extra onderzoek nodig om deelvaardigheden goed te kunnen definiëren (zowel psychometrisch als onderwijskundig). Dan kan ook geëvalueerd worden of de resulterende rapportages daadwerkelijk zinvol zijn voor stakeholders.

5.3 Omgaan met de resultaten

Toetsscores en resultaten worden veelal gepresenteerd in rapportages. Deze rapportages zijn het middel waarmee toetsresultaten vertaald worden naar betekenisvolle acties.⁶³ Bij het beantwoorden van de vragen over hoe te rapporteren, is het van groot belang wat gerapporteerd moet worden, wie er ingelicht moet worden, en wat de ontvanger van de rapportage ermee moet doen, oftewel wat het doel van de rapportage is. Deze zaken moeten op elkaar afgestemd zijn omdat anders ongewenste neveneffecten op de loer liggen.

Naast een keuze over 'wat' er gerapporteerd wordt, is het ook belangrijk over 'wie' de rapportage gaat: over een individuele leerling, over een klas, over alle leerlingen van één leerkracht, over alle leerlingen in een school of zelfs over het Nederlandse onderwijsstelsel als geheel. Afhankelijk van het beoogde doel kan een verschillend aggregatieniveau gekozen worden om de toetsresultaten te presenteren.

⁶² Additionele relevante deelvaardigheden om te rapporteren voor rekenen-wiskunde in een lvs die werden genoemd, zijn informatie uit tabellen en grafieken verwerken.

⁶³ Hopster-den Otter, Wools, Eggen & Veldkamp, 2017

Afhankelijk van het doel wordt bepaald welke informatie het best gepresenteerd kan worden. Als de leerling ingelicht wordt dan is andere informatie relevant dan als de leerkrachten of de scholen ingelicht worden. Dat geldt ook als de centrale toetsen dienen voor een systeemmeting, waarbij de centrale toetsen een deel van de functie van peilingsonderzoek hebben. Het is niet ongebruikelijk om in de praktijk verschillende vormen van rapportages te geven geschikt voor verschillende doelgroepen (leerling, ouders, leerkracht, school, bestuurder) en afgestemd op de informatiebehoefte.

5.3.1 Het terugkoppelen van resultaten aan leerkrachten en scholen

Valide toetsscores zijn scores die geschikt zijn voor de beoogde interpretatie en het gebruik van een toets. Om te bepalen of toetsscores valide zijn, is het noodzakelijk dat goed gedefinieerd is wat de score betekent en hoe men deze score zou willen gebruiken: het doel van de toets. Het is dan ook van belang dat aangetoond wordt dat de toetsscores geschikt zijn voor de beoogde interpretatie en het beoogde gebruik.⁶⁴ Bij het beoogde gebruik speelt de terugrapportage een zeer belangrijke rol, want een juiste interpretatie van toetsresultaten is een noodzakelijke voorwaarde voor een adequaat gebruik.⁶⁵ Het ondersteunen van de gebruikers bij het interpreteren van de toetsresultaten is zodoende een belangrijk aspect van validiteit.

Het is dus duidelijk dat we de rapportage beschouwen als belangrijk element van de validiteit. Het doel van de toets en de vorm van de rapportage zijn dan ook sterk aan elkaar gerelateerd. Het doel van de toets en het daadwerkelijke gebruik worden met elkaar verbonden door de rapportagevorm: daar waar de rapportage het gevolg is van het doel, zal het de oorzaak van het uiteindelijk gebruik zijn. Dat betekent dat de vraag of de toets wordt gebruikt volgens het doel grotendeels zal afhangen van de rapportage. Bij het ontwikkelen van de rapportage moet zodoende “achteruit gedacht” worden: levert deze rapportage wel het beoogde doel op? En nog breder: past deze vorm van de rapportage bij de uitgangspunten van de toets?

Naast de gegevens die de ontwikkelaar van de terugrapportage wil overbrengen, is het van belang om aan te sluiten bij de informatiebehoefte van de gebruiker. De rapportage kan gezien worden als vorm van communicatie: met een afzender, een boodschap en een ontvanger. De afzender van een toetsrapportage is meestal de toetsontwikkelaar die de resultaten wil presenteren. De boodschap heeft betrekking op de inhoud van de rapportage en het publiek bestaat uit mensen die de toetsresultaten willen gebruiken.⁶⁶

5.3.2 Rapportages in leerlingvolgsystemen

Zoals eerder beschreven kunnen bij het gebruik van een vaardigheidsschaal drie verschillende vormen van normeringen gebruikt worden: het vergelijken van de leerlingprestatie met een eerdere prestatie van zichzelf (groeigerichte normering), de vergelijking van de prestatie van een leerling met een inhoudelijk criterium (absolute normering) en de vergelijking van de prestatie van een leerling met andere (groepen) leerlingen (relatieve normering). Het is van belang om een normering te kiezen die leidt tot een rapportage die aansluit bij het beoogde gebruik van leerlingvolgsystemen.

⁶⁴ American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 2014

⁶⁵ Van der Kleij et al., 2014

⁶⁶ Ryan, 2006

Het doel van leerlingvolgsystemen is om systematisch leerlingprestaties te volgen in de tijd. Dit moet inzicht geven in de groei van de leerling en aanknopingspunten bieden voor formatief handelen. Volgsystemen rapporteren groei door het vaardigheidsniveau van een leerling op verschillende momenten te vergelijken, vaak aangevuld met normgroepvergelijkingen om de prestatieniveaus in de populatie te duiden. Deze relatieve normering is voor scholen met name zinvol om te gebruiken voor determinatie. Echter kan het ook bijdragen aan het zogenoemde 'hokjesdenken', waarbij leerlingen vroeg in hun schoolcarrière worden vastgezet in een bepaalde onderwijssoort en waarbij doorstroming naar een ander niveau moeilijk is. Bovendien geeft het niet noodzakelijk aanknopingspunten voor formatief gebruik. Dit geldt ook voor de rapportages in termen van groei. Een signalering van een achterblijvende groei vertelt niet waar de leerling moeite mee heeft of wat moet worden geredieerd. Om de rapportages voor de leerlingvolgsystemen te verbeteren adviseren we een inhoudelijke duiding van de gerapporteerde beheersingsniveau, bijvoorbeeld met diagnostische rapportages en rapportages gebaseerd op beheersing van specifieke doelen (zie ook hoofdstuk 5.2). Daarbij is het ook van belang om in ogenschouw te houden dat er niet gewoonweg gerefereerd kan worden naar het behalen van 'het niveau van een bepaalde onderwijssoort', gezien de grote overlap in vaardigheidsniveaus in de verschillende onderwijssoorten in het voortgezet onderwijs.⁶⁷

⁶⁷ Elffers, 2023

Conclusies en aanbevelingen

Het afgelopen jaar heeft een team van onderzoekers van Oberon, Stichting Cito en Universiteit Twente onderzoek gedaan naar de praktijk van het gebruik van leerlingvolgsystemen in het voortgezet onderwijs en naar richtlijnen voor een succesvolle implementatie. Daarbij is beschikbare documentatie geanalyseerd en zijn aanbieders, scholen, leerlingen, beleidsmakers en experts bevroegd. De bevindingen in deze ontwerpstudie leiden tot een aantal concrete aanbevelingen.

In het onderzoek komt naar voren dat 81% van de scholen al gebruikmaakt van een lvs. De manier waarop verschilt sterk. Dat kan te maken hebben met de manier waarop de informatie uit het lvs wordt ingezet in het onderwijs (op leerlingniveau, voor bijvoorbeeld feedback geven of overgangsbepalingen, of op schoolniveau, voor bijvoorbeeld verbeteren van onderwijsprocessen), maar kan ook gaan over voor welke vakken het lvs wordt ingezet (hoofdzakelijk Nederlands en rekenen-wiskunde) of voor welke leerjaren. Ook verschilt het wie de gegevens gebruikt. Op bijna alle scholen gebruiken leraren deze gegevens, terwijl locatie/schoolleiders deze informatie gebruiken op 30% - 40% van de scholen. Een kleine 20% van de scholen gebruikt nog geen lvs. Een deel van deze scholen is echter wel van plan een lvs te implementeren. Een van de voordelen van het gebruik van een lvs in het vo is dat leerlingprestaties vergeleken kunnen worden met een landelijke standaard. De vier bestaande lvs'en zijn voor een belangrijk deel gebaseerd op de referentieniveaus. Een inhoudelijke vergelijking met deze referentieniveaus laat zien dat de verschillende domeinen voor taal en rekenen inderdaad goed afgedekt worden in alle vier de lvs'en. Er zijn enkele, relatief kleine, verschillen in welke onderdelen worden afgedekt in de verschillende lvs'en. Recentelijk is er een nieuwe generatie kerndoelen door de SLO geformuleerd die concreter zowel aanbods- als beheersings- en ervaringsdoelen aangeven. De inhoud van de lvs-toetsen zijn ook vergeleken met deze conceptkerndoelen. Wat daarbij opvalt is dat de inhoudelijke dekking voor wiskunde zelfs toegenomen is, terwijl voor de talen de dekking van de concepteendoelen aanmerkelijk lager is dan die van de referentieniveaus.

De inhoudelijke aansluiting van lvs-inhoud op het lesprogramma wordt als een van de belangrijkste manieren genoemd om de intrinsieke motivatie van leraren en leerlingen voor het gebruik van lvs'en te verhogen. Daarentegen wordt het als motivatieverlagend ervaren dat de voorspellende waarde van lvs-toetsen voor toekomstige leerprestaties laag is en dat een verplichting van een lvs een gebrek aan vertrouwen in docenten kan uitstralen. Er wordt benadrukt dat er eigenaarschap nodig is bij alle betrokken medewerkers op de school, en de resultaten niet bij één persoon te laten liggen, om de lvs-resultaten ook daadwerkelijk zinvol in te zetten. Andere nadelen van het gebruik van een lvs zijn het risico op toetsstress bij leerlingen, werkdruk bij leraren en onduidelijkheid van het belang. Ook schuilen er gevaren in het sturende effect dat de invoering kan hebben op het curriculum, waardoor er meer focus komt op de makkelijk meetbare onderdelen van het curriculum en er minder vrijheid voor eigen invulling is voor scholen. Ten slotte zijn er zorgen over oneigenlijk gebruik van het lvs. Als voorbeelden van oneigenlijk gebruik worden verantwoording richting toezichthouder, vergelijking van docenten en het nemen van determinatiebepalingen genoemd. Lvs-toetsen zijn daar niet voor ontworpen en de data worden niet met dat oogmerk verzameld. Een eenduidig doel en een zeer begrijpelijke, heldere en duidelijke uitleg is hierbij nodig.

Een onderwerp dat bij lvs vo ook geregeld ter sprake komt, is de koppeling met gegevens uit lvs po en/of met de doorstroomtoetsen. Psychometrisch gezien is dit een uitdaging omdat de inhouden van de lvs-toetsen en lvs-doorstroomtoets vo op dit moment niet goed op elkaar aansluiten. Koppeling via referentieniveaus lijkt een mogelijkheid, maar daarvoor zouden deze niveaus fijnmaziger gedefinieerd moeten worden. Om een koppeling mogelijk te maken is het wel nodig om zorgvuldig aandacht te besteden aan gebruikersvriendelijkheid van een dergelijke koppeling en privacy-/databasebeheerissues. Over de wenselijkheid van een dergelijke koppeling zijn de meningen van leraren en stakeholders overigens verdeeld. De koppeling kan bijdragen aan de continuïteit, maar kan ook belemmeren dat leerlingen met een schone lei kunnen beginnen.

Naast het evalueren van het huidige gebruik van lvs in het vo, is in deze studie onderzocht wat nodig zou zijn om lvs succesvol te implementeren in het vo voor Nederlands, Engels, wiskunde, digitale geletterdheid en burgerschap. Hiervoor bieden de nieuwe conceptkerndoelen een goede basis. Allereerst is gekeken in hoeverre het mogelijk is om deze kerndoelen te meten in een lvs. Het zijn met name kerndoelen die een beroep doen op cognitieve processen die verdergaan dan reproduceren, begrijpen en toepassen die zich minder goed lijken te lenen voor toetsing in een lvs. Kerndoelen die een beroep doen op het vermogen van leerlingen tot analyseren, evalueren en creëren vragen meestal complexe en uitgebreide opdrachten die meer nakijktijd vragen. Dit blijkt vooral bij Nederlands en burgerschap het geval. Bij burgerschap speelt bovendien nog mee dat de meeste bestaande instrumenten zich veel meer richten op het meten van attitudes, wat niet terugkomt in de kerndoelen. Het meten van attitudes brengt daarnaast een aantal risico's met zich mee, zoals het meten van sociaal-wenselijke antwoorden en het beoordelen van waarden, die niet te meten zijn met een goed/fout-beoordeling. Het eenzijdig meten van kennisdoelen binnen burgerschap zou er daarentegen voor zorgen dat het construct 'burgerschap' niet valide gemeten wordt.

Om een lvs succesvol te verplichten in de onderbouw van het vo, kunnen ook lessen getrokken worden uit de implementatie en het gebruik in het po. In het onderzoek komt duidelijk naar voren dat het vanaf het begin duidelijk moet zijn waar een verplicht lvs voor ingezet gaat worden. Wenselijkheidsonderzoek geeft aan dat de voorkeur uitgaat naar een formatief gebruik en dat het als onwenselijk gezien wordt wanneer scholen, leraren of leerlingen beoordeeld worden op hun lvs-resultaten. Ook moet duidelijk gecommuniceerd worden wat onder een dergelijke verplichte invoering wordt verstaan (hoe vaak afnemen, welke vakken, welke leerlingen, etc.). Het is ten slotte belangrijk dat het onderwijsveld zich eigenaar voelt en het belang inziet. Daarbij is de directe toepasbaarheid en een inzichtelijke rapportage cruciaal. Rapportage op het niveau van deelvaardigheden via somscores of via cognitive diagnostic models biedt hiervoor mogelijkheden.

De vraag blijft wel in hoeverre het wenselijk is om het gebruik van een lvs te verplichten. Op basis van dit onderzoek concluderen we dat de meerwaarde van een verplichting wordt betwijfeld en dat het risico's oproept met betrekking tot weerstand vanuit het veld, toegenomen toetsdruk en verlies van autonomie van scholen en docenten. Als alternatief zou overwogen kunnen worden om scholen weliswaar te verplichten om basisvaardigheden te monitoren, maar hen vrij te laten in de manier waarop.

De studie heeft geresulteerd in een aantal aanbevelingen voor het implementeren en ontwikkelen van lvs'en en voor wanneer de landelijke verplichting voor het gebruik van een lvs in de onderwijs van het vo wel wordt ingevoerd.

Een lvs zou niet gebruikt moeten worden voor

- Verantwoording voor de onderwijskwaliteit door een toezichthouder (zowel intern als extern); Dit leidt af van het voornaamste doel van het lvs en kan zorgen voor druk om hoge prestaties te leveren. Dit kan leiden tot teaching-to-the-test.
- Summatieve oordelen (bijv. determinatie); Lvs-toetsen zijn ontwikkeld voor formatieve doeleinden en zijn dus niet geschikt voor summatief gebruik.
- Een vergelijking van resultaten tussen docenten. Ook dit is voor docenten een vorm voor externe verantwoording, waardoor ambivalentie in doelstellingen ontstaat. Dit kan ten koste gaan van een valide beeld van de ontwikkeling van de leerlingen.

Advies beleidsmakers

- Besluit in hoeverre vaardigheden (digitale geletterdheid) en/of attitudes (burgerschap) een plek moeten krijgen in het lvs en wees bewust van de verschillende voor- en nadelen en draagkracht die hiermee gepaard gaan
- Beperk het gebruik tot formatieve doeleinden op individueel of klasniveau. Zo voorkom je dat met het lvs meerdere doelen worden nagestreefd en behoud je de validiteit.
- Investeer in heldere communicatie rond een verplichte invoering. Scholen hebben behoefte aan duidelijke kaders, zodat ze weten wat wel en niet van hen verwacht wordt.
- Zie voorlopig af van het koppelen van lvs-gegevens uit po en vo. Deze koppeling kan momenteel nog niet betekenisvol en valide gelegd worden.
- Introduceer een kwaliteitsborging van lvs-toetsen. Dit garandeert de kwaliteit van de toetsen voor scholen.

Advies aanbieders

- Neem de nieuwe conceptkerndoelen als uitgangspunt voor een actuele invulling van de inhoud. Deze zijn actueler dan de referentieniveaus.
- Concretiseer de kerndoelen zodat ze geschikt zijn voor operationalisering binnen lvs-toetsen. Momenteel is het bij sommige kerndoelen nog onduidelijk hoe dit tot concreet uiting kan komen in de praktijk en hoe dit in verschillende leerjaren en niveaus naar voren kan komen. Concretisering van de kerndoelen draagt bij aan het valide ontwikkelen van de toetsen.
- Beperk de inhoud van het lvs tot objectief meetbare doelen. Dit komt ten goede van de betrouwbaarheid en validiteit van de toetsen.
- Verzorg een externe borging van de inhoudelijke en psychometrische kwaliteit van de lvs-toetsen.
- Onderzoek welke randvoorwaarden er zijn voor een rapportage die direct toepasbaar en inzichtelijk is. Scholen hebben namelijk baat bij een direct toepasbare rapportage om het zinvol in te kunnen zetten in hun onderwijspraktijk.
- Stimuleer dat het lvs dusdanig gebruiksvriendelijk vormgegeven is dat docenten gemotiveerd raken het te gebruiken.
- Neem wensen van docenten als uitgangspunt bij het ontwikkelen van de rapportages.

Advies leraren

- Verhoog motivatie van leerlingen door uit te leggen wat een lvs hen oplevert. Intrinsieke motivatie van leerlingen zal bijdragen aan het valide en betrouwbaar meten van hun ontwikkelingen.
- Integreer de afname van het lvs in het leerproces ter voorkoming van toets stress. Toets stress is namelijk al een probleem in het voortgezet onderwijs en het is onwenselijk om deze verder te vergroten.

Beperkingen

Wel zijn er beperkingen waar rekening mee gehouden moet worden:

- Een lvs is niet geschikt om de meer cognitieve vaardigheden te meten.
- Het draagvlak voor het verplicht invoeren van een lvs vo is laag bij schoolleiders, docenten en leerlingen

Ten slotte willen we nog opmerken dat er binnen het vo ook een toegenomen aandacht is voor praktijkgerichte vakken. Dit komt, los van het vakgebied digitale geletterdheid, nog niet terug in de lvs'en. Praktijkgerichte vaardigheden zijn vaak complexe constructen die vragen om een passende aanpak. Uitgangspunt bij een lvs is dat er een doorlopende leerlijn is die gemeten wordt op een concrete op itemresponstheorie gebaseerde vaardigheidsschaal en deze aanpak lijkt niet te passen bij het meten van praktische vaardigheden. Praktijkgerichte vaardigheden staan echter centraal binnen het vmbo en ook binnen havo en vwo komt er steeds meer aandacht voor. Bovendien is er bij deze vaardigheden ook sprake van een opbouw via het beheersen van deelvaardigheden en het succesvol uit kunnen voeren van subtaken tot de leerling het gevraagde beheersingsniveau heeft aan einde van het vierde, vijfde of zesde leerjaar. Het meten van praktische vaardigheden moet in een praktische context gebeuren, of opdrachten moeten een expliciet praktische context representeren. Bovendien moet helder zijn hoe praktische vaardigheden in het voortgezet onderwijs gedefinieerd zijn. Dit is echter nog niet volledig uitgekristalliseerd.

Een aanvullende aanbeveling is daarom:

- Verken hoe praktijkgerichte vaardigheden gemeten kunnen worden binnen een lvs.

Onderzoeksverantwoording

Interviews

Voor de beantwoording van de onderzoeksvragen van werkpakket 1, 5 en 6 is gebruik gemaakt van semi-gestructureerde interviews. Er zijn gesprekken gevoerd met experts op het gebied van lvs-gebruik alsmede met stakeholders in het veld. De stakeholders met wie gesproken is, staan in Tabel B1 (in willekeurige volgorde).

Tabel B1.1 Geïnterviewde stakeholders

Nederlandse Vereniging van Wiskundeleraren	Vereniging Levende Talen	Bureau ICE
SLO	Landelijk Actie Komitee Scholieren (LAKS)	Dia
OCW afdeling opo	Inspectie van het Onderwijs	Cito BV
OCW afdeling ovo	VO-raad	AMN
CvTE afdeling po	Stichting Cito divisie CTE	
Nederlandse Vereniging van Leraren Maatschappijleer	Scholen	

In de interviews zijn de volgende onderwerpen aan bod gekomen: het doel, definitie en huidig gebruik van een lvs, de wenselijkheid van het invoeren van een verplicht lvs in het vo (voor- en nadelen), de mogelijkheden die een lvs in het vo biedt (welke (deel)vaardigheden moeten opgenomen worden), oneigenlijk gebruik van de lvs-data, en de wenselijkheid van een koppeling tussen leerlingvolgsystemen in het po en vo. Daarnaast konden deelnemers aan de interviews zelf nog relevante punten over de verplichting of het gebruik van een lvs in het vo inbrengen.

De gesprekken konden online of live plaatsvinden. Er is vooraf toestemming gevraagd voor opname van het gesprek, op basis waarvan gespreksverslagen uitgewerkt zijn. De gespreksverslagen zijn voorgelegd aan de geïnterviewden voordat deze verwerkt zijn in deze rapportage.

Naast de interviews met lvs-gebruikers zijn er ongestructureerde interviews gehouden met toetsexperts om de mogelijkheden voor een psychometrische koppeling tussen leerlingvolgsystemen in het po en het vo in kaart te brengen.

Enquête

Voor de beantwoording van de onderzoeksvragen van onderzoeksvraag 4 is een vragenlijst uitgezet onder alle vo-schoolvestigingen in Nederland. De vragenlijst is per mail naar de schoolleiders verstuurd. Vervolgens is tweemaal per mail een rappel verstuurd en is tot slot een telefonisch rappel uitgevoerd. Telefonisch kregen de respondenten de optie aangeboden om samen een verkorte versie van de vragenlijst in te vullen.

Van de 1437 scholen hebben 545 scholen de vragenlijst ingevuld. Het responspercentage bedraagt daarmee 37,9%. In de responsanalyse hebben we gekeken naar verschillende schoolkenmerken, namelijk regio, denominatie en onderwijsaanbod. Uit onderstaande tabellen blijkt dat de afwijkingen in respons voor regio en denominatie tamelijk beperkt zijn. Voor deze kenmerken hebben we daarom niet gewogen in de analyses. Wat betreft het onderwijsaanbod zien we dat vestigingen die brugjaar of vmbo aanbieden iets ondervertegenwoordigd zijn en vestigingen die praktijkonderwijs of havo/vwo aanbieden iets oververtegenwoordigd. Omdat dit kenmerk vermoedelijk van invloed kan zijn op de (manier van) inzet van een lvs, hebben we gekozen om hiervoor een weegfactor toe te passen.

Tabel B1.2 Respons scholen naar regio

	Populatie	Respons
Noord	39,7%	39,5%
Midden	37,5%	38,0%
Zuid	22,8%	22,5%

Tabel B1.3 Respons scholen naar denominatie

	Populatie	Respons
Bijzonder	70,1%	71,6%
Openbaar	28,5%	27,5%
Overige	1,4%	0,9%

Tabel B1.4 Respons scholen naar onderwijsaanbod

	Populatie	Respons	Weegfactor
Brugjaar	4,2%	3,1%	1,33
Praktijk	9,7%	12,4%	0,79
VMBO	36,6%	33,0%	1,11
VMBO/HAVO	4,1%	4,8%	0,86
HAVO/VWO	12,9%	14,0%	0,92
Breed	32,5%	32,7%	1,00

Literatuurstudie

Voor de beantwoording van de onderzoeksvragen van onderzoeksvraag 1 (optionele deel) en 6 is gebruik gemaakt van bureauonderzoek (literatuurstudie).

De geraadpleegde literatuur voor het paragraaf 'Praktische vaardigheden' is verzameld door middel van een combinatie van I) gerichte zoekopdrachten ("practical skills" AND "assessment", "praktische vaardigheden", "praktische vaardigheden" AND "toetsing", "praktische vaardigheden" AND "toetsen") in Google Scholar en het Education Resources Information Center (ERIC) en II) de sneeuwbal methode waarmee de literatuurlijsten van reeds ontsloten relevante bronnen werden bestudeerd. Deze beperkte literatuurstudie betrof een kwalitatieve verkenning naar manieren waarop praktische vaardigheden binnen de onderwijscontext worden geconceptualiseerd en gemeten en was gedeeltelijk onderdeel van een promotieonderzoek naar praktische vaardigheden meten.

De literatuurstudie binnen werkpakket 6 bestond uit het verzamelen van achtergrondinformatie over het huidige gebruik van lvs'en en oneigenlijk gebruik. Hiervoor is ook informatie in het primair onderwijs verzameld. Voor het huidige gebruik van lvs'en is bijvoorbeeld de website van de Rijksoverheid geraadpleegd. Verder zijn geraadpleegde bronnen de kennisbank van onderwijskennis.nl (NRO), Didactief en SLO.

Focusgroepen

Om zicht te krijgen op de meetbaarheid van praktische vaardigheden heeft een focusgroep plaatsgevonden over dit onderwerp. Bij deze focusgroep waren toetsconstructeurs en experts op het gebied van CSPE's aanwezig. Gespreksonderwerpen hierbij waren: definitie van praktische vaardigheden, concretisering van praktische vaardigheden en het beoordelen van praktische vaardigheden.

Daarnaast hebben er twee focusgroepen plaatsgevonden om het oneigenlijk gebruik van het lvs vanuit het perspectief van docenten en leerlingen in kaart te brengen. Aan docenten hebben we onder andere gevraagd op welke wijze het lvs gebruikt wordt op school, hoe het gebruik van een lvs de motivatie van leerlingen en docenten beïnvloedt, welke risico's er zijn op oneigenlijk gebruik van een lvs en wat daar tegen gedaan kan worden. Aan leerlingen hebben we bijvoorbeeld ook gevraagd hoe het huidige gebruik van het lvs op hun school er uitziet en wat zij als voor- en nadelen zien van het afnemen van de lvs-toetsen.

Survey met vignetten

In de laatste fase van onze ontwerpstudie hebben we de respondenten die in de enquête hebben aangegeven (najaar 2024) dat we ze opnieuw mochten benaderen in het kader van deze ontwerpstudie, uitgenodigd om een survey in te vullen. Het doel van deze survey is om bij schoolleiders en vakdocenten te inventariseren wat hun ervaringen zijn met het gebruik van de huidige lvs'en, wat daarbij de problemen de problemen op hun school zijn en hoe zij aankijken tegen het mogelijk verplichte gebruik door scholen.

We hebben in februari 2025 geïnteresseerde schoolleiders via een mail benaderd en gevraagd de anonieme link door te sturen naar andere schoolleiders en vakdocenten. Na drie weken en een reminder is de survey gesloten. In totaal hebben 25 personen (74%) een start gemaakt met het invullen van de Qualtrics survey. Vier respondenten zijn verwijderd omdat zij een ander lvs gebruiken dan dat we in deze ontwerpstudie beschouwen of wanneer zij aan een leerlingadministratiesysteem refereerden (Managebac, Magister, TNT en onbekend). Vervolgens hebben we nog vijf respondenten verwijderd omdat zij gestopt zijn direct na het invullen van de startvraag naar hun functie. Dit resulteert in 16 bruikbare respondenten (47%) die als volgt zijn verdeeld over de lvs-aanbieders:

- Dia toetsen (n=6),
- Leerling in beeld (n=4),
- Jij! (n=4), of
- geen enkele lvs voor de onderbouw vo (n=2).

Om een concreet, realistisch en breed beeld te verkrijgen van de mogelijke bezwaren tegen het gebruik (casus 1), het moeite hebben met interpretatie en implementatie van de lvs-resultaten (casus 2) hebben we twee vignetten met open toelichtingsvragen toegevoegd aan de survey. De vignetten hebben de volgende probleemomschrijvingen:

Casus 1: weerstand tegen gebruik leerlingvolgtoetsen

Op school A maken meerdere vaksectieleiders bezwaar tegen het gebruik van het leerlingvolgsysteem. Ze willen het niet langer inzetten, want het is toch niet wettelijk verplicht. Ze vinden hun eigen toetsen en de methodetoetsen goed genoeg om de leervorderingen van hun leerlingen te kunnen monitoren. Als argumenten brengen zij naar voren dat de volgtoetsen Nederlands, Engels en rekenen-wiskunde voor de onderbouw vo:

- zorgen voor een te hoge toetsdruk voor zowel docenten als leerlingen;
- door veel leerlingen niet serieus worden ingevuld;
- een te beperkt en wisselend beeld geven van de leervorderingen van leerlingen;
- te weinig opleveren voor de inrichting van het onderwijs;
- door docenten vooral gezien worden als een afvinklijstje.

Gelden genoemde argumenten tegen het gebruik van volgtoetsen Nederlands, Engels, rekenen-wiskunde ook voor uw school?

Casus 2: moeite met interpretatie volgtoetsen en onderwijsaanpassingen

Op school B zijn onlangs de toetsen uit het leerlingvolgsysteem afgenomen bij alle leerlingen in leerjaar 2. In een vaksectieoverleg worden de resultaten besproken. Diverse docenten geven aan moeite te hebben met het begrijpen, analyseren en interpreteren van het functioneringsniveau en de groei in de rapportages. Ook vinden veel vakdocenten het lastig om op gepaste wijze aan de slag te gaan met onderwijsaanpassingen op basis van de rapportages op leerling- en klasniveau.

De redenen die zij opgeven zijn:

- de uitleg en voorbeelden in de docenthandleiding zijn lastig te begrijpen;
- het diepgaand analyseren, vergelijken en interpreteren van de resultaten met de behulp van de diverse rapportages is moeilijk;
- de rapportage over deelaspecten van leerdomeinen toont een te wisselend beeld om echt mee aan de slag te kunnen gaan;
- het beeld van veel leerlingen in de rapportages komt niet overeen met de eigen waarneming in de klas;
- de voorgestelde verbetering van het onderwijsaanbod in de docenthandleiding is onvoldoende handelingsgericht.

Gelden genoemde redenen ook voor uw school?

Gesprekken en documentenanalyse toetsaanbieders

Ten behoeve van de inhoudelijke inventarisatie van de leerlingvolgsystemen (zie paragraaf 3.4 Aanbod lvs) hebben we een gedetailleerd overzicht (Excel) gemaakt van de definitieve conceptkerndoelen voor Nederlands, Engels en rekenen en wiskunde⁶⁸ en de niveaubeschrijvingen 1F en 2F van het referentiekader taal en rekenen.⁶⁹ Op basis van de ontvangen en openbaar beschikbare informatie van de vier toetsaanbieders (AMN, Cito BV, Diataal en ICE) hebben we gecodeerd in welke mate de kerndoelen met bijbehorende doelzin en puntgewijze uitwerkingen van de doelzinnen worden gedekt door de huidige volgtoetsen voor leerlingen in onderbouw van het voortgezet onderwijs. We maakten daarbij onderscheidt in de codering 'ja', 'deels' en 'nee'. Dit hebben we ook gedaan voor het referentiekader taal en rekenen. De uitsplitsingen naar (hoofd)domeinen, niveau 1F en 2F, en bijbehorende subdoelomschrijvingen zijn vergeleken met de inhoudelijke informatie van de vier aanbieders en gecodeerd op mate van dekking. De middencategorie 'deels' is divers: uiteenlopend van bijna niet tot grotendeels wel gedekt.

De toegestuurde en beschikbare informatie was echter globaal van aard waardoor de compleetheid en juistheid van de codering werd bemoeilijkt. Om dit probleem op te lossen hebben we het conceptschema met iedere aanbieder afzonderlijk besproken (Teams-overleggen in januari en februari 2025). Iedere toetsaanbieder heeft zijn schema geverifieerd, aangepast en retour gestuurd. Met deze feedback hebben we het conceptschema uitgewerkt tot een definitief schema waarin de eerdere hiaten of onduidelijkheden zijn opgelost.

Naast de verificatie van het conceptschema hebben we met de individuele toetsaanbieders ook gesproken over de redenen waarom zij hun volgtoetsen in meer of mindere mate baseren op de definitieve conceptkerndoelen van de SLO en/of het referentiekader taal en rekenen en de ERK referentieniveaus voor Engels. Deze informatie geldt als toelichting op de gevonden resultaten over de dekkingsgraad en is samengevat in de paragrafen motivatie voor dekking van kerndoelen en referentiekader (3.4.4 en 3.4.6).

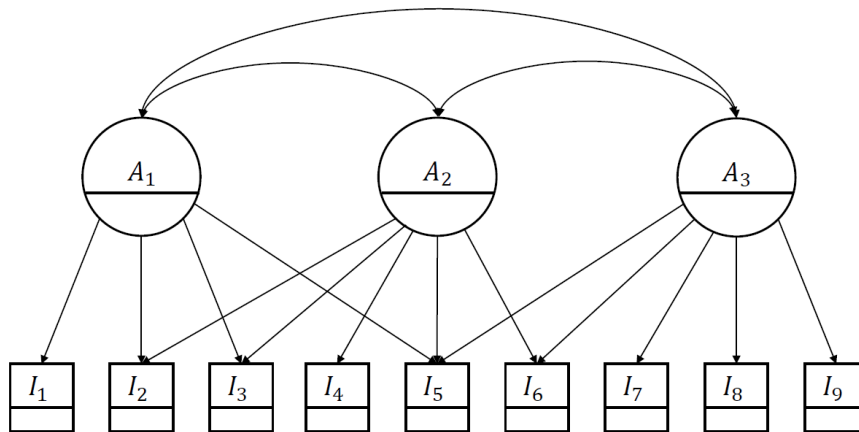
Technische achtergrond Cognitive Diagnostic Models

Cognitief-diagnostische modellen (CDM's) leiden tot probabilistische classificaties van respondenten voor een set discrete latente variabelen, de zogenoemde attributen. Figuur A.1 toont een grafische weergave van een CDM met drie attributen die worden gemeten door negen items. Volgens de standaardconventies voor dergelijke weergaven representeren de cirkels latente variabelen (de attributen *A*), de vierkanten geobserveerde variabelen (de items *I*), en horizontale strepen binnen cirkels en vierkanten dichotome variabelen. De pijlen lopen van de latente variabelen naar de itemresponsvariabelen, gezien aangenomen wordt dat variantie in de latente predictorvariabelen leidt tot variantie in de responsvariabelen. CDM's zijn confirmatieve latente klassemodellen, in de zin dat domeinexperts vooraf specificeren welke attributen door welke items gemeten worden. Dit wordt vastgelegd in een zogenoemde Q-matrix (Tatsuoka, 1983). Zoals te zien is in Figuur B1.1 kunnen items op meerdere attributen laden, wat weerspiegelt dat een item een responsproces activeert waarbij meerdere attributen een rol spelen.

⁶⁸ SLO, 2023a, 2023b en 2024

⁶⁹ Expertgroep doorlopende leerlijnen taal en rekenen, 2008, 2009

Figuur B1.1 Grafische weergave van een CDM met drie attributen en negen items



Algemene vorm van CDM's

Een CDM bestaat uit een structureel model en een meetmodel. Het structurele model beschrijft de samenhang tussen de latente attributen en het meetmodel legt de koppeling tussen de latente variabelen en de geobserveerde item responsdata. Stel we hebben een toets met I items die A binaire attributen meet, en x_{ri} is de dichotoom gescoorde respons van respondent r op item i , die gelijk is aan 0 of 1 voor een respectievelijk incorrect of correct antwoord. De wiskundige representatie van de kans op een geobserveerd responspatroon $\mathbf{x}_r = [x_{r1}, \dots, x_{rI}]$ is dan:

$$P(\mathbf{x}_r) = \underbrace{\sum_{c=1}^{2^A} \nu_c}_{\text{Structureel model}} \underbrace{\prod_{i=1}^I \pi_{ci}^{x_{ri}} (1 - \pi_{ci})^{1-x_{ri}}}_{\text{Meetmodel}}$$

Hier staat ν_c voor de proportionele grootte van latente klasse c , en π_{ci} voor de kans dat een respondent in latente klasse c een correct antwoord geeft op item i . De latente klasse c is een attribuutprofiel, welke wordt aangegeven met de vector $\alpha_c = [\alpha_{c1}, \dots, \alpha_{cA}]$, waarbij $\alpha_{ca} = 1$ als attribuut a beheerst wordt en $\alpha_{ca} = 0$ als dat niet zo is.

Het structurele model

Voor een toets met A attributen, zijn er $2^A - 1$ structurele componenten (met de restrictie $\sum_{c=1}^{2^A} \nu_c = 1$). Om het aantal structurele parameters te verminderen, kan de component ν_c verder gemodelleerd worden in verschillende vormen; zie Rupp et al. (2010, Ch. 8) voor een overzicht van zulke structurele modellen. Het toepassen van structurele modellen leidt tot een minder complexe parametrisatie van de latente attribuutruimte.

Het meetmodel

Verscheidende CDM's parametriseren de itemresponskansen π_{ci} op verscheidende manieren. Dit weerspiegelt de verscheidende mechanismen die ten grondslag liggen aan het itemresponsgedrag (bijvoorbeeld of gebrek aan beheersing van een attribuut kan worden gecompenseerd door beheersing van een ander attribuut). Er zijn vele CDM's die afzonderlijk van elkaar ontwikkeld zijn. Echter zijn er ook algemene raamwerken, zoals het *log-linear cognitive diagnosis model* (LCDM; Rupp et al., 2010), het *general diagnostic model* (GDM; von Davier, 2008) en het *generalized deterministic inputs, noisy "and" gate model* (G-DINA; de la Torre, 2011). Deze raamwerken omvatten meerdere specifiekere CDM's, die verkregen kunnen worden door statistische restricties op model parameters te zetten.

Het LCDM modelleert de conditionele kans π_{ic} dat een respondent met attribuutprofiel α_c een correct antwoord geeft op item i als volgt:

$$\pi_{ic} = P(X_{ic} = 1 \mid \alpha_c) = \frac{\exp(\lambda_{i,0} + \lambda_i^T h(\alpha_c, \mathbf{q}_i))}{1 + \exp(\lambda_{i,0} + \lambda_i^T h(\alpha_c, \mathbf{q}_i))}$$

Hier is $\mathbf{q}_i = [q_{i1}, \dots, q_{iA}]$ is de vector van binaire Q-matrix entries voor item i , die aangeeft welke attributen het item meet. De intercept parameter $\lambda_{i,0}$ is de logit (log-odds) van een correct antwoord gegeven dat geen enkel attribuut beheerst wordt door een respondent. De vector λ_i met lengte $2^A - 1$ bevat de parameters met hoofdeffecten en interactie-effecten voor item i , en $h(\alpha_c, \mathbf{q}_i)$ is een vector van lengte $2^A - 1$ met lineaire combinaties van α_c en $\mathbf{q}_i$. De itemparameters, attribuutprofielen van de respondenten en de Q-matrix entries worden in de exponent als volgt gecombineerd:

$$\lambda_{i,0} + \lambda_i^T h(\alpha_c, \mathbf{q}_i) = \lambda_{i,0} + \sum_{a=1}^A \lambda_{i,1,(a)} \alpha_{ca} q_{ia} + \sum_{a=1}^{A-1} \sum_{a'=a+1}^A \lambda_{i,2,(a,a')} \alpha_{ca} \alpha_{ca'} q_{ia} q_{ia'} + \dots$$

Voor item i bevat de exponent een intercept ($\lambda_{i,0}$), alle hoofdeffecten (bijv. $\lambda_{i,1,(a)}$ is de toename in de logit van een correct antwoord gegeven dat attribuut a beheerst wordt) en alle mogelijke (tweeweg tot A -weg) interacties tussen attributen (bijv. $\lambda_{i,2,(a,a')}$ is de tweeweginteractie tussen attributen a en a' , waardoor de logit van een correct antwoord kan veranderen bij beheersing van beide attributen). Er liggen monotoniciteitsrestricties op de parameters, zodat de itemresponskansen monotoon toenemen als het aantal beheerste attributen toeneemt.

Om van het algemene LCDM naar een specifiekere, simpeler CDM te gaan, kunnen bepaalde parameters in de vector λ_i in de itemresponsfunctie op nul worden gefixeerd. Het LCDM is de gesatureerde versie van veel verscheidende CDM's (zie Rupp et al., 2010, Ch. 7). De keuze voor een specifiek CDM kan worden gebaseerd op verscheidende overwegingen, zoals inhoudelijke overwegingen, modelcomplexiteit, eenvoud van interpretatie of modelfit (Ma, Iaconangelo & de la Torre, 2016). Simpelere CDM's, zoals het deterministic inputs, noisy "and" gate model (DINA; de la Torre, 2009), worden vaak geselecteerd vanwege hun eenvoudige schatting en interpretatie. Complexere modellen worden gekozen vanwege hun algemene toepasbaarheid en mogelijkheden tot modelverfijning. Bij het toepassen van psychometrische modellen moet er bij de selectie van een model een balans gevonden worden tussen statistische fit, meeteigenschappen en wensen van stakeholders (Ravand & Baghaei, 2020).

Attribuutprofielen schatten

CDM's schatten de kans dat een respondent in een bepaalde latente diagnostische klasse valt gegeven de itemresponsen. Dit wordt gedaan in een Bayesiaans raamwerk. De posterior kans dat respondent r in latente klasse c valt, ofwel attribuutprofiel α_c heeft, wordt gegeven door:

$$P(\alpha_c | \mathbf{x}_r) = \frac{P(\alpha_c) \cdot P(\mathbf{x}_r | \alpha_c)}{P(\mathbf{x}_r)} = \frac{\nu_c \prod_{i=1}^I \pi_{ci}^{x_{ri}} (1 - \pi_{ci})^{1-x_{ri}}}{\sum_{c=1}^C \nu_c \prod_{i=1}^I \pi_{ci}^{x_{ri}} (1 - \pi_{ci})^{1-x_{ri}}}$$

Classificaties kunnen worden gebaseerd op maximum a posteriori (MAP) of expected a posteriori (EAP) schattingen van de posterior verdeling. MAP gebruikt posterior kansen per profiel, terwijl EAP kansen per afzonderlijk attribuut schat. De EAP-schattingen zijn eenvoudiger te interpreteren en leiden tot hogere classificatie accuraatheid voor de individuele attributen (Huebner & Wang, 2011). Om marginale attribuut informatie te vinden, wordt de verwachtingswaarde per attribuut a over alle latente klassen berekend:

$$P(\alpha_{ra} = 1 | \alpha_r) = \sum_{c=1}^C \alpha_{rc} \alpha_{ca}$$

Als deze kans hoger is dan een bepaalde drempelwaarde, wordt gesteld dat de respondent het attribuut beheerst. Deze drempelwaarde wordt meestal op 0,5 gezet, zodat posterior kansen op beheersing groter dan 0,5 tot een classificatie als 'beheerst' leiden. Als misclassificaties in de ene richting zwaarder wegen dan in de andere richting, kan men voor een andere drempelwaarde kiezen (Bradshaw & Levy, 2019).

CDM-uitbreidingen

Polytome attributen

Bij de meeste CDM's zijn de attributen dichotoom: een respondent beheerst een attribuut wel of niet. Echter zijn er ook modellen waarin de attributen polytome gedefinieerd zijn, d.w.z. met meer dan twee beheersingsniveaus. Deze niveaus worden gerepresenteerd door gerangschikte gehele getallen, waarbij lagere getallen staan voor lagere beheersingsniveaus (bijv. $\alpha_{ca} \in \{0,1,2\}$ bij drie beheersingsniveaus).

De beheersingsniveaus kunnen ofwel inhoudelijk gedefinieerd worden voorafgaand aan de data-analyse, zoals in het *polytomous generalized deterministic inputs, noisy, "and" gate model* (pG-DINA; J. Chen & de la Torre, 2013), ofwel afgeleid worden na het modelleren van de data, zoals in het *polytomous diagnostic classification model* (PDCM; Bao, 2019). In het eerste geval spreken we over expert-gedefinieerde attributen. Zo worden bij het pG-DINA model items geconstrueerd die gericht zijn op het meten van een specifiek beheersingsniveau en dit beoogde niveau wordt vastgelegd in de Q-matrix, zodanig dat de entries q_{ia} niet binair zijn maar gehele getallen die passen bij de beheersingsniveaus (bijv. $q_{ia} \in \{0,1,2\}$ bij drie beheersingsniveaus). In het tweede geval spreken we over data-gedefinieerde attributen. Zo worden met het PDCM twee hoofdeffecten per attribuut geschat. Bijvoorbeeld, voor een unidimensioneel item i dat één attribuut a met drie beheersingsniveaus meet, ziet de itemresponskans voor een respondent in latente klasse c er als volgt uit:

$$\pi_{ic} = P(X_{ic} = 1 | \alpha_c) = \frac{\exp(\lambda_{i,0} + \lambda_{i,1,(a)}^1 \alpha_{ca}^1 + \lambda_{i,1,(a)}^2 \alpha_{ca}^2)}{1 + \exp(\lambda_{i,0} + \lambda_{i,1,(a)}^1 \alpha_{ca}^1 + \lambda_{i,1,(a)}^2 \alpha_{ca}^2)}$$

Hier zijn α_{ca}^1 en α_{ca}^2 dummy variabelen voor attribuut a (d.w.z. $\alpha_{ca} = \alpha_{ca}^1 + \alpha_{ca}^2$); $\lambda_{i,0}$ is de intercept en de logit (log-odds) van een correct antwoord voor degenen met het laagste beheersingsniveau ($\alpha_{ca}^1 = \alpha_{ca}^2 = 0$); $\lambda_{i,1,(a)}^1$ is het hoofdeffect voor het middelste beheersingsniveau ($\alpha_{ca}^1 = 1, \alpha_{ca}^2 = 0$), wat de toename in de itemresponskans bij het halen van dit beheersingsniveau weerspiegelt; en $\lambda_{i,1,(a)}^2$ is het hoofdeffect voor het hoogste beheersingsniveau ($\alpha_{ca}^1 = \alpha_{ca}^2 = 1$), wat de additionele toename in de itemresponskans weergeeft bij het behalen van het hoogste beheersingsniveau. Er liggen monotoniciteitsrestricties op de itemparameters, zodat de itemresponskansen monotoon toenemen als het beheersingsniveau toeneemt. Voor de algemene vorm van het PDCM verwijzen we naar Bao (2019).

Longitudinale CDM's

In een CDM-raamwerk vertaalt het meten van groei zich naar het meten van veranderingen in beheersingsstatus over tijd. Op individueel niveau gaat dit om een transitie van het niet beheersen van een attribuut naar het wel beheersen van een attribuut; op groepsniveau gaat het om de toename in het percentage respondenten dat een attribuut beheerst en welke attributen moeilijk zijn om te leren. Hiervoor hebben Madison en Bradshaw (2018) *transition diagnostic classification model* (TDCM) ontwikkeld. Hierin is het meetmodel een combinatie van het LCDM en latente transitie-analyse (LTA; Collins & Wugalter, 1992). Met het TDCM worden transitiekansen geschat, die de kans representeren dat een respondent wisselt van attribuutstatus tussen twee meetmomenten. Voor een respondent r die I items beantwoordt op twee meetmomenten wordt de kans op responspatroon $\mathbf{x}_r$ gegeven door:

$$P(\mathbf{X}_r = \mathbf{x}_r) = \sum_{c_1=1}^C \sum_{c_2=1}^C v_{c_1} \tau_{c_2|c_1} \prod_{t=1}^2 \prod_{i=1}^I \pi_{ic_t}^{x_{rit}} (1 - \pi_{ic_t})^{1-x_{rit}}$$

Hier is v_{c_t} de proportionele grootte van latente klasse c op meetmoment t . Er wordt gesommeerd over alle C latente klassen bij elk meetmoment; de producten gaan over beide meetmomenten en over de I items. De transitiekans $\tau_{c_2|c_1}$ is de kans om te wisselen tussen attribuutstatussen tussen meetmoment $t = 1$ en meetmoment $t = 2$; de π_{ic_t} 's zijn de itemresponskansen die worden geschat met het LCDM; en x_{rit} is het antwoord van respondent r op item i op meetmoment t . Voor de algemene vorm van het TDCM met meer meetmomenten, verwijzen we naar Madison en Bradshaw (2018).

Een belangrijk aspect in longitudinale psychometrische modellen is meetinvariantie. Dit betreft de mate waarin de betekenis van de gemeten vaardigheid en de relatie met itemresponsen consistent blijft over tijd. Als er meetinvariantie is, kunnen veranderingen in de prestaties van respondenten over tijd worden toegeschreven aan de veranderingen in vaardigheid. Formeel is er sprake van meetinvariantie als de verdeling van itemresponsen conditioneel op de beheersingsstatus niet verandert over tijd, ofwel als voor alle meetmomenten $t = 1, \dots, T$ (waarop dezelfde items afgenomen worden) geldt:

$$P(\mathbf{X}_{r1} = \mathbf{x}_{r1} | \alpha_c) = P(\mathbf{X}_{r2} = \mathbf{x}_{r2} | \alpha_c) = \dots = P(\mathbf{X}_{rT} = \mathbf{x}_{rT} | \alpha_c)$$

Hoewel CDM-resultaten in theorie item-invariant zijn (Bradshaw & Madison, 2016), is het zonder ankeritems heel moeilijk om af te dwingen dat de betekenis van beheersing hetzelfde is bij verschillende toetsen. Door ankeritems te gebruiken en de parameters hiervan gelijk te stellen over tijd, wordt ervoor gezorgd dat de beheersing van attributen dezelfde betekenis houdt op verschillende meetmomenten. Of deze parameterrestricties legitiem zijn, ofwel of er sprake is van meetinvariantie, kan getoetst worden door het model te vergelijken met een model waarin de parameters vrij geschat worden (bijvoorbeeld met een likelihood-ratiotoets).

Inhoudelijke dekking referentiekader taal en rekenen met doelomschrijvingen

Tabel B2.1 Overzicht van de mate waarin onderwerpen binnen (sub)domeinen het referentiekader taal inhoudelijke worden gedekt door de vier lvs-aanbieders

Onderwerp	Niveau	Doelomschrijving: De leerling ...	Dekking van referentieniveaus door aanbieders		
			wel	deels	niet
Domein: Mondelinge Taalvaardigheid					
Subdomein: Gesprekken					
Taken - Deelnemen aan discussie en overleg	1F en 2F	... kan bespreken wat er gedaan moet worden en bijdragen aan de planning. ... kan tijdens een discussie of overleg (op beleefde wijze) een probleem verhelderen, een overtuiging of mening, instemming of afkeuring uitdrukken en commentaar geven op de visie van anderen.	1	0	3
Taken - Informatie uitwisselen	1F en 2F	... kan informatie vragen en geven aan instanties binnen en buiten school. ... kan informatie verzamelen en verwerken via het houden van een vraaggesprek.	1	0	3
Subdomein: Luisteren					
Luisteren naar instructies	1F	... kan voldoende feitelijke informatie halen uit instructies en aanwijzingen om taken of handelingen uit te kunnen voeren.	2	0	2
	2F	... kan uitleg en instructies over concrete onderwerpen begrijpen.	2	0	2
Luisteren als lid van een live publiek	1F	... kan de hoofdlijn begrijpen van korte informatieve, instructieve en betogende teksten met een duidelijke structuur en voldoende herhaling. ... kan een eenvoudig, voorgelezen of verteld verhaal begrijpen.	2	0	2
	2F	... kan een helder gestructureerde voordracht, toespraak of les begrijpen over vertrouwde onderwerpen binnen het eigen vak- of interessegebied. ... kan een voorgelezen of verteld verhaal begrijpen.	2	0	2

Onderwerp	Niveau	Doelomschrijving: De leerling ...	Dekking van referentieniveaus door aanbieders		
			wel	deels	niet
Luisteren naar radio en televisie en naar gesproken tekst op internet	1F	... kan hoofdpunten van korte en duidelijke berichten op radio en televisie en via internet begrijpen als onderwerp en context bekend zijn. Kan selectief luisteren om de benodigde informatie zoals openingstijden, adres, telefoonnummer te halen uit korte berichten.	2	0	2
	2F	... kan de hoofdpunten begrijpen van (nieuws) berichten, documentaires, reclameboodschappen en discussieprogramma's over vertrouwde onderwerpen. ... kan films en televisieseries geschikt voor de eigen leeftijd volgen.	2	0	2
Subdomein: Spreken					
Een monoloog houden	1F	... kan alledaagse aspecten beschrijven, zoals mensen, plaatsen en zaken. ... kan verslag uitbrengen van gebeurtenissen, activiteiten en persoonlijke ervaringen. ... kan een kort, voorbereid verhaal of presentatie houden en daarbij op eenvoudige vragen reageren.	2	0	2
	2F	... kan in grote lijnen redenen en verklaringen geven voor eigen meningen, plannen en handelingen en kan een kort verhaal vertellen. ... kan informatie verzamelen om over een onderwerp uit eigen interessegebied een voorbereide presentatie te geven. ... kan vragen beantwoorden naar aanleiding van deze presentatie	2	0	2
Domein: Lezen					
Subdomein: Zakelijke Teksten (informatieve teksten, instructies en betogende teksten)					
Vloeiend lezen en woordenschat	1F	... kan teksten zodanig vloeiend lezen dat woordherkenning tekstbegrip niet in de weg staat. ... kent de meest alledaagse (frequente) woorden, of kan de betekenis van een enkel onbekend woord uit de context afleiden.	3	0	1
	2F	<i>Op dit niveau is de woordenschat geen onderscheidend kenmerk van leerlingen meer.</i>	--	--	--

Onderwerp	Niveau	Doelomschrijving: De leerling ...	Dekking van referentieniveaus door aanbieders		
			wel	deels	niet
Lezen en begrijpen van zakelijke teksten	1F	... herkent specifieke informatie, wanneer naar één expliciet genoemde informatie-eenheid gevraagd wordt (letterlijk begrip). ... kan (in het kader van het leesdoel) belangrijke informatie uit de tekst halen en kan zijn manier van lezen daar op afstemmen (bijvoorbeeld globaal, precies, selectief/gericht).	4	0	0
	2F	... kan de hoofdgedachte van de tekst weergeven en maakt onderscheid tussen hoofd- en bijzaken. ... legt relaties tussen tekstdelen (inleiding, kern, slot) en teksten. ... ordent informatie (bijvoorbeeld op basis van signaalwoorden) voor een beter begrip. ... herkent beeldspraak (letterlijk en figuurlijk taalgebruik).	4	0	0
Lezen en interpreteren van zakelijke teksten	1F	... kan informatie en meningen interpreteren voor zover deze dicht bij de leerling staan.	4	0	0
	2F	... legt relaties tussen tekstuele informatie en meer algemene kennis. ... kan de bedoeling van tekstgedeeltes en/of specifieke formuleringen duiden. ... kan de bedoeling van de schrijver verwoorden.	4	0	0
Lezen en evalueren van zakelijke teksten	1F	... kan een oordeel over een tekst(deel) verwoorden.	3	1	0
	2F	... kan relaties tussen en binnen teksten evalueren en beoordelen.	3	1	0
Lezen en samenvatten van zakelijke teksten	1F	<i>Op dit niveau is samenvatten niet van toepassing.</i>	--	--	--
	2F	... kan een eenvoudige tekst beknopt samenvatten.	3	0	1
Opzoeken en lezen van informatie	1F	... kan informatie opzoeken in duidelijk geordende naslagwerken, zoals woordenboeken, telefoongids e.d. ... kan schematische informatie lezen en relaties met de tekst expliciteren.	2	1	1
	2F	... kan systematisch informatie zoeken (op bijvoorbeeld het internet of de schoolbibliotheek) bijvoorbeeld op basis van trefwoorden.	2	0	2

Onderwerp	Niveau	Doelomschrijving: De leerling ...	Dekking van referentieniveaus door aanbieders		
			wel	deels	niet
Subdomein: Fictionele, narratieve en literaire teksten					
Lezen, begrijpen en samenvatten van fictionele, narratieve en literaire teksten	1F en 2F	... herkent basale structurelementen, zoals wisselingen van tijd en plaats, rijm en versvorm. ... kan meeleven met een personage en uitleggen hoe een personage zich voelt. ... kan gedichten en verhaalfragmenten parafraseren of samenvatten.	1	0	3
Lezen en evalueren van fictionele, narratieve en literaire teksten	1F en 2F	... kan relaties leggen tussen de tekst en de werkelijkheid. ... kan spannende, humoristische of dramatische passages in de tekst aanwijzen. ... herkent verschillende emoties in de tekst, zoals verdriet, boosheid en blijdschap.	1	0	3
Domein: Schrijven					
Schrijven met samenhang en afstemming op doel en publiek, met aandacht voor (gevarieerd) woordgebruik, correcte spelling, interpunctie, grammatica en leesbaarheid	1F	... kan korte, eenvoudige teksten schrijven over alledaagse onderwerpen of over onderwerpen uit de leefwereld. Het gaat om de taken: 1) correspondentie, 2) formulieren invullen, het schrijven van berichten, advertenties en aantekeningen, 3) het schrijven van verslagen, werkstukken, samenvattingen en artikelen, en 4) vrij schrijven.	2	0	2
	2F	... kan samenhangende teksten schrijven met een eenvoudige, lineaire opbouw, over uiteenlopende vertrouwde onderwerpen uit de (beroeps)opleiding en van maatschappelijke aard. Het gaat om de taken: 1) correspondentie, 2) formulieren invullen, het schrijven van berichten, advertenties en aantekeningen, en 3) het schrijven van verslagen, werkstukken, samenvattingen en artikelen.	2	0	2
Domein: Begrippenlijst en taalverzorging					
Subdomein: Taalverzorging					
Leestekens (interpunctie)	1F	... kan dubbele punt, punt, komma, puntkomma, uitroepetekens, vraagtekens, aanhalingstekens en hoofdletters toepassen.	3	0	1
	2F	... kan trema, accent, hoofdletter bij eigennaam en directe rede toepassen.	2	1	1

Onderwerp	Niveau	Doelomschrijving: De leerling ...	Dekking van referentieniveaus door aanbieders		
			wel	deels	niet
Woordsoorten	1F	... kan zelfstandig naamwoord, werkwoord (klankvast, klankveranderend (zwak, sterk)) en bijvoeglijk naamwoord onderscheiden.	3	0	1
	2F	<i>Op dit niveau is de woordsoortherkenning niet omschreven.</i>	--	--	--
Grammaticale kennis	1F	... kan onderwerp, lijdend voorwerp, hoofdzin, bijzin, gezegde en persoonsvorm herkennen.	3	0	1
	2F	... kan lijdende en bedrijvende vorm, vragende vorm herkennen.	3	0	1
Tekstkennis, stilistiek en semantiek, morfologie, opmaak en klanken	1F en 2F	... herkent standpunt, argument, feit, mening, gesprekvormen, paragraaf in teksten, onderscheidt tekstsoorten en vormen van stilistiek en semantiek, past morfologie, tekstopmaak en articulatie, klemtoon, intonatie, spreekpauze toe.	1	2	1
Spelling	1F en 2F	... kan diverse spellingvormen toepassen: alfabetische, orthografische, en eenvoudige en moeilijke morfologische spelling en kent afbreekregels en spelambigue woorden.	3	0	1
	2F	... kan morfologische spelling op syntactische basis correct toepassen (<i>verandering van persoonsvorm</i>).	3	0	1

Noot. De doelomschrijvingen voor leerlingen in de onderbouw van het voortgezet onderwijs zijn afkomstig van de Expertgroep doorlopende leerlijnen taal en rekenen (2009)

Meetbaarheid van de kerndoelen voor Nederlands, Engels, wiskunde, digitale geletterdheid en burgerschap

Naast een beeld van de vaardigheden die wenselijk zouden zijn om op te nemen in een lvs, is het voor een verplicht lvs ook van belang om te onderzoeken welke vaardigheden betrouwbaar en valide te meten zijn in leerlingvolgtoetsen. Daartoe hebben we de door SLO geformuleerde kerndoelen van de vakken Nederlands, Engels, rekenen-wiskunde en burgerschap en digitale geletterdheid geanalyseerd. Hierbij is nagegaan welke onderwijsdoelen meetbaar of niet meetbaar zijn in een lvs, en in hoeverre de vaardigheden onderliggend aan de doelen overlappen tussen de kernvakken. Hierbij zijn we ervan uitgegaan dat de doelen die in de huidige leerlingvolgsystemen opgenomen zijn (zie paragraaf 3.4) voldoende meetbaar zijn om op te kunnen nemen in een lvs. Binnen het huidige onderzoek is echter geen uitgebreid onderzoek gedaan naar de validiteit en betrouwbaarheid van de items die hierin zijn opgenomen.

In het algemeen geldt voor alle onderwijsdoelen dat ze doorgaans geformuleerd worden in termen van (a) een bepaalde inhoud van het onderwerp, en (b) een beschrijving van wat er met of aan die inhoud moet worden gedaan. Doelstellingen omvatten doorgaans een (groep van) zelfstandige naamwoord(en) - de inhoud van het onderwerp - en een (groep) werkwoord(en) - het cognitieve proces. Bijvoorbeeld, de leerlingen kunnen:

1. de hoofdsteden van alle Europese landen opsommen.
2. uitleggen hoe het watercyclusproces werkt.
3. een wiskundige formule toepassen om een vraagstuk op te lossen.
4. de elementen van een verhaal identificeren en hun onderlinge verbanden analyseren.
5. verschillende argumenten beoordelen en een onderbouwde conclusie trekken.
6. een eigen onderzoeksvraag formuleren en een bijpassend onderzoeksvoorstel opstellen.

In bovenstaande voorbeelden gaat het om verschillende typen kennis waarop een beroep wordt gedaan, maar eveneens om uiteenlopende cognitieve processen. Deze typen kennis en cognitieve processen worden in de herziene taxonomie van Bloom uitgewerkt en aan elkaar gerelateerd.⁷⁰

⁷⁰ Krathwohl, 2002

Wat betreft het type kennis worden in de taxonomie van Bloom de volgende vier onderscheiden:

1. **Feitenkennis:** de basiselementen die leerlingen moeten kennen om bekend te raken met een discipline of om problemen daarin op te lossen. Dit omvat kennis van terminologie en specifieke details.
2. **Conceptuele kennis:** de onderlinge relaties tussen de basiselementen binnen een grotere structuur. Dit omvat kennis van classificaties, van principes en generalisaties, van theorieën, modellen en structuren.
3. **Procedurele kennis:** kennis over toepassing, d.w.z. kennis van onderzoeksmethoden en criteria voor het gebruik van vaardigheden, algoritmen, technieken en methoden. Dit omvat kennis van vakspecifieke vaardigheden en methoden en tevens kennis van criteria om te bepalen wanneer geschikte procedures moeten worden gebruikt.
4. **Metacognitieve kennis:** kennis over kennis. Dit omvat kennis over cognitieve taken, inclusief passende contextuele en voorwaardelijke kennis. Ook strategische kennis en zelfkennis vallen hieronder.

Wat betreft cognitieve processen worden zes typen onderscheiden:

1. **Onthouden:** relevante kennis ophalen uit het langetermijngeheugen. Dit gaat om herkennen en herinneren.
2. **Begrijpen:** de betekenis bepalen van aangeboden informatie. Dit omvat interpreteren, voorbeelden geven, classificeren, samenvatten, conclusies trekken, vergelijken en uitleggen.
3. **Toepassen:** een procedure uitvoeren of gebruiken in een bepaalde situatie.
4. **Analyseren:** materie uitsplitsen naar samenstellende delen en bepalen hoe de delen zich tot elkaar en tot een groter geheel of doel verhouden.
5. **Evalueren:** oordelen vellen op basis van criteria en normen.
6. **Creëren:** elementen samenvoegen om een nieuw, samenhangend geheel te vormen of een origineel product te maken.

De verschillende typen kennis en cognitieve processen worden in de gereviseerde taxonomie gecombineerd in de matrix in Tabel B3.1. In de matrix zijn ook de eerder genoemde algemene voorbeelden van onderwijsdoelen geplaatst. In sommige gevallen zal blijken dat een onderwijsdoel niet eenduidig in een vak kan worden geplaatst. Het zal geregeld voorkomen dat een onderwijsdoel zich qua kennis en/of cognitief proces zich in een grensgebied bevindt. Dat geldt hier voor het zesde onderwijsdoel.

Tabel B3.1 Dimensies op basis van de gereviseerde taxonomie van Bloom⁷¹

	1 Feitenkennis	2 Conceptuele kennis	3 Procedurele kennis	4 Metacognitieve kennis
1 Onthouden	de hoofdsteden van alle Europese landen opsommen			
2 Begrijpen		uitleggen hoe het watercyclusproces werkt		
3 Toepassen			een wiskundige formule toepassen om een vraagstuk op te lossen	
4 Analyseren		de elementen van een verhaal identificeren en hun onderlinge verbanden analyseren		
5 Evalueren		verschillende argumenten beoordelen en een onderbouwde conclusie trekken		
6 Creëren		een eigen onderzoeksvraag formuleren en een bijpassend onderzoeksvoorstel opstellen		

Om zicht te krijgen op de meetbaarheid van de kerndoelen die door SLO geformuleerd zijn, is aan vakexperts op het gebied van Nederlands, wiskunde, Engels, digitale geletterdheid en burgerschap gevraagd om de door SLO geformuleerde kerndoelen voor hun vakgebied te categoriseren volgens de herziene taxonomie van Bloom. Daarnaast werd de vakexperts gevraagd aan te geven in hoeverre de kerndoelen gemakkelijk te operationaliseren (ofwel: vertalen) zijn naar toetsitems die geschikt zijn om te gebruiken in een lvs, bijvoorbeeld omdat ze eenduidig te beoordelen zijn. Op deze manier wordt inzichtelijk in hoeverre de kerndoelen in theorie gedekt kunnen worden door een lvs.

⁷¹ Krathwohl, 2002

Er is geen dwingende reden om een lvs te beperken tot eenvoudig scorebare leerdoelen. Wanneer ook moeilijker scorebare leerdoelen opgenomen worden, lijkt extra werkdruk voor docenten echter onvermijdelijk; ook al is inmiddels mogelijk om ook complexe producten (bijv. essays) geautomatiseerd te laten beoordelen dankzij de snelle ontwikkelingen op het gebied van artificiële intelligentie. Belangrijke beperkingen in dit verband liggen vooralsnog op het vlak van transparantie en privacy. Het is onduidelijk hoe door machine-learning gedreven algoritmes tot een oordeel komen. Daarnaast is evenmin duidelijk of en hoe de privacy van de data in dit verband gewaarborgd kan worden. Om een duidelijk beeld te krijgen van een onderscheid tussen relatief makkelijk te operationaliseren doelen, moeilijker te operationaliseren kerndoelen en doelen die helemaal niet te operationaliseren zijn, zijn de kerndoelen ingedeeld in de volgende drie categorieën:

- mogelijk zonder grote problemen,
- mogelijk maar erg moeilijk,
- niet realistisch.

De herziene taxonomie van Bloom maakt het mogelijk om leerdoelen van verschillende vakgebieden op een vergelijkbare manier te categoriseren. Zodoende kan per vakgebied een overzicht gegeven worden van de cognitieve processen waarop de leerdoelen een beroep doen. Verschillen en overeenkomsten tussen de vakgebieden in dit opzicht worden hiermee in kaart gebracht. Leerdoelen die een beroep doen op complexe cognitieve processen, zoals analyseren, evalueren en creëren, vragen over het algemeen een andere benadering binnen toetsing dan doelen gericht op onthouden, begrijpen en toepassen. De indeling naar cognitieve processen kan dus gezien worden als oplopend van relatief eenvoudig naar meer complex. Onthouden kan beschouwd worden als minder complex dan begrijpen; begrijpen als minder complex dan toepassen, etc. Dit neemt niet weg dat bepaalde onderdelen uit een meer eenvoudige categorie als meer complex beschouwd moeten worden dan onderdelen uit een meer complexe categorie. Zo valt uitleggen in de categorie begrijpen. Dit kan gezien worden als een meer complexe taak dan het uitvoeren van een eenvoudige procedure. Uitleggen hoe een watercyclus werkt kan men als complexer beschouwen dan het uitvoeren van een eenvoudige berekening (bijvoorbeeld het bepalen van de omtrek van een rechthoek). Er is dus in enkele gevallen sprake van enige overlap tussen de diverse categorieën⁷². In de verdere rapportage ligt de nadruk op het onderscheid tussen kerndoelen die een beroep doen op relatief eenvoudige cognitieve processen (onthouden, begrijpen en toepassen) versus de meer geavanceerde processen (analyseren, evalueren en creëren).

Voor de vakexperts op het gebied van Nederlands, Engels en wiskunde bleek het categoriseren op de voorgestelde manier geen wenselijke methodiek. Zij achtten het gehanteerde raamwerk niet passend voor het in kaart brengen van meetbaarheid voor de doelen van hun specifieke vakgebied. Hierover zijn we met de vakexperts in gesprek gegaan. Daarbij gaf de wiskunde-expert bijvoorbeeld aan dat de wiskundige denk- en werkwijzen, zoals modelleren en probleem oplossen, moeilijk te combineren zijn met vaste indelingen als typen kennis of cognitieve processen. Bij dergelijke vaardigheden spelen meerdere cognitieve processen samen, wat integratie van verschillende typen kennis en cognitieve processen en complexiteit vereist. Ook gaven de experts aan dat de mate waarin de kerndoelen geoperationaliseerd kunnen worden, sterk afhankelijk is van de toetsomgeving. Bij een papieren toets is er bijvoorbeeld meer ruimte voor het opschrijven van formules, vergelijkingen of het tekenen van meetkundige figuren. Daarnaast zijn denkwijzen en procedurele stappen vaak moeilijker gestandaardiseerd te meten. Er komt dus ook een afweging bij kijken tussen (theoretische) meetbaarheid en de haalbaarheid van het gestandaardiseerd meten en nakijken.

⁷² Krathwohl, 2002

Voor alle vijf vakgebieden zijn de leerdoelen desondanks gecategoriseerd volgens de herziene taxonomie van Bloom.⁷³ Zie Tabel B3.1 voor enkele voorbeelden. Echter, alleen voor Burgerschap en Digitale Geletterdheid is dit gebeurd in samenspraak met inhoudelijke vakexperts. Voor Nederlands, Engels en wiskunde is de indeling gemaakt door de onderzoekers. Hierbij moeten dus de nodige kanttekeningen geplaatst worden, zoals door de vakexperts benoemd en hierboven beschreven. Daarnaast is voor elk kerndoel uit de vijf vakgebieden ingeschat hoe eenvoudig het te vertalen is naar toetsitems die geschikt zijn voor een leerlingvolgsysteem. Met als doel om een duidelijk beeld te krijgen van welk deel van de leerdoelen zich leent voor een lvs en welk deel van de leerdoelen een beroep doet op meer geavanceerde cognitieve processen (d.w.z. verder dan onthouden, begrijpen of toepassen).

In totaal zijn 316 leerdoelen gecategoriseerd op hun operationaliseerbaarheid op basis van bovenstaande methode. In deze paragraaf worden de belangrijkste bevindingen gerapporteerd. Hierbij zijn aanboddoelen buiten beschouwing gelaten. Dit zijn doelen die aangeven dat van scholen wordt verwacht dat ze bepaalde leerinhouden aanbieden, zonder dat een beheersingsniveau voor leerlingen wordt gespecificeerd. Enkele voorbeelden hiervan zijn:

- De school zorgt voor een rijke taal- en leesomgeving in het Engels. Bijvoorbeeld stimuleren van de leesmotivatie en durf om te spreken en schrijven.
- De school stimuleert een wiskundige attitude bij leerlingen. Bijvoorbeeld stimuleren van een onderzoekende en kritische houding ten aanzien van getallen en andere wiskundige informatie.
- De school zorgt voor een democratische cultuur. Bijvoorbeeld aanbieden van activiteiten die de betekenis van democratische waarden laten zien.

Het aantal leerdoelen dat gecategoriseerd is, verschilt per vakgebied. Zie Tabel B3.2 voor een overzicht.

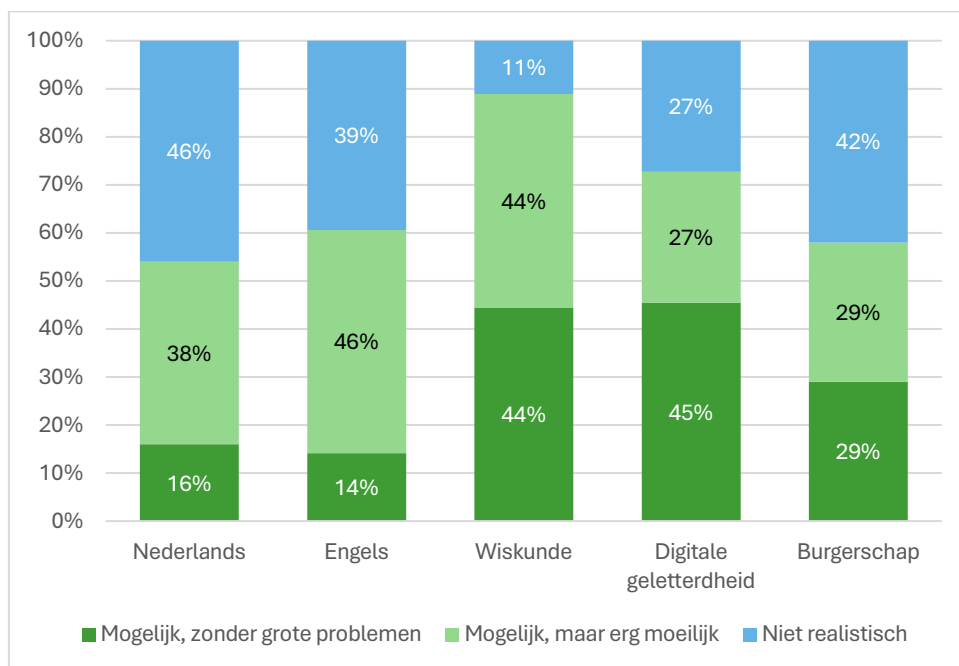
Tabel B3.2 *Leerdoelen per vakgebied*

Vakgebied	Aantal leerdoelen
Nederlands	95
Engels	90
Wiskunde	51
Digitale geletterdheid	55
Burgerschap	24

Het aandeel relatief makkelijk te operationaliseren leerdoelen varieert per vakgebied. Maar voor elk vakgebied geldt dat het slechts voor een minderheid van de leerdoelen zonder grote problemen haalbaar lijkt ze te operationaliseren naar toetsitems die geschikt werden geacht voor een leerlingvolgsysteem. Het aantal leerdoelen waarvoor operationalisering naar lvs-items niet realistisch lijkt is relatief groot voor Nederlands, Engels en burgerschap. Voor wiskunde is deze groep leerdoelen het kleinst, maar voor dat vakgebied is de groep leerdoelen die wellicht te operationaliseren zijn, maar alleen met grote moeite, relatief groot. Figuur B3.1 biedt een overzicht van de operationaliseerbaarheid van leerdoelen per vakgebied.

⁷³ Krathwohl, 2002

Figuur B3.1 Operationaliseerbaarheid leerdoelen per vakgebied



Het zijn met name leerdoelen die een beroep doen op cognitieve processen die verdergaan dan reproduceren, begrijpen en toepassen die zich over het algemeen niet goed lijken te lenen voor toetsing met items die snel en gemakkelijk objectief gescoord kunnen worden. Leerdoelen die een beroep doen op het vermogen van leerlingen tot analyseren, evalueren en creëren vragen meestal complexe en uitgebreide opdrachten die van de beoordelaars een aanzienlijke tijdsinvestering vragen.

Ook in dit opzicht zijn duidelijke verschillen te zien tussen de vijf vakgebieden. Voor Nederlands en burgerschap geldt dat een ruime meerderheid van de leerdoelen een beroep doet op het vermogen van de leerling tot analyseren, evalueren en creëren. Bij Engels en digitale geletterdheid is het aantal leerdoelen dat een beroep doet op reproduceren, begrijpen of toepassen ongeveer gelijk aan het aantal leerdoelen dat een beroep doet op analyseren, evalueren en creëren. Bij wiskunde is het aandeel van leerdoelen die een beroep doen op de moeilijker te meten leerprocessen relatief het laagst. Ook hier maken deze leerdoelen nog steeds 40% uit van alle leerdoelen m.b.t. vakgebied. Zie Figuur 4.2 voor een overzicht.

Wat zijn referentieniveaus?

In 2010 is de Wet referentieniveaus Nederlandse taal en rekenen⁷⁴ van kracht geworden. Deze wet geldt voor alle onderwijssectoren en draagt daarmee zorg voor inhoudelijke samenhang tussen deze sectoren. Het rapport van de Expertgroep doorlopende Leerlijnen⁷⁵ is de basis hiervoor en geeft een systematische beschrijving van de kennisdomeinen taal en rekenen. Hierbij wordt aangegeven wat leerlingen op belangrijke momenten in hun schoolloopbaan moeten kennen en kunnen. Een voorbeeld: op referentieniveau 1F, onder zakelijke teksten kan een leerling eenvoudige teksten over alledaagse onderwerpen en over onderwerpen die aansluiten bij de leefwereld lezen. Deze belangrijke momenten noemt de commissie “drempels”. Er zijn vier drempels vastgesteld in een schoolloopbaan van primair onderwijs tot uitstroom naar de arbeidsmarkt of instroom in het hoger onderwijs. Het niveau dat leerlingen moeten behalen om een drempel goed te kunnen nemen, wordt het fundamentele niveau (F) genoemd. Er zijn ook streefniveaus voor leerlingen die meer aan kunnen (S). De referentieniveaus die gelden voor de verschillende onderwijssoorten zijn te vinden in Tabel B4.1.

Tabel B4.1 Referentieniveaus per onderwijssoort

Onderwijssoort	Lezen	Rekenen	Taalverzorging
primair onderwijs	1F en 2F	1F en 1S	1F en 2F
speciaal onderwijs	1F en 2F	1F en 1S	1F en 2F
praktijkonderwijs	1F	1F	1F
vmbo bb	2F	2F	n.v.t.
vmbo kb	2F	2F	n.v.t.
vmbo-gt/tl	2F	2F	n.v.t.
havo	3F	3F	n.v.t.
vwo	4F	3F	n.v.t.
mbo niveau 1	2F	2F	n.v.t.
mbo niveau 2	2F	2F	n.v.t.
mbo niveau 3	2F	2F	n.v.t.
mbo niveau 4	3F	3F	n.v.t.

Om te bepalen of leerlingen en studenten het gevraagde referentieniveau beheersen, zijn toetsen nodig. De verplichte centrale toetsen en examens zijn gerelateerd aan deze doorlopende leerlijnen. Bij de eindtoetsen in po wordt sinds 2015 gerapporteerd over de beheersing van de referentieniveaus van lezen, taalverzorging en rekenen. De centrale eindexamens Nederlands in vo zijn sinds 2015 afgestemd op de referentieniveaus. En bij de centrale examens Nederlands in het mbo wordt de beheersing van de referentieniveaus getoetst. Ook de rekentoets vo (2012-2019) en de centrale rekenexamens in het mbo (2013-2019) toetsten de beheersing van de referentieniveaus.

⁷⁴ <https://wetten.overheid.nl/BWBR0027679/2022-08-01>

⁷⁵ Expertgroep doorlopende Leerlijnen, 2008

Voor deze doorlopende toetslijn zijn binnen het project Referentiesets opgaven ontwikkeld die inhoudelijk bij de referentieniveaus passen. De openbare opgaven zijn in 2016 gepubliceerd. Ook is een bijbehorende prestatiestandaard beschreven die aangeeft hoeveel opgaven goed gemaakt moeten worden om aan te tonen dat een leerling een bepaald referentieniveau beheerst. Het is de bedoeling dat leerlingen uit verschillende onderwijssectoren aan dezelfde prestatiestandaarden moeten voldoen, maar dat de opgaven die die niveaus meten om inhoudelijke redenen wel van elkaar verschillen. Dat kan met behulp van de referentiesets.

Een referentieset bestaat uit meerdere opgaven die samen per referentieniveau het inhoudelijk beschreven kennisdomein goed meten. De opgaven hebben hun oorsprong in verschillende onderwijssectoren. Alle opgaven zijn afgenomen in de onderwijssectoren die voor de set relevant zijn.

Er is een prestatiestandaard vastgesteld voor de referentieset. De prestatiestandaard geeft weer hoeveel opgaven van de referentieset een leerling goed moet hebben om een bepaald referentieniveau te halen. Er zijn in totaal zes referentiesets ontwikkeld voor rekenen en taal: rekenen 1F, 2F, 3F; taal (begrijpend lezen) 1F, 2F, 3F. Verder zijn met behulp van aanvullende projecten ankersets rekenen 1S opgeleverd voor het referentieniveau rekenen 1S, evenals ankersets taalverzorging 1F en 2F. Deze ankersets zijn alleen van toepassing op het primair onderwijs. Ze zijn geconstrueerd om tegemoet te komen aan de visie van experts om de betreffende niveaus te operationaliseren in verzamelingen opgaven met bijbehorende prestatiestandaarden. Er vindt bij deze sets geen vergelijking tussen schooltypen plaats. Dit betekent dat het niet nodig was om dezelfde opgaven in verschillende schooltypen af te nemen.

Referentieniveaus taal en rekenen

In Nederland is in de wet vastgelegd dat alle leerlingen op verschillende momenten in hun schoolloopbaan laten zien of zij voldoen aan de 'referentieniveaus'. Deze niveaus zijn voor Nederlandse taal en rekenen inhoudelijk beschreven. In de periode 2011-2014 is voor deze inhoudelijke niveaus een standaard vastgesteld waarmee ongeacht het toetsinstrument, dezelfde 'lat' kan worden gehanteerd voor iedere leerling.

Voor Nederlandse taal zijn drie niveaus beschreven: 1F, 2F en 3F (F staat voor fundamenteel). Voor rekenen zijn vier niveaus bepaald: 1F, 1S (S staat voor streefniveau), 2F en 3F. Dit zijn in beide gevallen oplopende niveaus waarbij 1F het laatste niveau is en 3F het hoogste. Het 2F-niveau wordt zowel gebruikt aan het eind van het basisonderwijs (11 jaar), als aan het eind van het voortgezet onderwijs (16 jaar) en aan het eind van het middelbaar beroepsonderwijs (18 jaar en ouder).

Om de niveaus vast te leggen is per niveau een expertpanel samengesteld met daarin acht tot tien leden. Deze leden waren evenredig verdeeld over de verschillende relevante onderwijssectoren, waren afkomstig uit de wetenschap, uit de praktijk en waren alleen inhoudelijk expert. De expertpanels werden ingezet voor twee activiteiten: het vaststellen van een 'referentieset' en het vaststellen van een inhoudelijke standaard.

Een referentieset bevat een groot aantal opgaven (in dit voorbeeld 130) die gezamenlijk een goede representatie zijn van het inhoudelijke domein. De opgaven zijn afkomstig uit de verschillende onderwijssectoren: basisonderwijs, voortgezet onderwijs en beroepsonderwijs. Zij dekken gezamenlijk alle inhoudelijke aspecten van het domein. Het expertpanel had een rol in de selectie en het vaststellen van deze referentieset.

Vervolgens zijn alle opgaven eenmalig afgenomen in alle onderwijssectoren. De data die hiermee beschikbaar kwamen zijn onder andere gebruikt in de standaardbepalingsprocedure. Voor de standaardbepalingsprocedure is het expertpanel aangevuld met tien leraren uit verschillende sectoren. Het panel heeft vervolgens via een extended Angoff (rekenen) en de 3DC-methode (taal) bepaald wat inhoudelijk het niveau moet zijn dat een leerling die een bepaald referentieniveau heeft, moet halen. In de volgende fase is op basis van de verzamelde data gepresenteerd wat de impact van de gezette standaard was: gegeven deze standaard, hoeveel % van de leerlingen behaalt dan op dit moment het niveau? Dit werd gebruikt om de standaard eventueel nog iets bij te stellen.

In de praktijk worden de referentiesets gebruikt om in verschillende lopende toets- en examenprogramma's te bepalen of leerlingen een bepaald referentieniveau behaald hebben. Hiervoor is bij elk programma de standaard eenmalig overgebracht en dit wordt in iedere toets omgezet naar een cesuur: hoeveel opgaven moet een leerling goed maken om niveau 2F te behalen. Belangrijk is dat voor elke toets, in elk onderwijsniveau, dit niveau altijd te vertalen is naar de oorspronkelijke nationale standaard die gezet is via de referentiesets.

Bronnenlijst

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Bao, Y. (2019). *A diagnostic classification model for polytomous attributes* (Academisch proefschrift). University of Georgia.
- Bjork, I. & Romy, D. (1999) What Constitutes a Nursing Practical Skill?, *Western Journal of Nursing Research*, 21:1, 51-70.
- Bradshaw, L. & Levy, R. (2019). Interpreting probabilistic classifications from diagnostic psychometric models. *Educational Measurement: Issues and Practice*, 38 (2), 79–88.
<https://doi.org/10.1111/emip.12247>
- Bradshaw, L. & Madison, M.J. (2016). Invariance properties for general diagnostic classification models. *International Journal of Testing*, 16 (2), 99–118. <https://doi.org/10.1080/15305058.2015.1107076>
- Chalmers, R.P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48, 1–29. <https://doi.org/10.18637/jss.v048.i06>
- Chen, H. & Chen, J. (2016). Retrofitting non-cognitive-diagnostic reading assessment under the generalized DINA model framework. *Language Assessment Quarterly*, 13 (3), 218–230.
<https://doi.org/10.1080/15434303.2016.1210610>
- Chen, J. & de la Torre, J. (2013). A general cognitive diagnosis model for expert-defined polytomous attributes. *Applied Psychological Measurement*, 37 (6), 419–437.
<https://doi.org/10.3389/fpsyg.2018.01474>
- Chen, J., de la Torre, J. & Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*, 50 (2), 123–140.
<https://doi.org/10.1111/j.1745-3984.2012.00185.x>
- Clarke, L. & Winch, C. (2006) A European skills framework? – but what are skills? Anglo-Saxon versus German Concepts, *Journal of Education and Work*, 19:3, 255-269.
- Collins, L.M. & Wugalter, S.E. (1992). Latent class models for stage-sequential dynamic latent variables. *Multivariate Behavioral Research*, 27 (1), 131–157. <https://doi.org/10.1207/s15327906mbr27018>
- Daas, R., Ten Dam, G. & Dijkstra, A.B. (2016). Contemplating modes of assessing citizenship competences. *Studies in Educational Evaluation*, 51, 88-95.
- De la Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of Educational and Behavioral Statistics*, 34 (1), 115–130. <https://doi.org/10.3102/1076998607309474>

- De la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76 (2), 179–199.
<https://doi.org/10.1007/S11336-011-9207-7>
- Deegan, D., Wims, P. & Pettit, T. (2016) Practical Skills Training in Agricultural Education – A Comparison between Traditional and Blended Approaches, *The Journal of Agricultural Education and Extension*, 22(2), 145-161.
- Dijkstra, A.B., Daas, R., Munniksma, A. & Ten Dam, G. (2023). Citizenship in public, non-religious private and religious private school. A comparison of student citizenship in Dutch secondary education. *Educational Review*, 77(2), 427-445. <https://doi.org/10.1080/00131911.2023.2203393>
- Elffers, L. (2023). De grote (on)gelijkmaker: Onderwijs in een ongelijke samenleving [Oratie]. Amsterdam: Universiteit van Amsterdam.
- Evers, A., Lucassen, W., Meijer, R. & Sijsma, K. (2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. Verkregen van
<https://psynip.nl/wp-content/uploads/2021/11/COTAN-Beoordelingssysteem-2010.pdf>
- Expertgroep doorlopende leerlijnen taal en rekenen (2008). *Over de drempels met taal en rekenen, hoofdrapport van de expertgroep doorlopende leerlijnen taal en rekenen*. Enschede: SLO.
- Expertgroep doorlopende leerlijnen taal en rekenen. (2009). *Referentiekader taal en rekenen. De referentieniveaus*. Enschede: SLO.
- Fan, T. & Yan, X. (2020). Diagnosing English reading ability in Chinese senior high schools. *Studies in Educational Evaluation*, 67, 100931. <https://doi.org/10.1016/j.stueduc.2020.100931>
- George, A.C., Robitzsch, A., Kiefer, T., Groß, J. & Üllü, A. (2016). The R package CDM for cognitive diagnosis models. *Journal of Statistical Software*, 74 (2), 1–24. <https://doi.org/10.18637/jss.v074.i02>
- Godaert, E., Assaert, K., Voogt, J. & Van Braak, J. (2022). Assessment of students' digital competences in primary school: a systematic review. *Education and Information Technologies*, 27(4), 9953-10011
<https://doi.org/10.1007/s10639-022-11020-9>
- Groot, A.D. de. (1946) Het denken van den schaker: een experimenteel-psychologische studie, Amsterdam: NV Nederlandse Uitgeversmaatschappij.
- Heitink, M.C., Luyten, H., Meelissen, R.M., Veldkamp, B.P., Keuning, J., Noordhof, R. & Van Langen, A. (2024). *Digitale geletterdheid in het basisonderwijs: Technisch rapport Peil.Digitale Geletterdheid 2022*. Enschede: Universiteit Twente.
- Henson, R.A. & Templin, J. (2008). Implementation of standards setting for a geometry end-of-course exam using cognitive diagnosis. American Educational Research Association Conference. New York.
- Hopster-den Otter, D., Wools, S., Eggen, T.J.H.M. & Veldkamp, B.P. (2017). Formative use of test results: A user's perspective. *Studies in Educational Evaluation*, 52, 12–23.
<https://doi.org/10.1016/j.stueduc.2016.11.002>

- Huebner, A. & Wang, C. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement*, 71 (2), 407–419.
<https://doi.org/10.1177/0013164410388832>
- Johnson, M.S. & Sinharay, S. (2020). The reliability of the posterior probability of skill attainment in diagnostic classification models. *Journal of Educational and Behavioral Statistics*, 45 (1), 5–31.
<https://doi.org/10.3102/1076998619864550>
- Krathwohl, D.R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, 41(4), 212–218.
- Ma, W. & de la Torre, J. (2020). GDINA: An R package for cognitive diagnosis modeling. *Journal of Statistical Software*, 93 (14), 1–26. <https://doi.org/10.18637/jss.v093.i14>
- Ma, W., Iaconangelo, C. & de la Torre, J. (2016). Model similarity, model selection, and attribute classification. *Applied Psychological Measurement*, 40 (3), 200–217.
<https://doi.org/10.1177/0146621615621717>
- Maas, L., Brinkhuis, M.J.S., Kester, L. & Wijngaards-de Meij, L. (2022a). Cognitive diagnostic assessment in university statistics education: Valid and reliable skill measurement for actionable feedback using learning dashboards. *Applied Sciences*, 12 (10). <https://doi.org/10.3390/app12104809>
- Maas, L., Brinkhuis, M.J.S., Kester, L. & Wijngaards-de Meij, L. (2022b). Diagnostic classification models for actionable feedback in education: Effects of sample size and assessment length. *Frontiers in Education*, 7, article 802828. <https://doi.org/10.3389/feduc.2022.802828>
- Madison, M.J. & Bradshaw, L. (2018). Assessing growth in a diagnostic classification model framework. *Psychometrika*, 83 (4), 963–990. <https://doi.org/10.1007/s11336-018-9638-5>
- Madison, M.J., Haab, S., Jeon, M. & Cotterell, M.E. (2024). Package 'TDCM'. Verkregen van <https://cran.r-project.org/web/packages/TDCM/TDCM.pdf>
- Maydeu-Olivares, A. & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, 71 (4), 713. <https://doi.org/10.1007/s11336-005-1295-9>
- Maydeu-Olivares, A. & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research*, 49 (4), 305–328. <https://doi.org/10.1080/00273171.2014.911075>
- Min, S. & He, L. (2022). Developing individualized feedback for listening assessment: Combining standard setting and cognitive diagnostic assessment approaches. *Language Testing*, 39 (1), 90–116.
<https://doi.org/10.1177/0265532221995475>
- Norris, S.P., Macnab, J.S. & Phillips, L.M. (2007). Cognitive modeling of performance on diagnostic achievement tests: A philosophical analysis and justification. In: J.P. Leighton & M.J. Gierl (red.), *Cognitive diagnostic assessment for education* (pp. 61–84). New York: Cambridge University Press.
<https://doi.org/10.1017/CBO9780511611186.003>
- OCW Skills Platform (2016) *Skills voor de toekomst: een onderzoeksagenda*, Ministerie van Onderwijs, Cultuur & Wetenschap.

- OECD. (2023). Pisa 2022 results (volume i): The state of learning and equity in education. *PISA, OECD Publishing, Paris*.
- OECD(2024). *PISA2022 Database* [Data set]. <https://www.oecd.org/en/data/datasets/pisa-2022-database.html>
- Paul, M. L. J. (2024). Kamerbrief 'De waardering van praktijkgericht onderwijs en de toekomst van het vmbo'. 7 juni 2024.
- Paul, M. L. J. (20204). Kamerbrief 'Brief van de minister voor primair en voortgezet onderwijs' (nr. 732). 30 mei 2024.
- Ravand, H. & Baghaei, P. (2020). Diagnostic classification models: Recent developments, practical issues, and prospects. *International Journal of Testing*, 20 (1), 24–56.
<https://doi.org/10.1080/15305058.2019.1588278>
- Reichert, R., Pan, Q. & Chen, L.L. (2023). *Digital Literacy Assessment – Background paper prepared for the Global Education Monitoring Report*, UNESCO.
- Roduta Roberts, M. & Gierl, M.J. (2010). Developing score reports for cognitive diagnostic assessments. *Educational Measurement: Issues and Practice*, 29 (3), 25–38.
<https://doi.org/10.1111/j.1745-3992.2010.00181.x>
- Rupp, A.A., Templin, J. & Henson, R.A. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York: The Guilford Press.
- Ryan, J.M. (2006). Practices, issues, and trends in student test score reporting. In: S.M. Downing & T.M. Haladyna (red.), *Handbook of Test Development* (pp. 677–710). Mahwah, NJ: Lawrence Erlbaum.
- Siddiq, F., Hatlevik, O.E., Throndsen, I. & Scherer, R. (2016) Taking a future perspective by learning from the past – A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educational Research Review* 19, 58-84.
- SLO (2016). *Karakteristieken en kerndoelen. Onderbouw voortgezet onderwijs*.
<https://www.slo.nl/@6310/karakteristieken/>
- SLO (2021). *Startnotitie moderne vreemde talen. Bovenbouw voortgezet onderwijs*.
<https://www.slo.nl/zoeken/@20066/startnotitie-moderne-vreemde-talen/>
- SLO (2023a). *Conceptkerndoelen Nederlands en rekenen en wiskunde*. Geraadpleegd op 2 maart 2025 via <https://open.overheid.nl/documenten/dpc-1ed205485182719028fc2c97959bf0eb3d1ba85a/pdf>
- SLO (2023b). *Startnotitie moderne vreemde talen, actualisatie kerndoelen*. Geraadpleegd op 2 maart 2025 via <https://www.slo.nl/@20066/startnotitie-moderne-vreemde-talen/>
- SLO (2024a). *Definitieve conceptkerndoelen Nederlands, inclusief toelichtingsdocument*. Geraadpleegd op 2 maart 2025 via <https://www.slo.nl/@24174/definitieve-conceptkerndoelen-nederlands/>

- SLO (2024b). *Definitieve conceptkerndoelen rekenen en wiskunde, inclusief toelichtingsdocument*. Geraadpleegd op 2 maart 2025 via <https://www.slo.nl/@24175/definitieve-conceptkerndoelen-rekenen/>
- SLO (2024c). *Conceptkerndoelen Moderne vreemde talen, actualisatie kerndoelen*. Geraadpleegd op 2 maart 2025 via <https://www.slo.nl/thema/meer/actualisatie-kerndoelen-examenprogramma/actualisatie-kerndoelen/conceptkerndoelen-moderne-vreemde-talen/>
- Sol, Y. & Visser, A. (2023) *Vaardigheden in het landelijke curriculum*. Amersfoort: SLO.
- Ten Dam, G., Geijssel, F. & Ledoux, G. (2010). Burgerschapscompetenties: de ontwikkeling van een instrument. *Pedagogische Studiën* 87, 313-333.
- Tatsuoka, K.K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of educational measurement*, 345–354.
<https://doi.org/10.1111/j.1745-3984.1983.tb00212.x>
- Templin, J. & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika*, 79 (2), 317–339.
<https://doi.org/10.1007/s11336-013-9362-0>
- Tricot, A. & Sweller, J. (2014) Domain-specific knowledge and why teaching generic skills does not work, *Educational Psychology Review*, 26:2, 265–283.
- Van der Kleij, F.M., Eggen, T.J. & Engelen, R.J. (2014). Towards valid score reports in the computer program loves: A redesign study. *Studies in Educational Evaluation*, 43 , 24–39.
<https://doi.org/doi.org/10.1016/j.stueduc.2014.04.004>
- Van Til, A. & Van Boxtel, H. (2015). *Wetenschappelijke verantwoording toets 0 t/m 3, tweede generatie*. Arnhem: Cito BV.
- Vogel, D. & Harendza, S. (2016) Basic practical skills teaching and learning in undergraduate medical education – a review on methodological evidence, *GMS Journal for Medical Education*, 33:4, Doc64.
- Von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology* , 61 (2), 287–307.
<https://doi.org/10.1348/000711007X193957>
- Wang, W., Song, L., Chen, P., Meng, Y. & Ding, S. (2015). Attribute-level and pattern-level classification consistency and accuracy indices for cognitive diagnostic assessment. *Journal of Educational Measurement* , 52 (4), 457–476. <https://doi.org/10.1111/jedm.12096>

