



UvA-DARE (Digital Academic Repository)

Het meten van skills: Verkenning van mogelijkheden voor een skills module gekoppeld aan het Nationaal Cohortonderzoek Onderwijs

Daas, R.J.M.; Dijkstra, Anne Bert; Roelofs, Erik; Sluijter, Cor

[Link to publication](#)

Citation for published version (APA):

Daas, R. J. M., Dijkstra, A. B., Roelofs, E., & Sluijter, C. (2020). *Het meten van skills: Verkenning van mogelijkheden voor een skills module gekoppeld aan het Nationaal Cohortonderzoek Onderwijs*.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

HET METEN VAN SKILLS

Verkenning van mogelijkheden
voor een skills module
gekoppeld aan het Nationaal
Cohortonderzoek Onderwijs

Remmert Daas
Anne Bert Dijkstra
Erik Roelofs
Cor Sluiter

Het meten van skills:

**Verkenning van mogelijkheden voor een skills module gekoppeld
aan het Nationaal Cohortonderzoek Onderwijs**

Remmert Daas

Anne Bert Dijkstra

Erik Roelofs

Cor Sluijter

Colofon



UNIVERSITEIT VAN AMSTERDAM



Juli, 2020

Ontwerp omslag: Maedium, Utrecht

© De auteurs

Gesubsidieerd door het Ministerie van Onderwijs, Cultuur en Wetenschap

Naar verwijzen als:

Daas, R., Dijkstra, A. B., Roelofs, E., & Sluijter, C. (2020). *Het meten van skills: Verkenning van mogelijkheden voor een skills module gekoppeld aan het Nationaal Cohortonderzoek Onderwijs*. Universiteit van Amsterdam & Cito.

Samenvatting separaat gepubliceerd als:

Daas, R., Dijkstra, A. B., Roelofs, E., & Sluijter, C. (2020). *Het meten van skills: Verkenning van mogelijkheden voor een skills module gekoppeld aan het Nationaal Cohortonderzoek Onderwijs – Samenvatting*. Amsterdam University Press.

Inhoudsopgave

Inleiding	3
Burgerschapscompetenties.....	17
Digitale geletterdheid.....	26
Sociaal-emotionele ontwikkeling.....	34
Zelfregulatie.....	40
Conclusie.....	47
Referenties	63
Bijlage 1: Resultaten instrumentanalyses	83
Bijlage 2: Gehanteerd screeningsinstrument voor skills-instrumenten	116

Inleiding

Skills zijn van belang voor het succesvol functioneren op een breed scala aan terreinen (OECD, 2011, 2015, 2019). Skills worden wel omschreven als “het geheel van kennis, houding en vaardigheden dat van belang is voor realisering van een bepaald doel” (Van den Berge et al., 2014, p.16). Vanwege het grote belang van skills voor individu en samenleving is de ontwikkeling ervan een centraal aandachtspunt in het funderend onderwijs. De in de afgelopen jaren sterk toegenomen aandacht voor de verwerving van sociale en maatschappelijke competenties is daarvan een illustratie, en kreeg onder meer vorm in een wettelijke opdracht tot bevordering van burgerschap. In de aandacht voor skills wordt daarbij opnieuw zichtbaar dat kwalificeren en socialiseren, als centrale functies van onderwijs, nauw verbonden zijn (Dijkstra, 2012).

Meer recent is de aandacht voor een reeks skills die bekendheid verwierf onder de noemer ‘21st century skills’. Hierbij gaat het om skills als samenwerken, communiceren en ICT geletterdheid (Voogt & Pareja Roblin, 2012). Al wat langer wordt ook het begrip ‘advanced skills’ gebruikt als aanduiding voor skills die tot op zekere hoogte leerdomeinneutral zijn, maar in samenhang met de inhoudelijke leerdomeinen worden ontwikkeld (Skills-platform, 2016). Het gaat daarbij bijvoorbeeld om kritisch en creatief denken, en zelfregulatie. Van zowel 21st century skills als advanced skills doen uiteenlopende definities de ronde. Hoewel de verschillende conceptuele modellen inhoudelijk op elkaar lijken, wordt niet altijd duidelijk waar de betreffende begrippen naar verwijzen en hoe de samenstellende delen (onder te delen in kennis, houding en vaardigheden) samenhangen. Zoals ook elders aangeduid gaat het om skills die, onder dezelfde of andere benamingen, reeds lang aandacht krijgen (cf. Dijkstra, 2012; Ledoux et al., 2013; Van den Berge et al., 2014; Voogt & Pareja Roblin, 2012).

In de literatuur rond skills is een veelheid aan begrippen te vinden. Naast de genoemde paraplu termen *advanced skills* en *21st century skills* worden ook andere, meer of minder verwante aanduidingen gebruikt (zoals *lifelong learning competencies*, *key competencies* of *life skills* en andere; Rychen & Salganik, 2003) en worden uiteenlopende uitwerkingen en afbakeningen gebruikt. Ook van de begrippen die daartoe gerekend worden, zoals zelfregulatie, sociale en maatschappelijke competenties, en de conceptuele afbakening van en verbanden tussen concepten bestaat vooralsnog geen eenduidig beeld (Daas & Dijkstra, 2016; Opdenakker & Ledoux, 2016).

Het veelvormige conceptuele beeld rond bovenstaande begrippen lijkt tevens illustratief voor het gebrek aan empirisch onderzoek naar hoe deze skills zich in samenhang ontwikkelen (vgl. Skills-platform, 2016). Naast de pas relatief recent ontstane aandacht voor skills, is dat niet in de laatste plaats vanwege de complexiteit van begripsafbakening en integrale conceptualisering, de ontwikkeling van valide meetinstrumenten en, mede als gevolg daarvan, gebrek aan empirisch onderbouwde kennis over de verwerving en ontwikkeling ervan (Pepper, 2011; Skills-platform, 2016). Het Nationaal Cohortonderzoek Onderwijs (NCO) wil dan ook bijdragen aan de verdere ontwikkeling van inzicht in de verwerving en ontwikkeling van skills. De mogelijkheden tot longitudinale dataverzameling en koppeling aan bestaande gegevens maken het in beginsel mogelijk om te komen tot inzicht in de wijze waarop leerlingen zich op verschillende domeinen effectief, breed, en samenhangend ontwikkelen, en welke invloed de school daarop uitoefent. Dat leidt tot de vraag hoe binnen dat kader een module, een (set van) instrument(en), aan de hand waarvan deze skills gemeten kunnen worden, vorm zou kunnen krijgen.

Om deze vraag te beantwoorden wordt in deze verkennende studie een reeks voorwaardelijke stappen verkend en uitgewerkt. We gebruiken het begrip ‘advanced skills’ als verzamelterm voor meer ‘geavanceerde’ competenties. We verstaan daaronder de

competenties waarvoor een redelijke mate van consensus lijkt te bestaan dat deze van belang zijn voor succesvolle onderwijsdeelname en die vanwege de technologische, economische en maatschappelijke ontwikkelingen in de samenleving nodig zijn voor een succesvolle individuele levensloop en participatie op de arbeidsmarkt en deelname aan de samenleving (Dijkstra, 2012; Ledoux et al., 2013; Rychen & Salganik, 2003; Voogt & Pareja Roblin, 2012).

Onderzoeksvragen

Dit onderzoek richt zich op de vraag: Hoe zou meting van advanced skills kunnen worden geïntegreerd in een module als onderdeel van het NCO?

Om deze vraag te beantwoorden zullen de volgende deelvragen uitgewerkt worden:

- 1) Welke (advanced) skills worden verondersteld een bijdrage te leveren aan een succesvolle schoolloopbaan c.q. leeropbrengsten?
- 2) Wat is er bekend over de ontwikkeling van deze skills gedurende de levensloop en de relatie tussen deze skills en de ontwikkeling van andere leeropbrengsten?
- 3) Hoe kunnen deze skills geoperationaliseerd worden in valide meetinstrumenten?
- 4) In hoeverre maken bestaande meetinstrumenten valide meting van deze skills mogelijk?
- 5) In hoeverre zijn meetinstrumenten voor advanced skills geschikt voor de integratie in het NCO?

Proces van selectie van skills

Voornoemde vragen impliceren inzicht in de longitudinale ontwikkeling van skills, de (cumulatieve) bijdrage van onderwijs gedurende de schoolloopbaan en de wederzijdse relatie tussen de ontwikkeling van skills en andere competenties. Voor uitwerking van longitudinale meting van skills lijkt het om boven beschreven redenen momenteel te vroeg om in te zetten

op ontwikkeling van een integraal theoretisch kader rond (de ontwikkeling van) advanced skills. De afwezigheid van basale empirische kennis en het ontbreken van de voorwaarden om daarin te kunnen voorzien vormt daarbij een van de belangrijke belemmeringen. Gezien de huidige staat van kennis en de beperkte (of afwezige) beschikbaarheid van adequate instrumenten (vgl. Ledoux et al., 2013; Skills-platform, 2016; Van den Berge et al., 2014) is deze verkennende studie naar de inrichting van een skills module in het NCO gericht op een (beperkte) selectie van skills. Deze selectie komt tot stand op basis van een aantal generieke criteria (gericht op de conceptuele inzichten rond de verwerving) en meet-technische criteria (gericht op het faciliteren van valide metingen).

Generieke criteria

- Er is sprake is van (voldoende) *conceptuele helderheid* om tenminste een voorlopige definitie van de betreffende advanced skill mogelijk te maken als voorwaarde voor valide meting;
- De betreffende advanced skills lijken zich te lenen voor *betekenisvolle beïnvloeding door onderwijs*, zoals blijkt uit de literatuur omtrent (aanwijzingen voor) effectief gebleken educatieve interventies die kunnen gelden als indicatie van de *leerbaarheid* van de betreffende skill;
- Er is *inzicht verworven in de bijdrage van deze skills* aan andere leeropbrengsten of later in de levensloop.

Meet-technische criteria

De verkenning is pragmatisch georiënteerd, in de zin dat verkenning van de uitvoerbaarheid van een NCO module het primaire oogmerk is, om ook aanwijzingen te geven binnen welke contouren dat zou kunnen plaatsvinden. Dat betekent dat vragen rond meetbaarheid in deze fase een leidende rol spelen. Omdat de te selecteren advanced skills operationaliseerbaar moeten zijn en via (verder te ontwikkelen of reeds beschikbare) meetinstrumenten in kaart

gebracht moeten kunnen worden, spelen de volgende meet-technische criteria een belangrijke rol:

- De mogelijke effecten van interventies op de advanced skill veronderstellen een *aanzienlijke range van leeftijden*, waardoor leer/ontwikkelingsprogressies in de advanced skill te onderscheiden zijn. Dat creëert ook de mogelijkheid om instrumenten voor longitudinale metingen te ontwikkelen.
- Van de te selecteren instrumenten is *voldoende informatie bekend om de validiteit van uitspraken gedaan op grond van gebruik van het instrument te kunnen beoordelen*. Hoewel informatie ontleend aan de Nederlandse situatie voorkeur heeft, volstaan we vooralsnog met het criterium dat informatie verzameld in het funderend onderwijs, ongeacht de nationale contexten waarin de gegevensverzameling plaatsvond, volstaat. We gaan daarbij uit van de richtlijnen van het Evidence-Centered Design (ECD).

Een beschrijving van het Evidence-Centered toetsontwerp is te vinden in Roelofs en Sluiter (2016). Het ECD-kader fungeert in deze studie als uitgangspunt bij de beschrijving en selectie van de instrumenten waarmee skills in kaart gebracht kunnen worden. Het ijkpunt is daarbij in hoeverre de instrumenten voldoen aan de eisen voor Evidence-Centered ontwerp, waarbij de kwaliteit van vijf deelmodellen een centrale rol speelt: het leerlingmodel, taakmodel, bewijsmodel, assemblagemodel en presentatiemodel. We lichten deze modellen hier kort toe en verwijzen voor meer informatie naar Roelofs en Sluiter (2016).

Het *leerlingmodel* beschrijft de aard van het construct (of de doelvaardigheid) dat centraal staat. Het model vat de essentiële benodigde constructen samen, dat kunnen kenniselementen, vaardigheden, strategieën, en attitude-elementen zijn, waarover het instrument moet rapporteren. In de praktijk resulteert dit in de regel in een verzameling te toetsen onderwijsdoelen, samengevat onder rubrieken of domeinen.

Het *taakmodel* bevat de operationalisering van de constructen in het leerlingmodel. Dit model kunnen we zien als de blauwdruk voor een te ontwikkelen instrument. Het taakmodel beschrijft de sleutelkenmerken van de taken en de taaksituaties die nodig zijn om in kaart te brengen wat de leerling denkt en doet; de wijze waarop de taken worden aangeboden, zoals instructies, structurering, stimuli, prompts; en de van de leerling verwachte responsen (als acties of producten).

Het *bewijsmodel* vormt de brug tussen het taakmodel en het leerlingmodel. Het beschrijft de methode waarlangs de verkregen uitwerkingen van opdrachten door leerlingen moeten worden gescoord en gekwantificeerd tot scores op hoofdattributen, deelattributen en subattributen die weer herleidbaar zijn tot de constructen en daarmee samenhangende leerdoelen. Hierbij spelen drie zaken een rol: 1) betrouwbare scoring van afzonderlijke responsen; 2) schaalbaarheid van de afzonderlijke opgaven tot betekenisvolle en constructen, zoals beoogd in het leerlingmodel; en 3) betekenisvolle standaarden voor niveaus van beheersing.

Het *assemblagemodel* maakt duidelijk hoe opgaven worden geselecteerd voor een instrument en hoe de samenstelling van het instrument als geheel eruit ziet. Opgaven dienen de boogde constructen, en hoe die moeten worden aangetoond in taken en taaksituaties (het domein) zoals beoogd in het leerlingmodel goed te dekken.

Het *presentatiemodel* beschrijft hoe de afname van instrumentonderdelen georganiseerd is. In dit model wordt onder andere uitgewerkt hoe de afname van het instrument plaatsvindt, hoe bewijzen worden verzameld, geregistreerd, gescoord, opgeslagen en verwerkt, en hoe de rapportages tot stand komen en worden gevisualiseerd.

Op basis van de voornoemde criteria kunnen instrumenten worden geconstrueerd. Evenzo kunnen bestaande instrumenten beoordeeld worden op de mate waarin ze voorzien in de voorwaarden die elk van de hierboven genoemde model aan het instrument stelt.

Zoals al naar voren kwam, wordt een veelheid aan skills onderscheiden. Hoewel we in het ideale geval een brede doorsnede daarvan zouden willen meenemen in deze verkenning, maken overwegingen van praktische uitvoerbaarheid keuzes onvermijdelijk. Voor zover bruikbare instrumenten ontbreken wordt zo'n keuze gemakkelijker, in zoverre dat meting niet mogelijk is zonder voorafgaande investering in instrumentontwikkeling. Verdere ontwikkeling van meetinstrumenten is van belang (cf. Daas & Dijkstra, 2017; Daas, Dijkstra, & De Korte, 2019; Dijkstra, 2015) maar bemoeilijkt realisering van metingen op korte termijn. Onder herhaling van het pleidooi voor verdere ontwikkeling van instrumenten, richten we ons hier dan ook op selectie van instrumenten die reeds beschikbaar zijn of die met aanpassingen (met name vertaling naar het Nederlands en de Nederlandse context) beschikbaar gemaakt zouden kunnen worden.

Wat betreft de methodologische overwegingen die van belang zijn voor de selectie van instrumenten hanteren we de in de vorige paragraaf besproken generieke en meet-technische criteria. Zoals verderop zal blijken leidt toepassing van deze criteria tot meerdere instrumenten die bruikbaar zijn (eventueel na aanpassingen). Daarnaast is een op inhoudelijke overwegingen gebaseerde keuze van skills nodig. In dat verband moet in de eerste plaats worden opgemerkt dat die keuze in essentie arbitrair is en wordt ingegeven door de beoogde doelen van de meting en het gebruik dat van de gegevens zal worden gemaakt. Eerder formuleerden we in dit verband dat het gaat om skills waarbij de gecombineerde inzet van conceptuele kennis, vaardigheden en houdingen belangrijk zijn, en die vanwege snelle ontwikkeling op uiteenlopende domeinen nodig zijn voor een succesvolle individuele levensloop, maatschappelijke participatie en deelname aan de arbeidsmarkt. Hoewel daarmee enige richting gegeven wordt, is deze omschrijving nog onvoldoende specifiek om een precieze bepaling van te meten skills op te kunnen baseren. We hebben daarom voor een strategie gekozen die niet zozeer op een (arbitraire) expliciete inhoudelijke afwegingen is

gebaseerd, maar uitgaat van de gedachte dat in deze fase een bredere benadering de voorkeur heeft, waarin vooralsnog voor een over verschillende domeinen gespreide meting gekozen wordt. Daarmee ontstaat enerzijds een brede en uiteenlopende eerste indruk, op basis waarvan in de toekomst verdere keuzes mogelijk zijn. Anderzijds wordt bij wijze van ‘risicospreiding’ een onbedoeld smalle of eenzijdige selectie tegengegaan. Op deze manier kunnen skills van uiteenlopende aard in kaart worden gebracht zodat, zo lang nog weinig kennis beschikbaar is over de ontwikkeling en relevantie van de verschillende skills, inzicht verkregen wordt in meetmogelijkheden voor skills op uiteenlopende domeinen.

Een en ander neemt niet weg dat de gemaakte selectie dus enigszins arbitrair van karakter is. Gegeven de nog beperkte beschikbaarheid van kennis over het belang van de te selecteren skills valt dat, denken we, in deze fase niet goed te vermijden. De voorgestelde selectie moet dan ook als een eerste, verkennende stap worden gezien. Op grond van de te verzamelen gegevens en empirische analyse van de bijdrage van deze skills aan succes op de genoemde maatschappelijke domeinen kan in latere fasen worden bezien welke uitbreidingen of aanpassingen vervolgens wenselijk zijn.

Selectie van skills en instrumenten

Bij de selectie is voortgebouwd op inzichten ontleend aan analyses zoals die in de afgelopen jaren beschikbaar kwamen omtrent de skills die vanwege maatschappelijke modernisering belangrijk lijken (o.a. Dijkstra, 2012; European Communities, 2007; Lai, 2011; OECD, 2015; Peschar et al., 2010; Rychen & Salganik, 2003; Voogt et al., 2019). In ons streven naar dekking van een brede range domeinen richten we ons op de skills die van belang geacht worden door veranderingen in informatievoorziening, persoonsontwikkeling en in menselijk samenleven.

De snelle en vergaande digitalisering en het beschikbaar komen van moderne communicatiemiddelen (zoals in de vorm van sociale media) voor grote groepen mensen leidt ertoe dat zowel de omvang en snelheid waarmee informatie beschikbaar is, als de ruime mogelijkheden ook zelf informatie te verspreiden, een groot beroep doet op het vermogen om met digitale middelen om te gaan, grote hoeveelheden informatie te verwerken en relevante informatie te selecteren, benutten en op verantwoorde wijze te delen. Zowel het belang van informatie als hulpbron voor individueel functioneren als de maatschappelijke effecten van de snelle en ongefilterde verspreiding van informatie, leidt tot nieuwe opgaven en maakt digitale geletterdheid tot een belangrijke competentie, waarbij zowel de ‘technische’ als ‘sociale’ aspecten daarvan aandacht vragen (Voogt et al., 2019).

Een belangrijk gegeven is de toegenomen snelheid van moderniseringsprocessen en maatschappelijke verandering als gevolg daarvan. Dit heeft mede als gevolg dat gedurende de levensloop meer dan voorheen een beroep wordt gedaan op het aanpassingsvermogen van mensen, zoals op de arbeidsmarkt. Die veranderingen laten zich bovendien niet goed voorspellen. De combinatie van snelle verandering en onbekendheid met de verandering die nodig kan zijn, doet een beroep op de skills die nodig zijn om daarmee te kunnen omgaan, zoals het vermogen tot zelfsturing en goed ontwikkelde sociaal-emotionele vaardigheden, die vanuit zowel vanuit individueel als maatschappelijk perspectief waardevol zijn (OECD, 2015, 2018; Rychen & Salganik, 2003). In dit verband richten we ons op zelfregulatie en sociaal-emotionele ontwikkeling.

Eén van de belangrijke kenmerken van de maatschappelijke transformaties die zich voordoen is de sterk toegenomen diversiteit en onderlinge verbondenheid van samenlevingen die, in combinatie met het afgenomen socialiserend vermogen van instituties zoals kerk en buurt, leiden tot nieuwe vragen rond effectieve waardenoverdracht en bevordering van maatschappelijke samenhang. Aandacht voor competenties zoals omgaan met verschil en het

oplossen van conflict is daarvan het gevolg. Meer algemeen gaat het om burgerschapscompetenties, die enerzijds bijdragen aan individueel functioneren (zoals autonomie, kritisch vermogen, omgaan met verschil) en anderzijds gericht zijn op realisering van collectieve belangen. Bij de laatste gaat het zowel om functionele overwegingen zoals het maatschappelijk belang van sociaal kapitaal, als om borging van democratische waarden als voorwaarde voor bestendiging van open, vrije samenlevingen (Dijkstra, 2012; OECD, 2007, 2010).

Al met al richt deze studie zich zo op vier skills: burgerschapscompetenties, digitale geletterdheid, sociaal-emotionele ontwikkeling en zelfregulatie. Deze skills bestrijken verschillende domeinen van het advanced skills concept, en bieden zodoende ook aanknopingspunten voor de uitwerking van andere skills. De uitwerking is bovendien actueel gezien de ontwikkelingen van een nieuw onderwijscurriculum waarin deze skills terugkomen in de ‘leergebieden’ (t.w. burgerschap en digitale geletterdheid), als ook in de ‘brede denkvaardigheden’ (t.w. sociaal-emotionele ontwikkeling, en zelfregulatie).

De keuze voor juist deze set skills is bovendien tot stand gekomen in een aantal overlegondes met vertegenwoordigers uit verschillende geledingen van de Nederlandse onderwijssectoren, onderzoeksinstituten, universiteiten en onderwijsverzorgingsinstellingen verenigd in het Skills-platform. Uiteindelijk is gekozen voor nadere verkenning van de vier genoemde skills als meest direct relevant, zonder daarbij afbreuk te doen aan het belang van andere advanced skills.

Voor de verkenning zijn in totaal 36 instrumenten geanalyseerd: 7 voor burgerschap, 14 voor digitale geletterdheid, 9 voor sociaal-emotionele ontwikkeling en 6 voor zelfregulatie. Per skill is getracht de meest gebruikte instrumenten in de Nederlandse context mee te nemen, en deze waarin relevant geacht aan te vullen met internationale instrumenten. Het gaat daarbij om instrumenten die potentieel geschikt zijn voor gebruik op stelselniveau.

Met deze selectie beogen we dus geen systematische beschrijving te geven van alle beschikbare instrumenten, maar een onderbouwd beeld van de reikwijdte van het relevante beschikbare instrumentarium en de kwaliteiten daarvan. De analyses per skill zijn opgenomen als hoofdstukken in dit rapport. De tabellen met de analyseresultaten, geordend naar aspecten van Evidence-Centered Design, zijn opgenomen in Bijlage 1.

Ontwikkeling van het screeningskader voor Skills-instrumenten

Om systematische te kunnen vaststellen of skills-instrumenten voldoen aan de hierboven genoemde generieke en meettechnische criteria, is een screeningskader uitgewerkt in een aantal rubrieken met specifieke evaluatieve vragen. Daarbij is gebruik gemaakt van een combinatie van het beoordelingssysteem voor Competentie-instrumenten, ontwikkeld door Wools et al. (2011) en de richtlijnen voor Evidence-Centered toetsontwerp van skills-instrumenten zoals beschreven door Roelofs & Sluijter (2016). Een aantal van de rubrieken van Wools et al. is daarbij specifiek uitgewerkt met behulp van richtlijnen voor ECD.

In het Beoordelingsinstrument Competentie-assessment van Wools et al. (2011) zijn de volgende hoofdruubrieken opgenomen:

1. Uitgangspunten van testconstructie (doel, doelgroep, competentie, theoretische onderbouwing, relevantie);
2. Kwaliteit van het testmateriaal (inhoud, ontwerp, vormgeving);
3. Kwaliteit van de handleiding (Informatieverstrekking over het instrument, voor assessoren, voor kandidaten);
4. Normen en standaarden (representativiteit normgroep, correcte berekeningen, duidelijkheid normschalen, correctheid keuze en uitvoering standaardbepalingsmethode);

5. Betrouwbaarheid (representatieve doelgroep, berekening juist, voldoen voor gebruiksdoeleinden);
6. Validiteit (interpretatief argument bevat aannemelijk en juiste inferenties, validiteitsargument: voldoende bewijzen)

Een aantal van de rubrieken is ten behoeve van deze studie aangepast of aangevuld aan de hand van de richtlijnen voor Evidence-Centered toetsontwerp. Daarin volgen we de gedachtegang van Mislavy (2007a) dat het specificeren van het toetsontwerp met daarin het basisraamwerk van leerlingmodel-taakmodel-bewijsmodel-assemblagemodel-presentatiemodel de basis legt om valide gevolgtrekkingen te kunnen doen over het presteren op toetstaken. ECD beoogt hiermee om al vanaf de fase van instrumentontwerp voor te bereiden op een validiteitsargumentatie (Mislavy, 2007b).

De rubriek 'Validiteit' wordt in het nieuwe skills-screeningskader vervat in de ECD-deelmodellen en is niet als aparte rubriek opgenomen. De rubriek 'Uitgangspunten van testconstructie' wordt vervat in het *Leerlingmodel*. De rubriek 'Kwaliteit van het testmateriaal' wordt vervat in de rubriek *Taakmodel*. De rubrieken 'Gebruikte normen en standaarden', 'Betrouwbaarheid' zijn integraal overgenomen in de brede rubriek *Bewijsmodel*. Bovendien zijn meer vragen gesteld over de beoordelaar, omdat verwacht werd dat beoordelaars een belangrijke rol zouden hebben bij het gebruik van skill-instrumenten. Daarnaast is de rubriek 'Kwaliteit van de handleiding' verkort tot 'Beschikbaarheid van de handleiding' in het screeningskader. Zodoende resulteert een screeningssysteem met in totaal 8 rubrieken, die dus deels worden gevat in bovenliggende ECD-rubrieken:

1. Leerlingmodel
2. Taakmodel
3. Bewijsmodel: beoordelaar en aard responsedata
4. Bewijsmodel: gebruikte normen en standaarden

5. Bewijsmodel: betrouwbaarheid
6. Assemblagemodel
7. Presentatiemodel
8. Beschikbaarheid van de handleiding

Tabel 1 geeft een overzicht van alle screeningsrubrieken en aandachtspunten die er deel van uitmaken. De nummers verwijzen naar de bijbehorende vragen die door de screener bij elk instrument zijn beantwoord. In Bijlage 2 zijn alle evaluatieve vragen opgenomen. Deze zijn opgenomen in een web-based vragenlijst, die tijd en plaats-onafhankelijke invulling mogelijk maakt, waarbij de data direct beschikbaar komen voor export en nadere analyses.

Tabel 1. Screeningskader skills-instrumenten en herkomst van de rubrieken

Rubriek	Herkomst
Algemeen (0)	Ontwikkeld ten behoeve van de inventarisatie.
1. Naam	
2. Auteurs	
3. Jaar	
4. Land	
5. Skill	
6. Type	
Leerlingmodel (1)	Vragen 1 t/m 8 zijn gebaseerd op een uitwerking van de richtlijnen voor skills-instrumenten volgens het ECD-model (Roelofs & Sluijter, 2016)
11. Conceptueel helder onderbouwd leerlingmodel	
12. Aanleerbaarheid	
13. Criteriumvaliditeit	Dit onderdeel komt deels overeen met de rubriek 'uitgangspunten van testconstructie' van het Beoordelingsinstrument Competentie-assessment Wools et al. (2011)
14. Taakprocessen	
15. Ontwikkelingsstadia	
16. Deel of geheelbenadering?	
17. Specificering taken en situaties	
18. Doelgroep: breedte leeftijdsrange en schooltype	
Taakmodel (2)	Vragen 21 t/m 25 zijn gebaseerd op een uitwerking van de richtlijnen voor skills-instrumenten volgens het ECD-model (Roelofs & Sluijter, 2016)
21. Operationalisatie in taken	
21a. Operationalisatie in taakrelevant gedrag	
21a. Operationalisatie situatiekenmerken	Dit onderdeel is een aanscherping van de rubriek 'kwaliteit van het testmateriaal' van het Beoordelingsinstrument Competentie-assessment Wools et al. (2011)
21c Basis van score of uitslag	
21d. Directheid van taken	
22. Prestatiecondities	
23. Taakcomplexiteitsfactoren	
24. Responseformat	
25. Geschiktheid responsekanaal	
Bewijsmodel: beoordelaar en aard responsedata (3)	Vragen 31 & 32 zijn gebaseerd op een uitwerking van de richtlijnen voor skills-instrumenten volgens het ECD-model (Roelofs & Sluijter, 2016)
31a. Beoordeling van indirecte prestatie	
31b. Beoordeling van directe prestatie	
32a1. Data-invoer registratie uitvoering	Vragen 33 t/m 38 komen uit het beoordelingsinstrument Competentie-assessment Wools et al. (2011)
32a2. Data-invoer scoring aanpak	
32a3. Data-invoer registratie antwoord	
32a4. Data-invoer scoring open antwoord	
32a5. Data-invoer scoring gesloten antwoord	
32b. Berekening totaalscores	

33. Beoordelaarstraining	
34. Beoordelingsmodel of scoremodel	
35. Expliciete beoordelingsregels	
36. Beoordelaars-betrouwbaarheid	
37. Beoordelingsalgoritme	
38. Onderbouwing samenstelling toetsscores	
Bewijsmodel: Gebruikte normen en standaarden (4)	Vragen 41 t/m 48 komen uit het beoordelingsinstrument Competentie-assessment Wools et al. (2011)
41. Validiteitsbewijzen	
42. Representativiteit normgroep	
43. Toereikende steekproefgrootte	
44. Gegevens over scoreverdeling	
45. Betekenis normschaal	
46. Normschaal en gebruiksdoel	
47. Gebruikte normen en standaarden	
48. Correctheid standaardbepalingsmethode	
Bewijsmodel: betrouwbaarheid (5)	Vragen 51 & 52 komen uit het Beoordelingsinstrument Competentie-assessment Wools et al. (2011)
51. Gerichtheid op doelpopulatie	
52a. Betrouwbaar voor belangrijke beslissing	
52b. Betrouwbaar voor minder belangrijke beslissing	
52c. Betrouwbaar voor conclusies op stelselniveau	
Assemblagemodel (6)	Vragen 61 & 62 zijn gebaseerd op een uitwerking van de richtlijnen voor skills-instrumenten volgens het ECD-model (Roelofs & Sluijter, 2016)
61. Specificatie matrijs-domein naar kenmerken van te dekken taaksituaties	
62. Representativiteit samenstelling naar taaksituaties en omstandigheden	
Presentatiemodel (7)	Vragen 71 & 72 zijn gebaseerd op een uitwerking van de richtlijnen voor skills-instrumenten volgens het ECD-model (Roelofs & Sluijter, 2016)
71. Locatie taakuitvoering	
72. Locatie vastleggen assessmentdata	
Beschikbaarheid van de handleiding (8)	Vraag 81 komt uit het Beoordelingsinstrument Competentie-assessment Wools et al. (2011)
81. Beschikbaarheid handleiding	

Burgerschapscompetenties

Burgerschapscompetenties hebben betrekking op de ontwikkeling van jongeren als deelnemer in de samenleving. Afhankelijk van de omvang van de ‘gemeenschap’ wordt ook wel gesproken over Europees burgerschap of wereldburgerschap, maar doorgaans wordt de opdracht voor scholen opgevat als ontwikkeling van jongeren tot burger in de nationale samenleving, waarbij het volgens de wet (in navolging van het advies van de Onderwijsraad, 2003, p. 10) gaat om de ontwikkeling van “de bereidheid en het vermogen deel uit te maken van een gemeenschap en daar een actieve bijdrage aan te leveren”. Algemener geformuleerd gaat het daarbij om het evenwicht tussen individuele vrijheid en sociale omgang, waarvan persoonlijke autonomie, interne cohesie (de binding met deelgroepen waarin mensen leven) en externe cohesie (de samenleving) de componenten zijn (Dijkstra, 2012).

Hoewel het op het eerste gezicht lijkt alsof er van veel uiteenlopende opvattingen over wat onder ‘burgerschap’ wordt verstaan rekening moet worden uitgegaan, lijkt dat mede het gevolg van het accent dat lange tijd lag op de verschillen in burgerschapsconcepties. Daarbij moet echter niet uit het oog verloren worden dat over de kern van waarom het bij burgerschap gaat – kort samen te vatten als het beginsel van de ‘*golden rule*’ gebaseerd op wederkerig altruïsme – brede overeenstemming bestaat. Hoewel dat niet betekent dat er geen verschillende accenten mogelijk zijn en de legitimaties uiteenlopen, kan zeker van een breed gedeelde kern worden gesproken, die zich richt op de procedurele waarden die voor vreedzaam menselijk samenleven nodig zijn (Dijkstra, 2012; Eindhoven et al., 2015). Dit is ook het uitgangspunt dat hier bij de selectie van instrumenten wordt gevolgd.

Dat neemt niet weg dat, zoals gezegd, de breedte van het begrip burgerschap – die mede het gevolg is van de veelheid van antwoorden op wat het betekent om met elkaar ‘samen te leven’ – ertoe leidt dat sprake is van verschillende kaders en accenten. Dit is ook terug te zien in de theoretisch kaders die voor burgerschap ontwikkeld zijn (zie ook Visser,

2018). Zo legt het kader ontwikkeld door Ten Dam et al. (2010) vanwege het uitgangspunt dat burgerschap niet iets voor ‘later’ is, maar ook al de situatie betreft waarin jongeren leven, het accent meer op sociale aspecten van burgerschap. Dit kader wordt gebruikt bij het instrument ‘Burgerschap meten’, welke onder andere gebruikt is in het Cohortonderzoek onderwijsloopbanen (COOL5-18) en de Scholenpanels Burgerschap (Peschar et al., 2010). Ter vergelijking legt het kader ontwikkeld door Schulz et al. (2008, 2016) meer nadruk op democratische structuren. Dit kader wordt gebruikt in de International Civic and Citizenship Education Study (ICCS). Het kader ontwikkeld door Wagenaar et al. (2011) ten behoeve van het peilingsonderzoek in het primair onderwijs werkt burgerschapskennis uit vanuit de onderwerpen sociaal-cultureel, politiek en economisch burgerschap, en komt daarmee ook op een andere operationalisering in kennistoetsen dan Burgerschap meten en ICCS (voor overzichten zie onder meer Daas, Dijkstra & de Korte, 2019; Daas & Mennes, 2018).

In vrijwel alle empirische onderzoeken naar de ontwikkeling van burgerschap onder jongeren wordt onderscheid gemaakt naar de te ontwikkelen kennis, houdingen en vaardigheden. Kennis wordt doorgaans gemeten door middel van meerkeuzevragen; soms aangevuld met enkele korte open vragen. Houdingen worden doorgaans gemeten door middel van vragen met Likert-schaal antwoordmogelijkheden. Vaardigheden worden doorgaans indirect gemeten, door te vragen hoe goed jongeren zichzelf in staat achten bepaalde taken uit te voeren met Likert-schaal antwoordmogelijkheden. Sommige meetinstrumenten identificeren hiernaast nog componenten zoals gedrag, waarden of reflectie (Schulz et al., 2008, 2016; Ten Dam et al., 2010).

1. Ontwikkeling gedurende de levensloop

Longitudinaal onderzoek in Nederland en andere landen biedt enig inzicht in de ontwikkeling van burgerschapscompetenties gedurende de levensloop, bijvoorbeeld het

COOL5-18 onderzoek in Nederland, de Citizenship Education Longitudinal Study (CELS) in Engeland, en de Youth & Society (YeS) studie in Zweden. Veel van dit onderzoek spitst zich toe op de adolescentie (circa 12 tot 18 jaar). Vanuit onderzoek is er enige steun voor de ‘formatieve jaren hypothese’ (Alwin & Krosnick, 1991; Dijkstra et al., 2004; Sears & Levy, 2003). Empirisch onderzoek naar de ontwikkeling van burgerschapscompetenties van jongeren laat zien dat deze met name in de vroege adolescentie veranderlijk zijn, en steeds stabiel worden (Eckstein, Noack, Gniewosz, 2012; Geboers et al., 2015; Keating et al., 2010; Russo & Stattin, 2017). Bij nadere beschouwing van deze onderzoeken gaat het hierbij met name om de ontwikkeling van burgerschapshoudingen. Voor burgerschapskennis geldt dat de ontwikkeling hiervan vooral samen lijkt te hangen met verschillen in cognitieve ontwikkeling (c.q. prestaties) van leerlingen, en dat burgerschapskennis van leerlingen gedurende de adolescentie gestaag toeneemt (Geboers et al., 2015). Over de ontwikkeling van burgerschapsvaardigheden zijn (behalve door meting van ‘*self efficacy*’) nauwelijks empirische gegevens beschikbaar.

2. Bijdrage van deze skills aan de ontwikkeling van andere leeropbrengsten

Over de wederzijdse beïnvloeding van de ontwikkeling van burgerschapscompetenties en andere skills is nog weinig bekend. Eidhof et al. (2017) laten zien dat burgerschapscompetenties aan het eind van de basisschool positief samenhangen met zowel taalvaardigheid als ontwikkeling in taalvaardigheid vanaf groep 5. De samenhang van burgerschapscompetenties met rekenvaardigheid en algemene intelligentie is lager. Dit lijkt te wijzen op de verwevenheid van taalvaardigheid en burgerschapscompetenties, bijvoorbeeld wanneer het gaat om het omgaan met anderen of gezamenlijk tot een beslissing kunnen komen. Meer algemeen kan worden opgemerkt dat verwacht wordt dat sociale en maatschappelijke competenties bijdragen aan schoolsucces, onder meer in de voorwaardelijke

sfeer zoals in de vorm van (op individueel niveau) zelfregulatie en prestatiemotivatie, en (op groepsniveau) via een veilig en op leren gericht schoolklimaat. Wanneer het gaat om interventies in situaties van achterstand of gemankeerde individuele ontwikkeling heeft bevordering van sociale competenties bovendien sterke positieve effecten op de sociale en cognitieve ontwikkeling van leerlingen. Bovendien is er veel evidentie dat sociale competenties bijdragen aan een succesvolle levensloop, zoals in de vorm van arbeidsmarktkansen (voor een samenvatting, zie Dijkstra, 2012).

3. Operationalisering in meetinstrumenten

In onze inventarisatie van meetinstrumenten bespreken we vijf instrumenten die in Nederland gebruikt zijn om op stelselniveau uitspraken te doen over de burgerschapscompetenties van jongeren: ADKS (Adolescentenpanel Democratische Kernwaarden en Schoolloopbanen), Burgerschap meten, ICCS (International Civic and Citizenship Education Study), PPON 45: Actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland), en PPON 49: Oordelen en argumenteren als onderdeel van actief burgerschap en sociale integratie. Deze selectie is verder aangevuld met de CELS en YeS studies die hierboven genoemd werden vanwege de inzet van longitudinale dataverzameling (vgl. Daas, Dijkstra, & de Kort, 2019). De resultaten van de instrumentanalyse zijn te vinden in de tabellen in Bijlage 1.

De instrumenten die in Nederland afgenomen zijn richten zich alle op de leeftijd 10 tot 16 jaar. De CELS en YeS meten ook op ontwikkelingen op latere leeftijd. Met uitzondering van PPON 49 gaan alle instrumenten uit van een uitsplitsing van competenties naar relevante kennis, houding en vaardigheden. Daarbij wordt kennis gemeten met behulp van (meerkeuze) toetsen, en houding en vaardigheden voor middel van zelfinschatting op vragenlijsten. PPON 49 kende een meer integratieve meting op een deeldomein, namelijk

oordelen en argumenteren. Daarbij kregen leerlingen een aantal kleine taken voorgelegd waarin zij gevraagd werden hun mening te geven en onderbouwen.

In termen van het leerlingmodel dat voor de instrumenten uitgewerkt wordt is er meestal sprake van een heldere onderbouwing van de inhoud van het competentiebegrrip, en de aanleerbaarheid hiervan bij leerlingen. Ook de taken en contexten die hierbij relevant zijn voor de uiting van de competentie zijn doorgaans voldoende onderbouwd. Een zwakte van de meeste instrumenten is dat niet of nauwelijks duidelijk is hoe de competentieontwikkeling voorspellend is voor daadwerkelijk handelen in het heden of de toekomst, en dat weinig inzicht lijkt te bestaan in de volgtijdelijke ontwikkeling van de competenties; oftewel welke ontwikkeling leerlingen veronderstelt worden te hebben doorgemaakt vooraf aan de onderzochte leeftijdsgroep.

Wat betreft het taakmodel maakt alleen PPO 49 gebruik van directe taken. Voor alle andere instrumenten geldt dat indirecte taken gebruikt worden om te infereren hoe ver de competentie van de leerling ontwikkeld is. De systematische uitwerking van burgerschapscompetenties in ‘sociale taken’ of ‘inhoudsdomeinen’ maakt dat de instrumenten wel goed scoren in termen van omschrijving van prestatiecondities en responsformat. De mate waarin daar bovenop ook beschreven wordt op welke manier variatie is aangebracht in de complexiteit van taken, om onderscheid te kunnen maken in de mate waarin de skill is ontwikkeld, wisselt tussen instrumenten.

De uitwerking van het bewijsmodel van de meeste instrumenten is vergelijkbaar. Burgerschapscompetenties worden doorgaans beschouwd als een samengesteld construct met verschillende deelaspecten, wat zich uit in het feit dat geen ‘totaalscore’ wordt gegeven, maar een rapportage van specifieke subschalen. Voor zover er normering plaatsvindt is dit vooral in vergelijking met gemiddelden voor de hele populatie. Waar wel sprake is van een absolute norm is deze ontwikkeld voor burgerschapskennis (PPO 45), maar ook hier geldt dat

vanwege het ontbreken van een extern criterium, deze betrekkelijk vrijblijvend blijft. Gegevens over de betrouwbaarheid zijn niet voor alle instrumenten beschikbaar, maar waar dat wel het geval is geldt dat de betrouwbaarheid van zowel hoofdschalen als subschalen voldoende is voor landelijke peilingen, maar ongeschikt voor individuele beoordeling van leerlingen. In het geval van houdingen of vaardigheden beoordeelt de leerling meestal zichzelf. Bij kennisopgaven bepalen ontwikkelaars vooraf het juiste antwoord.

Met betrekking tot het assemblagemodel geldt dat de meeste instrumenten voldoende duidelijk uitwerken op welke combinaties van situaties en handelingen de conceptualisering van burgerschapscompetenties betrekking heeft, en hoeveel taken of opgaven daarop betrekking hebben.

Kijkend naar het presentatiemodel zien we in oudere onderzoeken vooral afnames met ‘pen en papier’; meer recent wordt gebruik gemaakt van digitale afnames. Waar deze data vervolgens worden opgeslagen is in veel gevallen niet geheel duidelijk.

4. Taxatie van mogelijkheden tot valide meting

Voor het meten van burgerschapskennis van leerlingen tussen 10 en 16 jaar zijn verschillende instrumenten beschikbaar, waarbij (vooral) gebruik gemaakt wordt van zelfinschatting en meerkeuze-kennisvragen. Aan de Universiteit van Amsterdam vindt momenteel onderzoek plaats om het instrumentarium tevens verder uit te breiden voor bovenbouw vo en het mbo. Bovendien wordt het kennisinstrument voor de leeftijdsfase 10 - 12 jaar geactualiseerd en gekoppeld aan een model voor opbrengstgericht werken voor scholen. Ook wordt gewerkt aan aanpassing van dit instrument (Burgerschap meten) voor gebruik in het speciaal onderwijs.

Voor burgerschapskennis zijn verschillende toetsen ontwikkeld, waarin vooral gebruik wordt gemaakt van meerkeuzevragen. Op basis van de toetsen zijn ook pogingen

gedaan om met behulp van item-respons theorie een indeling naar ontwikkelingsniveaus uit te werken, die verschillende maten van (burgerschap)begrip beschrijven (ICCS & PPON 45). Omdat verondersteld wordt dat een leerling een toets maar één keer kan maken is er behoefte aan meer items of instrumenten om zo ook herhaaldelijk meten te faciliteren.

Voor het meten van burgerschapshoudingen is op basis van het huidige instrumentarium al een veelvoud aan vragen beschikbaar, waarbij steeds een set van vier tot zes vragen een houding ten aanzien van een specifiek aspect van burgerschap meet (zoals bijvoorbeeld steun voor democratische besluitvorming, of houding ten aanzien van gelijk kansen voor immigranten). De beperkingen genoemd onder burgerschapskennis zijn hier kleiner, omdat de validiteit van vragen niet (c.q. minimaal) beïnvloed wordt door herhaalde afname. De verschillende onderzoeken bieden daarbij een breed palet aan gevalideerde schalen.

Voor het meten van burgerschapsvaardigheden is weinig materiaal voorhanden. Bestaande instrumenten meten op basis van een zelfinschatting door de leerling (*self-efficacy*), en er is nauwelijks sprake van uitwerking van verschillende domeinen waarop burgerschapsvaardigheden ingezet kunnen worden. De inventarisatie van Daas, Dijkstra en de Kort (2019) laat zien dat projecten lopen waarin gepoogd wordt hier meer invulling aan te geven, maar de inzet daarvan voor gestandaardiseerde afname zal naar verwachting (op korte termijn) niet tot op grote schaal toepasbare instrumenten leiden.

Er is nauwelijks sprake van instrumentarium dat een integrale meting van burgerschapscompetenties mogelijk maakt, waarin leerlingen in (authentieke) specifieke taken zowel kennis, houding, als vaardigheden demonstreren. Het instrument t.b.v. PPON 49 meet een relevant aspect van burgerschap, maar laat tevens zien dat de ontwikkeling van dit type instrumenten weerbarstig is, dat dergelijke instrumenten schaars zijn, en dat verwerking, analyse en interpretatie niet eenvoudig is.

5. Conclusie

Op basis van bovenstaande verkenning komen we in deze paragraaf tot een korte samenvattende schets van het beeld met betrekking tot het meten van burgerschapscompetenties.

De meeste instrumenten voor burgerschapscompetenties geven een heldere onderbouwing voor de uitwerking van de doelvaardigheid (begripsvaliditeit), voor zover het losse samenstellende componenten van competentie betreft. De uitwerking hebben echter niet het karakter van (integratief) competent handelen, waarbij verschillende deelprocessen worden doorlopen om een burgerschapstaak uit te voeren. Er is in de conceptualisering en uitwerking in meetinstrumenten voldoende consensus herkenbaar over de inhoud van het begrip burgerschapscompetenties om zinvolle metingen te verrichten. De inhoudelijk verschillende uitwerkingen plaatsen vooral verschillende accenten. De verschillende uitwerkingen betekenen echter wel dat resultaten van verschillende instrumenten niet direct met elkaar vergelijkbaar zijn. Het is dus van belang een heldere afweging te maken voor welke uitwerking van burgerschapscompetenties het meest relevant wordt geacht.

De meeste instrumenten meten kennis, houdingen en vaardigheden afzonderlijk en focussen veel minder op competent handelen in taken, waarin deze losse aspecten worden gecombineerd, waar het bij de formulering van de burgerschapscompetenties juist om ging. Met name voor burgerschapshoudingen worden vervolgens diverse subschalen onderscheiden afhankelijk van het aspect of onderwerp dat bevroegd wordt. Er lijken geen gevalideerde instrumenten beschikbaar die een meer integrale meting van burgerschapscompetenties beogen, in de zin zoals hierboven beschreven. Daarnaast worden metingen zelden gerelateerd aan daadwerkelijk (toekomstig) gedrag, waardoor de criteriumvaliditeit van de meeste instrumenten onvoldoende is.

De meeste instrumenten maken bovendien gebruik van zelfbeoordeling om houding en vaardigheden te meten. 'PPON 49: Oordelen en argumenteren als onderdeel van actief burgerschap en sociale integratie' laat zien dat ook voor deze aspecten een directe meting mogelijk is, al blijkt dat een nauwgezette uitwerking te vragen (Daas, Ten Dam, & Dijkstra, 2016; Wagenaar et al., 2012). De interbeoordelaarsbetrouwbaarheid van de oordelen over houdingen op basis van toelichtingen die leerlingen geven blijkt (mede vanwege de achterblijvende kwaliteit van toelichtingen) tot op heden kwetsbaar (Daas, Dijkstra, & Karsten, 2019; Wagenaar et al., 2012).

Ten slotte zijn de meeste instrumenten ontwikkeld om burgerschapscompetenties te meten in de leeftijd 10 tot 16 jaar. Dit wordt in de literatuur ook beschouwd als een belangrijke periode in de formatie van burgerschapscompetenties. Vanwege het ontbreken van longitudinale inzichten is echter niet te beantwoorden hoe ontwikkelingen in deze periode zich verhouden tot noodzakelijke ontwikkeling in eerdere levensfasen, en wat de implicaties zijn voor opvolgende ontwikkeling. Het ontbreekt met andere woorden aan empirisch gegevens voor de uitwerking van een helder ontwikkelingsmodel dat inzicht geeft in de (normale) ontwikkeling van burgerschapscompetenties gedurende de levensloop.

Digitale geletterdheid

De reviewstudie van Voogt et al. (2019) laat zien dat in de literatuur (nog lang) geen overeenstemming bereikt is over de definitie van digitale vaardigheden, als ook welke kennis, houding en vaardigheden daar onderdeel van uitmaken. De auteurs verbinden daar de conclusie aan dat het leergebied nog in ontwikkeling is, zoals ook te zien is aan de verschillende begrippen of componenten die aan het concept verbonden worden (van Laar et al., 2017). Er is gezien die onduidelijkheid rond begripsafbakening ook geen duidelijkheid welke competenties relevant zijn voor leerlingen in de Nederlandse onderwijscontext (Voogt et al. 2019). De auteurs zijn bovendien kritisch op de mate waarin meetinstrumenten voor digitale geletterdheid rekening houden met de tijdsafhankelijkheid van het concept door veranderingen in technologie.

Het raamwerk van DigComp (Ferrari, 2013) beschrijft vijf domeinen waarbinnen doelen geformuleerd zijn voor de ontwikkeling van digitale geletterdheid. De review van Voogt laat – net als een eerdere overzichtsstudie (Voogt & Pareja Roblin, 2012) – zien dat de verschillende conceptuele uitwerkingen van digitale geletterdheid op verschillende manieren overlappen. Ook de uitwerking die in het kader van Curriculum.nu gedaan werd overlapt (grotten-)deels (zie Kader 1).

Omdat digitale geletterdheid als competentie beschouwd wordt, kan voor de uitwerking van het concept een matrix opgesteld worden met enerzijds de aan competentie onderliggende elementen kennis, houding en vaardigheden, en anderzijds domeinen of toepassingsgebieden. De review van Voogt et al. (2019) laat zien dat de meeste studies naar (aspecten van) digitale geletterdheid focussen op vaardigheden. Die oververtegenwoordiging wordt verklaard doordat digitale geletterdheid vaak gerelateerd wordt aan “het vermogen om iets te ‘kunnen’ doen met de technologie als middel om bepaalde doelen te bereiken” (ibid., p. 36), waarbij de focus op ‘kunnen’ verraadt dat vaardigheden meer aandacht krijgen.

Wanneer kennis en houding onderzocht worden gebeurt dat doorgaans met afzonderlijke kennisvragen of een vragenlijst.

Kader 1. Raamwerk digitale geletterdheid DigComp (Ferrari, 2013) & Curriculum.nu

INFORMATIE- EN DATAGELETTERDHEID

- 1.1 Navigeren, zoeken en filteren van gegevens, informatie & digitale inhoud
- 1.2 Evalueren van gegevens, informatie & digitale inhoud
- 1.3 Beheer van gegevens, informatie & digitale inhoud

COMMUNICATIE EN SAMENWERKING

- 2.1 Interactie via digitale technologieën
- 2.2 Delen via digitale technologieën
- 2.3 Burgers betrekken bij burgerschap via digitale technologieën
- 2.4 Samenwerken via digitale technologieën
- 2.5 Netiquette
- 2.6 Beheer van de digitale identiteit

CREËREN VAN DIGITALE INHOUD

- 3.1. Ontwikkeling van digitale inhoud
- 3.2 Integreeren & bewerken digitale inhoud
- 3.3 Auteursrechten & licenties
- 3.4 Programmeren

VEILIGHEID

- 4.1 Beveiligingsvoorzieningen
- 4.2 Bescherming persoonsgegevens & privacy
- 4.3 Bescherming gezondheid & welzijn
- 4.4 Bescherming milieu

PROBLEEMOPLOSSING

- 5.1 Technische problemen oplossen*
- 5.2 Vaststelling van behoeften & technologische antwoorden
- 5.3 Creatief gebruik van digitale technologieën
- 5.4 Vaststellen van leemtes in digitale competenties*

ANDERE DOELEN

- 1. Bedienen en kiezen van toepassingen**
- 2. De technische werking van technologie**
- 3. Reflectie op gevolgen van technologie voor zichzelf en de samenleving**

* Deze doelen zijn volgens Voogt et al. (2019) niet onderdeel van Curriculum.nu-po bovenbouw.

** Deze doelen zijn volgens Voogt et al. (2019) niet onderdeel van het DigComp raamwerk, maar wel van Curriculum.nu-po bovenbouw.

1. Ontwikkeling van deze skills gedurende de levensloop

Argumenten voor het belang van (het meten van) digitale geletterdheid spitsen zich doorgaans toe op het belang van deze skill naar aanleiding van technologische

ontwikkelingen; niet op evidentie over inzichten in de werving van deze competenties door jongeren gedurende de schoolloopbaan. De meeste meetinstrumenten voor digitale geletterdheid richten zich op de bovenbouw van het primair onderwijs of de onderbouw van het voortgezet onderwijs, maar het is onvoldoende duidelijk of dit ook de leeftijd is waarop digitale competenties zich ontwikkelen en beïnvloedbaar zijn. Doordat instrumenten voor andere leeftijdsgroepen niet of nauwelijks beschikbaar zijn, lijkt het voorlopig nog niet goed mogelijk om meer inzicht te ontwikkelen in de ontwikkeling van digitale geletterdheid over langere termijn.

2. Bijdrage van deze skills aan de ontwikkeling van andere leeropbrengsten

De vele conceptuele uitwerkingen van digitale geletterdheid maken dat het concept aan een veelvoud van begrippen gerelateerd wordt; niet in de laatste plaats aan 21st century skills (vgl. van Laar et al., 2017). Het lijkt hier echter alleen te gaan om conceptuele verbanden; van operationalisering in meetinstrumenten en empirische validering lijkt geen sprake. In algemene zin lijken er dus vele verbanden mogelijk, zoals ook vaak terug te zien in het feit dat technologische ontwikkelingen als één van de voornaamste aanjagers beschouwd wordt voor het belang van 21st century skills. ICILS 2013 liet zien dat jongeren op hogere onderwijsniveaus ook beter scoren op digitale geletterdheid, wat lijkt te wijzen op een verband tussen digitale geletterdheid en cognitieve ontwikkeling (Meelissen et al., 2014). Het gebrek aan empirische gegevens maakt echter dat hierover vooralsnog geen gestaafde uitspraken mogelijk zijn.

3. Operationalisering in meetinstrumenten

Op basis van de overzichtsstudie van Voogt et al. (2019) bekeken we veertien instrumenten die in die studie naar voren komen. Details van de analyses zijn terug te vinden in Bijlage 1. Het gaat om de ST²L (Student Tool for Technology Literacy), Performance

based test ICT competences, KERIS 2014 (Korea Education Research and Information Service, K-12 test), ICILS, ICT literacy test, New ICT Literacy test, MMTL (Methods for Measuring Technological Literacy), ICT literacy assessment, Effectiveness study: digital literacy program, NAP-ICT Literacy (National Assessment Program: ICT Literacy), Measurement of actual digital literacy, TEL Framework (Technology and Engineering Literacy), Authentic assessment online information literacy, en Testen digitale geletterdheid. De twee laatstgenoemde instrumenten zijn in Nederland ontwikkeld, de anderen verspreid over de hele wereld. De resultaten van de instrumentanalyse zijn te vinden in Bijlage 1.

Alle instrumenten zijn toegespitst op leerlingen in de bovenbouw basisonderwijs en/of onderbouw voortgezet onderwijs. Alleen KERIS onderzoekt ook de digitale geletterdheid van jongere leerlingen. Bij alle instrumenten speelt het meten van vaardigheden een grote rol, en veelal wordt daarbij gebruik gemaakt van directe meting, dat wil zeggen met behulp van taken waarin leerlingen rechtstreeks hun kunnen tonen. Daarbij verschillen instrumenten wel in hun conceptuele uitwerking. Sommige instrumenten maken gebruik van taken waarbij sprake is van integratie van kennis en vaardigheden. Bij andere instrumenten worden taken ingezet om losse kennis en vaardigheden te meten. Weer andere instrumenten bevragen naast de directe meting van competenties of vaardigheden tevens de attitudes van leerlingen in een separate vragenlijst.

Wat betreft het onderliggende leerlingmodel aan de instrumenten verschilt de kwaliteit van de conceptuele helderheid ervan tussen instrumenten. Tevens valt op dat veel instrumenten van verschillende concepten uitgaan en dat niet geheel duidelijk is hoe deze concepten zich tot elkaar verhouden (zoals *technology literacy*, *digital literacy*, *engineering literacy* en *ICT literacy*). Zoals hierboven ook aangegeven is, lijkt onderzoek naar digitale geletterdheid vaak aangejaagd te worden door veranderende (technologische) omstandigheden. Conceptuele uitwerkingen van hoe jongeren digitale geletterdheid

ontwikkelen in termen van zowel aanleerbaarheid als ontwikkelingsstadia staan minder centraal. De uitwerking daarvan ontbreekt niet geheel, maar laat regelmatig wel te wensen over. In het verlengde daarvan is ook de criteriumvaliditeit van een groot deel van de instrumenten matig te noemen.

Wat betreft het taakmodel valt op dat de meerderheid van de instrumenten gebruik maakt van directe taken. Daarbij wisselt tussen instrumenten nog wel in welke mate uitgewerkt is hoe de taken relevant en representatief zijn voor het doeldomein. Hetzelfde geldt voor de mate waarin prestatiecondities zijn beschreven en hoe variatie is aangebracht in de complexiteit van taken. Enkele instrumenten zoals ICILS en NAP-ICT Literacy vallen in positief opzicht op vanwege de nauwgezette theoretische onderbouwing die aan de instrumenten gegeven wordt.

Bij de uitwerking van het bewijsmodel zien we vergelijkbare verschillen tussen instrumenten voor wat betreft de mate waarin deze een onderbouwde rationale geven voor hoe scores van leerlingen gecombineerd worden tot totaalscores voor de doelvaardigheid. Ongeveer de helft van de instrumenten rapporteert (relatieve) normen voor de scores van leerlingen, maar alleen NAP-ICT Literacy geeft hiervoor een goede onderbouwing die duidelijk maakt wat de mogelijkheden en beperking van de normen zijn. Zoals ook Siddiq et al. (2016) en Voogt et al. (2019) aangeven ontbreekt het voor een aanzienlijk deel van de instrumenten aan rapportage van betrouwbaarheden van verkregen (totaal)scores. De instrumenten waarvoor deze wel gerapporteerd worden voldoen doorgaans aan betrouwbaarheidseisen op dit vlak; zeker wanneer het gaat om metingen ten behoeve van peilingsonderzoek, waarbij minder stringente eisen gelden dan wanneer op grond van scores belangrijke beslissingen moeten worden genomen over individuele leerlingen.

Met betrekking tot het assemblagemodel zien we dat bij de helft van de instrumenten de specificering van de taaksituaties waarin de leerling de vaardigheid moet demonstreren te

wensen overlaten. Datzelfde geldt voor de mate waarin de taaksituaties representatief zijn voor het brede doeldomein. Het is dan ook de vraag of over het doeldomein en de daarbij behorende specificaties van beoogde taken en taaksituaties wel voldoende duidelijkheid bestaat. Voor de meeste instrumenten lijkt bovendien geen handleiding beschikbaar.

Met betrekking tot het presentatiemodel zal het tot slotte niet verbazen dat de meeste instrumenten afgenomen worden op de computer. Een deel van de onderzoeken stelt de geanonimiseerde data beschikbaar voor secundaire analyses.

4. Taxatie van mogelijkheden tot valide meting

Bij meetinstrumenten voor digitale geletterdheid wordt doorgaans de nadruk op het meten van vaardigheden (in concrete taaksituaties). Voor het meten van kennis en houding is binnen de conceptuele uitwerkingen van digitale geletterdheid weinig aandacht. Dat verklaart ook de beperkte mate waarin deze componenten van de skill in meetinstrumenten uitgewerkt zijn. Bij het meten van kennis worden doorgaans meerkeuzeopgaven gebruikt, waarbij door het gebruik van authentieke (digitale) toetsomgevingen een focus lijkt te zijn op (het toepassen van) procedurele kennis (vgl. Fraillon et al., 2014; Heitink, 2018). Het meten van houdingen gebeurt vooral aan de hand van stellingen met antwoordmogelijkheden op een Likert-schaal. Een grotere aandacht voor authentiek-integratieve taken, waarbij de leerling naast procedurele vaardigheden, ook kennis en houdingaspect van het handelen kan tonen, vraagt om een andere conceptualisering van digitale geletterdheid als integratieve advanced skill. Die ontwikkeling is nog niet waargenomen in de gescreende instrumenten.

De in de reviewstudies van Voogt et al. (2019) en Siddiq et al. (2016) opgenomen instrumenten maken (vrijwel) alle gebruik van een computergestuurde omgeving voor het meten van digitale geletterdheid. Daarbij wordt zowel gebruik gemaakt van meerkeuze opgaven, alsook ‘constructed response’ items waar de acties van de student op een nog

directere wijze gemeten en beoordeeld worden. Vooral oudere instrumenten meten vaardigheden indirect met behulp van meerkeuze vragen of door zelfinschatting (Siddiq et al., 2016). Bij instrumenten waar de vaardigheden van de student beoordeeld worden op basis van interactie met de digitale omgeving vindt beoordeling (deels) handmatig plaats door een beoordelaar, soms door gebruik van rubrics. De interbeoordelaarsbetrouwbaarheid lijkt bij deze instrumenten voldoende gewaarborgd.

5. Conclusie

De conceptuele onduidelijkheid rond digitale geletterdheid bemoeilijkt vooralsnog een heldere uitwerking van de inhoud ervan, en als gevolg daarvan ook de meting van competent handelen door leerlingen. Met name de veelheid aan modellen en de opdeling in subcomponenten leidt tot onduidelijkheid. Voor de meeste modellen lijkt te gelden dat ze opgebouwd zijn naar aanleiding van ontwikkelingen in (gebruik van) technologieën; veel minder aandacht gaat uit naar de manier waarop leerlingen veronderstelt worden digitale geletterdheid te ontwikkelen, en hoe deze ontwikkeling begrepen kan worden.

Een positiever beeld ontstaat wanneer gekeken wordt naar de ontwikkeling van meetinstrumenten. Veel van de instrumenten die we in onze verkenning meenamen zijn recent ontwikkeld, wat lijkt te duiden op een hoge mate van ontwikkelingen op dit terrein. Daarbij zien we ook dat verschillende instrumenten expliciet beogen om directe metingen van digitale geletterdheid te realiseren. Aansluitend bij de aandacht die er in de conceptualisering van digitale geletterdheid is voor vaardigheden van leerlingen, zien we ook in instrumenten veel aandacht voor deze component.

Kijkend naar de mate van onderbouwing bij de verschillende instrumenten worden wel verschillen duidelijk. Met name ICILS en NAP-ICT Literacy vallen op in de relatief uitgebreide onderbouwingen die aan deze instrumenten ten grondslag liggen. De

instrumenten laten ook mooie voorbeelden zien van directe taken die leerlingen vervullen in gesimuleerde digitale omgevingen. In dat opzicht biedt digitale afname van dit soort instrumenten nog meer voordelen dan het voor de andere skills in deze verkenning ook lijkt te bieden, omdat de digitale simulatie inherent verbonden is aan de doelvaardigheden die gemeten worden.

Sociaal-emotionele ontwikkeling

Het domein sociaal-emotionele ontwikkeling wordt vooral vanuit de ontwikkelingspsychologie veelvuldig bestudeerd. Er is dan ook een veelheid aan instrumenten voor sociaal-emotionele ontwikkeling gericht op jonge kinderen en op het signaleren van afwijkende / problematische ontwikkeling (Halle & Darling-Churchill, 2016). Tegelijkertijd wordt geconstateerd dat er weinig consensus is over wat sociaal-emotionele ontwikkeling precies inhoudt, hoe deze ontwikkelingen het best gemeten kunnen worden, en of bestaande meetinstrumenten niet vooral persoonlijkheid meten (Darling-Churchill & Lippman, 2016; Wigelsworth et al., 2010). Een veelgebruikte indeling van CASEL (Collaborative for Academic Social and Emotional Learning; www.casel.org) gaat uit van vijf componenten: zelfbewustzijn, sociaal bewustzijn, besluiten kunnen nemen, zelfmanagement, en het vermogen relaties te hanteren. Instrumenten voor sociaal-emotionele ontwikkelen richten zich op verschillende aspecten daarvan (Halle & Darling-Churchill, 2016).

1. Ontwikkeling van deze skills gedurende de levensloop

Veel onderzoek naar sociaal emotionele ontwikkeling richt zich op de ontwikkeling van jonge kinderen voordat zij naar school gaan. Over de psychosociale ontwikkeling die jonge kinderen doormaken lijkt over het algemeen consensus. Jonge kinderen worden zich bewust van hun omgeving, van hun lichaam, gaan sociale relaties aan, en beginnen hun identiteit te vormen. Tijdens de kindertijd worden verschillende levensfasen doorlopen (baby, peuter, kleuter, schoolkind, puber). Deze levensfasen hangen ook sterk samen met de sociaal-emotionele ontwikkeling. Vanwege de breedte van deze ontwikkeling valt het buiten het bestek van deze verkenning om die in detail te beschrijven.

2. Bijdrage van deze skills aan de ontwikkeling van andere leeropbrengsten

Van met name sociaal-emotionele ontwikkeling op jonge leeftijd is veel bewijs dat deze op latere leeftijd bijdraagt aan een veelheid van wenselijke uitkomsten (Halle & Darling-Churchill, 2016). Ook voor latere leeftijden laat onderzoek zien dat sociaal-emotionele ontwikkeling bijdraagt aan schooluitkomsten (Durlak et al., 2011). Hierbij wordt verondersteld dat betere sociaal-emotionele ontwikkeling leidt tot meer wenselijk gedrag in de klas, wat vervolgens bijdraagt aan betere leeropbrengsten. De meta-analyse van Durlak et al. (2011) laat tevens zien dat programma's voor sociaal-emotionele ontwikkeling hieraan kunnen bijdragen.

3. Operationalisering in meetinstrumenten

Verschillende meetinstrumenten zijn relevant om in ogenschouw te nemen voor het meten van sociaal-emotionele ontwikkeling vanwege de focus op 'sociaal functioneren in de klas'. We nemen hierbij de instrumenten SCOL (Sociale Competentie Observatie Lijst), Viseon 2.0 (Volginstrument Sociaal-Emotionele Ontwikkeling), Zien!, Kijk! op sociale competentie, SVL/SAQI (SchoolVragenLijst / School Attitude Questionnaire Internet), en Kanjervragenlijst mee. Deze instrumenten zijn veelal onderdeel van een leerlingvolgsysteem en komen ook in eerdere inventarisaties naar voren (Inspectie van het Onderwijs, 2016, Ledoux et al., 2013). De CBSA (Competentie belevingsschaal voor adolescenten) is als enige specifiek ontwikkeld voor het voortgezet onderwijs. Daarnaast evalueren we ook twee instrumenten die ontwikkeld zijn in de Verenigde Staten en gericht zijn op directe meting van sociaal-emotionele ontwikkeling. Zowel SELweb als Zoo U maken gebruik van een digitale omgeving om directe meting te faciliteren. De resultaten van de instrumentanalyse zijn te vinden in de tabellen in Bijlage 1.

De meeste instrumenten hebben als doelgroep leerlingen uit de bovenbouw van het primair onderwijs, en zijn in veel gevallen ook bruikbaar voor de onderbouw. De SCOL en SAQI zijn tevens ontwikkeld om ook in de onderbouw van het voortgezet onderwijs gebruikt te kunnen worden. De CBSA is gericht op het voortgezet onderwijs. De beide instrumenten met directe metingen zijn ontwikkeld voor de onderbouw van het primair onderwijs. Vrijwel alle instrumenten bestaan uit een samenstelling van meerdere elkaar aanvullende instrumenten; in veel gevallen een combinatie van een vragenlijst voor leerlingen en een vragenlijst voor docenten.

Kijkend naar het leerlingmodel, kennen de meeste instrumenten een heldere conceptuele uitwerking van wat onder de sociaal-emotionele ontwikkeling van leerlingen verstaan wordt. De verdere uitwerking van het leerlingmodel laat in de meeste gevallen echter te wensen over. Er wordt maar in beperkte mate uitgewerkt hoe de leerling zich op aspecten van deze sociaal-emotionele competentie ontwikkelt. Ook wordt de meting door de meeste instrumenten niet gerelateerd aan externe criteria om de criteriumvaliditeit te onderbouwen. Bovendien wordt in de meeste gevallen niet specifiek uitgewerkt welke (sociaal-emotionele) taakprocessen de leerling in welke situaties zou moeten leren toepassen.

Wat betreft het taakmodel gaan bijna alle instrumenten uit van een combinatie van vragenlijsten waarin de te beantwoorden gedragsstellingen slechts beperkt zijn gespecificeerd naar situaties en taken. Het is in de meeste gevallen niet duidelijk op welke wijze verondersteld wordt hoe een bepaalde mate van ontwikkeling zich verhoudt tot gedrag in een bepaalde context. Ook is niet duidelijk of en in welke mate variatie aangebracht wordt in de complexiteit of belasting die taken of prestatiecondities met zich meebrengen in de toepassing van sociaal-emotionele competentie. Alleen voor de instrumenten waarin leerlingen met direct taken worden geconfronteerd is dit helderder uitgewerkt.

Bij de uitwerking van het bewijsmodel wordt voor de meeste instrumenten een onderbouwing gegeven voor hoe subschalen worden gecombineerd tot een totaalscore. De instrumenten waarvoor dat niet gebeurt rapporteren alleen scores voor de subschalen. De drie instrumenten waarvoor interbeoordelaarsbetrouwbaarheid gerapporteerd wordt voor de docentenvragenlijst laten een matige overeenstemming zien, maar voldoende voor een instrument dat als doel heeft ontwikkelingsproblemen te signaleren. De meeste instrumenten kennen vanwege dat doel ook een relatieve normscore, waarbij vooral problematische ontwikkeling aangeduid wordt, uitgaand van een normgroep van leerlingen van dezelfde leeftijd. Eventuele beperkingen van een dergelijke normschaal worden niet door alle instrumenten even duidelijk toegelicht. Vrijwel alle instrumenten rapporteren betrouwbaarheden van de schalen, en die zijn doorgaans voldoende betrouwbaar voor minder belangrijke individuele beslissingen en voor uitspraken op stelselniveau.

Voor het assemblagemodel ontstaat het volgende beeld. De aard van de taaksituaties waarin leerlingen de doelvaardigheden moeten laten zien is maar beperkt gespecificeerd. Omgekeerd geldt dat wel onderbouwd is hoe de taaksituaties uitgewerkt zijn om generaliseerbaar te zijn naar andere relevante situaties (in de meeste gevallen sociaal-emotioneel gedrag op school). Bij de instrumenten die gebruik maken van directe taken zijn de specificaties van de te dekken taken en taaksituaties doorgaans nauwkeuriger uitgewerkt.

Tot slot het presentatiemodel. In de meeste gevallen wordt het instrument door zowel leerlingen als docenten op de computer ingevuld. Het is vaak onduidelijk waar en hoe de gegevens opgeslagen worden. Voor de meeste instrumenten zijn handleidingen beschikbaar.

4. Taxatie van mogelijkheden tot valide meting

Voor het meten van sociaal-emotionele ontwikkeling wordt veel gebruik gemaakt van zelfrapportage en het invullen van observatieformulieren door docenten. Een fundamentele

vraag die zich hierbij aandient is of deze aanpak, die vooral gericht lijkt te zijn op het meten van ‘kenmerkend gedrag’, valide informatie oplevert om uitspraken te doen over competentie, waarbij het vooral gaat om ‘maximaal gedrag’ (Wigelsworth et al., 2010). Instrumenten voor sociaal-emotionele ontwikkeling die uitgaan van vragenlijsten lijken vooral toegespitst op kenmerkend gedrag dat leerlingen in alledaagse situaties laten zien. Deze instrumenten lijken vooral als doel te hebben om leerlingen die (risico op) problematisch gedrag vertonen te identificeren.

Het meten van sociaal-emotionele ontwikkeling met behulp van directe meting lijkt meer geschikt voor het meten van maximaal gedrag, door een situatie te simuleren waarin de leerling de benodigde competenties moet inzetten. Het gesimuleerde karakter maakt het echter moeilijk om te generaliseren naar andere situaties, omdat de leerling zich in de meeste gevallen – en zeker waar het computersimulaties betreft – zal realiseren dat het niet om een authentieke situatie gaat. Het instrument Zoo U probeert hier een oplossing voor te bieden door te verbergen welke gedragingen van leerlingen beoordeeld worden.

5. Conclusie

De conceptuele modellen voor sociaal-emotionele ontwikkeling verschillen in de ontwikkelingsgebieden die daartoe gerekend worden (Darling-Churchill & Lippman, 2016; Wigelsworth et al., 2010). Ook de begrippen die daarbij gehanteerd worden kennen verschillende definities of gebruiken verschillende begrippen voor hetzelfde construct (Jones et al., 2016). Een van de moeilijkheden die zich daarbij voordoet is dat de relatie met persoonlijkheidskenmerken niet altijd duidelijk is en onduidelijk blijft in hoeverre deze onderdeel uitmaken van het sociaal-emotionele ontwikkelingsmodel dat wordt verondersteld (Wigelsworth et al., 2010).

Voor het meten van sociaal-emotionele ontwikkeling van leerlingen lijken, ook in internationaal perspectief, veel instrumenten beschikbaar. Een inventarisatie van CASEL biedt een aanzienlijk palet aan instrumenten (vgl. Ledoux et al., 2013). Bovendien zijn instrumenten toegespitst op verschillende gebruiksdoelen (McKown, 2019). Zo zien we dat instrumenten die gebruikt worden in leerlingvolgsystemen vaak gericht zijn op signalering van afwijkende sociaal-emotioneel gedrag van leerlingen, om vroegtijdig nader onderzoek te doen of ondersteuning te bieden.

De meeste instrumenten voor het meten van sociaal-emotionele ontwikkeling maken gebruik van zelfrapportage of docentenrapportage (McKown, 2019). Het gaat in die gevallen dus om indirecte metingen, die afhankelijk zijn van de kwaliteit van de inschatting van de beoordelaar. Deze afhankelijkheid is problematisch omdat onderzoek van Kankaraš et al. (2019) laat zien dat er een sterker effect uitgaat van wie de sociaal-emotionele ontwikkeling beoordeelt (hier leerling, leraar of ouder) dan van welk aspect van sociaal-emotionele ontwikkeling beoordeeld wordt.

SELweb en Zoo U laten zien dat digitale instrumenten interessante mogelijkheden bieden voor directe metingen van de sociaal-emotionele ontwikkeling van leerlingen. Hiermee komt ook meer focus te liggen op de capaciteiten van leerlingen, dan op typisch gedrag wat zij laten zien. Een probleem dat zich voordoet is tenslotte in hoeverre gedrag in digitale platformen zich laat vertalen naar uitspraken over competentie in de sociale werkelijkheid daarbuiten.

Zelfregulatie

Het is moeilijk een eenduidige definitie te geven van zelfregulatie. Binnen de onderzoeksliteratuur staan (vooral) de begrippen metacognitie, zelfregulatie en zelfregulerend leren centraal, en in aanverwant (vooral cognitief-psychologisch / neurologisch) onderzoek ook het begrip executieve functies. De uitwerkingen van deze begrippen zijn niet altijd eenduidig en de begrippen kennen verschillende onderzoeksachtergronden. Grofweg kan gesteld worden dat metacognitie met name vanuit de cognitieve psychologie onderzocht is, en zich richt op cognitieve processen. Metacognitie kent doorgaans twee componenten: kennis over cognitie en sturing van cognitie, hoewel de gehanteerde terminologie kan wisselen (Lai, 2011). Zelfregulatie en zelfregulerend leren hebben een meer sociaalpsychologische of pedagogische achtergrond, waarbij een belangrijk onderscheid met metacognitie is dat zelfregulatie zich meer op de interactie van het individu met de omgeving richt (Opdenakker & Ledoux, 2016). Zelfregulerend leren kan gezien worden als een verbijzondering van zelfregulatie in onderwijssettingen (Dinsmore, Alexander, & Loughlin, 2008). Ook executieve functies kennen verschillende definities en invulling, maar kunnen beschouwd worden in het verlengde van zelfregulatie in zoverre dat ze uitgaan van de interactie tussen individu en omgeving, en zijn toegespitst op de zelfregulatie van cognitieve processen; in het bijzonder inhibitie, werkgeheugen en cognitieve flexibiliteit (Diamond, 2013; Hofmann, Schmeichel, & Baddeley, 2012).

Ook voor de bovengenoemde begrippen geldt dat niet altijd helder is wat er onder verstaan wordt. Dat de definities van deze begrippen niet altijd even helder zijn blijkt ook uit de review van Dinsmore et al. (2008): in ongeveer een derde tot de helft van de empirische studies die zij onderzochten naar zelfregulatie en zelfregulerend leren werd het centrale begrip niet expliciet gedefinieerd. Dit bleek zelfs in twee derde van de studies naar metacognitie het geval.

1. Ontwikkeling van deze skills gedurende de levensloop

Eerder werd vooral aangenomen dat jonge kinderen nog niet in staat zijn hun eigen leren te reguleren, en richtte meting van zelfregulatie zich vooral op leerlingen (en studenten) in het voortgezet onderwijs of daarna (Winne & Perry, 2000). Recent onderzoek laat zien dat ook jongere kinderen zelfregulatie toepassen, en dat ook zelfregulatie van leerlingen in het primair onderwijs te meten is (Vandevelde et al., 2013). Voor metacognitie geldt dat kinderen tussen 3 en 6 jaar grote ontwikkelingen door lijken te maken (Lai, 2011). Onderzoek van Van der Stel en Veenman (2014) laat zien dat de ontwikkeling van metacognitie tussen 12 en 15 jarige leeftijd niet lineair verloopt. De resultaten lieten tevens zien dat domeinspecifieke metacognitieve vaardigheden van leerlingen rond hun 15^e meer generiek worden. Samenvattend kan gesteld worden dat er verschillende inzichten zijn in de ontwikkeling van zelfregulatie, maar dat het beeld gefragmenteerd is.

2. Bijdrage van deze skills aan de ontwikkeling van andere leeropbrengsten

Voor alle aan zelfregulatie verwante begrippen die hierboven beschreven worden geldt dat er een aanzienlijke hoeveelheid evidentie is dat zelfregulatie als skill bijdraagt aan andere leeropbrengsten (vgl. Hattie, 2009; Lai, 2011). De overzichtsstudie van Diamond (2013) naar executieve functies toont effecten aan op mentale gezondheid, fysieke gezondheid, lees- en rekenvaardigheid et cetera. Van der Stel en Veenman (2014) laten zien dat metacognitie bijdraagt aan leerresultaten voor geschiedenis en wiskunde, ook wanneer gecontroleerd wordt voor intelligentie (de grootte van het effect lijkt echter tussen vakken, en afhankelijk van leeftijd, te wisselen). Rademacher en Koglin (2019) laten zien dat de verschillende uitwerkingen van zelfregulatie het moeilijk maken om heldere uitspraken te doen, maar dat in algemene zin gesteld kan worden dat zelfregulatie van jonge kinderen bijdraagt aan hun sociaal-emotionele ontwikkeling.

In een meta-analyse van 61 studies onderzoeken Dent en Koenha (2016) de effecten van zelfregulerend leren op schoolprestaties in primair en voortgezet onderwijs. Zij maken daarbij onderscheid tussen cognitieve strategieën, zoals uitleggen en herhalen, en metacognitieve processen: doelen stellen, plannen, zelfmonitoren, zelfcontrole en zelfevaluatie. Zowel in primair als voortgezet onderwijs dragen cognitieve strategieën en metacognitieve processen bij aan leeropbrengsten. De correlaties zijn hoger voor metacognitieve processen. De auteurs zijn echter terughoudend in het formuleren van implicaties op basis van deze conclusie, omdat onderzochte cognitieve strategieën vaak oppervlakkige aanpakken zijn, waardoor wellicht geen grote effecten te verwachten zijn. ‘Diepere’ cognitieve strategieën hadden een groter effect dan meer oppervlakkige strategieën. Metacognitieve processen hebben het meest effect wanneer sprake is van een samengestelde meting. Van de specifieke processen heeft plannen het grootste effect op leeruitkomsten. Voor beide aspecten geldt dat effecten verschillen tussen schoolvakken, waarbij effecten doorgaans groter zijn voor mens- en maatschappijvakken dan voor wiskunde. De relatie tussen cognitieve strategieën en leeruitkomsten neemt geleidelijk toe in de loop van het primair en voortgezet onderwijs. Voor metacognitieve processen werd geen duidelijk patroon gevonden.

3. Operationalisering in meetinstrumenten

De meest gebruikte instrumenten om zelfregulatie te meten werken met zelfrapportages. Daarnaast worden ook observaties, hardop denk protocollen, interviews, zelfinschatting (van verwachte leerprestaties) en dagboeken gebruikt. Voor zelfregulatie en metacognitie zijn diverse vragenlijsten voor leerlingen en leerkrachten beschikbaar. We bespreken vier veelgebruikte instrumenten: de MSLQ (Motivated Strategies for Learning Questionnaire), CP-SRLI (Children’s Perceived use of Self-Regulated Learning Inventory),

MAI (Metacognitive Awareness Inventory), en jr.MAI (junior Metacognitive Awareness Inventory).

Vragenlijsten worden veel bekritiseerd omdat directe metingen van metacognitie een duidelijker relatie vertonen met leeruitkomsten. Vaak blijken de directe metingen (aan de hand van een concrete taak) nauwelijks samen te hangen met indirecte metingen. Om die reden wordt in de literatuur de voorkeur gegeven aan directe metingen (Veenman & van Cleef, 2019). Daarbinnen zijn ook weer uiteenlopende ervaringen met verschillende methoden van dataverzameling. Zo heeft het gebruik van digitale ‘logfiles’ als voordeel dat verwerking ervan minder arbeidsintensief is dan de hardop denk methode (Van Hout-Wolters, 2009). We bespreken daarom ook het nStudy en gStudy instrument. Deze gaan beide uit van dezelfde aanpak, namelijk het digitaal ‘loggen’ van alle handelingen van een leerling bij het studeren of andere aan studie gerelateerde activiteiten. Het zijn in die zin niet zozeer instrumenten waar direct een beoordeling uit voortkomt, maar een taakstructuur en daarbij behorende registratie van taakaanpak door de leerling.

Executieve functies worden doorgaans gemeten met behulp van directe taken in een laboratoriumsetting (Diamond, 2013). We nemen in onze inventarisatie het AMES (Assessment of Motivation, Effort, and Self-regulation) instrument mee, waarbij leerlingen in een klassetting een aantal directe taken krijgen voorgelegd, aangeboden op een tablet. De resultaten van de instrumentanalyse zijn te vinden in de tabellen in Bijlage 1.

De MSLQ en MAI zijn ontworpen voor studenten in het hoger onderwijs. De MSLQ wordt daarnaast ook gebruikt in het voortgezet onderwijs. De CP-SRLI, jr.MAI en AMES zijn ontworpen voor bovenbouw primair onderwijs, en jr. MAI tevens voor onderbouw voortgezet onderwijs. gStudy en nStudy zijn vanwege hun opzet voor alle leeftijden geschikt. De CP-SRLI is in Vlaanderen ontwikkeld, gStudy en nStudy in Canada. Alle andere instrumenten zijn de in VS ontwikkeld.

Kijkend naar het leerlingmodel zien we onduidelijkheid terug over wat zelfregulatie precies inhoudt, de vraag of de ontwikkeling van vaardigheden domeinspecifiek of generiek verloopt, en hoe de ontwikkeling van zelfregulatie bij leerlingen (in de tijd) verloopt, resp. welke leerprogressies optreden. Omdat bij het meten van zelfregulatie vaak ook andere leerresultaten meegenomen worden is de criteriumvaliditeit doorgaans wel voldoende of zelfs goed. De meeste instrumenten lijken te veronderstellen dat met de verschillende componenten die gemeten worden de complete doelvaardigheid representeren. Gezien de vele begrippen en definities is het de vraag in hoeverre dat het geval is.

Wat betreft het taakmodel maken de meeste instrumenten gebruik van indirecte meting (namelijk vragenlijsten). Wat opvalt is dat de onderbouwing van de selectie van relevant domeinspecifiek of generiek gedrag vaak weinig onderbouwd is. Ook wordt vaak niet omschreven op welke manier variatie is aangebracht in de complexiteit van (bevroegde) taken. Daarbij valt te denken aan taken of taakomstandigheden waarin het gemakkelijker of juist moeilijker is om zelfregulatie te tonen.

Omdat de instrumenten geen totaalscore geven voor de doelvaardigheid, en geen normen gegeven worden om betekenis te geven aan de scores van leerlingen, blijft dat deel van het bewijsmodel nagenoeg leeg. Aan de psychometrische eigenschappen wordt doorgaans veel aandacht besteed en de instrumenten scoren daar goed op. Voor gStudy en nStudy ontbreken deze gegevens vanwege de aard van het instrument. Deze leveren immers geen beoordelingsscores op. Voor alle instrumenten met vragenlijsten is de leerling degene die zichzelf beoordeelt. Voor de directe instrumenten beoordeelt de onderzoeker en/of een algoritme de prestatie.

Met betrekking tot het assemblagemodel zien we dat voor de meeste instrumenten geen handleiding beschikbaar is. Voor de vragenlijsten is matig tot onvoldoende uitgewerkt in welke situaties de doelvaardigheid moet worden getoond. Voor de directe instrumenten is

dit wel voldoende uitgewerkt. Het is voor de meeste instrumenten wel voldoende aannemelijk dat resultaten gegeneraliseerd worden naar relevante situaties (veelal studiegedrag voor school of opleiding).

Kijkend naar het presentatiemodel, zien we dat afname van de vragenlijsten ‘met pen en papier’ plaatsvindt, de directe instrumenten maken gebruik van een computer / tablet. Waar de data opgeslagen wordt, is niet gespecificeerd.

4. Taxatie van mogelijkheden tot valide en betrouwbare meting

Ook voor zelfregulatie geldt dat het leerlingmodel aandacht behoeft. Zoals ook de review van Dinsmore et al. (2008) laat zien, wordt een leerprogressiemodel van zelfregulatie zelden uitgewerkt in termen van ontwikkeling van leerlingen. Daarnaast zijn er conceptuele vragen over wat zelfbeoordeling door middel van vragenlijsten meet wanneer de uitkomsten hiervan nauwelijks samenhangen met directe metingen. Een mogelijk probleem dat hierbij optreedt is de vraag in hoeverre zelfregulatie zich domein-/taakspecifiek of generiek ontwikkelt (McCoy, 2019). Directe metingen zijn gericht op taakspecifieke metingen, terwijl vragenlijsten meer generiek meten.

Er lijken momenteel stappen gezet te worden om de directe metingen die voorheen vooral in laboratorium-setting afgenomen werden, ook digitaal afneembaar te maken (Obradović et al., 2018). Dit lijkt mogelijkheden te bieden voor een meer grootschalige afname van dit type instrumenten. Ook het digitaal monitoren van leerprocessen van leerlingen biedt mogelijkheden, maar hierbij lijken moeilijkheden op te treden bij het begrijpen en scoren van het gedrag van leerlingen (Hadwin et al., 2007; Nesbit et al., 2006).

5. Conclusie

De conceptualisering van zelfregulatie is complex vanwege de vele onderscheiden concepten die aan elkaar gerelateerd worden. Voor metacognitie, zelfregulatie, zelfregulerend

leren en executieve functies lijkt evidentie te bestaan dat deze van belang zijn. Gezien de grote variatie in theoretische uitwerkingen is de kennis over deze concepten vooralsnog echter gefragmenteerd.

Voor het meten van zelfregulatie worden vooral zelfbeoordelingen gebruikt. Omdat ook het bewust toepassen van zelfregulatie als onderdeel van de competentie beschouwd kan worden, wordt een combinatie van zelfrapportage en directe meting wenselijk geacht (Boekaerts & Corno, 2005; Van Hout-Wolters, 2005; Vandervelde et al., 2013; Winne & Perry, 2000). Dit levert echter problemen op bij de interpretatie van verschillende scores, ook gezien de relatief lage correlaties tussen zelfinschattingen en observaties. Meer inzicht in deze relatie tussen deze metingen is van belang voor het verder verhogen van de begripsvaliditeit van instrumenten.

Het is bovendien de vraag in hoeverre deze verschillen het gevolg zijn van de meetbenadering (direct of indirect), of een conceptueel onderscheid tussen zelfregulatie als generieke of taakspecifiek. Dit heeft ook implicaties voor de geschiktheid van instrumenten (McCoy, 2019; Van Hout-Wolters, 2009). Directe metingen van zelfregulatie en aanverwante begrippen wordt per definitie taakspecifiek gemeten, terwijl vragenlijsten vaak generieke metingen beogen. Van der Stel en Veenman (2014) vonden aanwijzingen dat de transitie van taakspecifieke naar generieke metacognitieve vaardigheden rond 15-jarige leeftijd plaatsvindt. We hebben hier te maken met het praktische probleem dat zelfregulatie per definitie in taak- en domeinspecifieke contexten gebeurt, en dat het generieke karakter van deze skill pas aangetoond kan worden als de scores over deze contexten heen duidelijk generaliseerbaar zijn. Anders gezegd, de skill is generiek, wanneer een leerling, ongeacht de context of het domein waarin hij opereert, vergelijkbare scores behaalt voor zelfregulatie.

Conclusie

Om inzicht te krijgen in de ontwikkeling van skills zijn metingen van beheersingsniveaus onmisbaar. In deze verkenning inventariseren we de mogelijkheden om vier skills te meten. Aanleiding voor deze studie is de ontwikkeling van het Nationaal Cohortonderzoek Onderwijs (NCO) als vruchtbare setting om de stand van zaken en ontwikkeling van skills van leerlingen inzichtelijk te maken. Het NCO maakt gebruik van koppelingen van databronnen. In deze verkenning staat de vraag centraal hoe de meting van advanced skills zou kunnen worden geïntegreerd in een of meer modules die in zo'n context kan worden gerealiseerd.

Inzichten in de inhoud en ontwikkeling van advanced skills (of 21st century skills) worden ernstig bemoeilijkt door onduidelijkheid in de conceptualisering van de skills waar het om gaat (vgl. Van den Berge et al., 2014). Er bestaat inmiddels een veelvoud aan kaders en raamwerken van de skills waar het om zou gaan (Voogt & Pareja Roblin, 2010). Een eerste praktische afweging was dan ook het maken van een keuze voor de in de verkenning op te nemen skills. Hoewel zo'n selectie een enigszins arbitrair karakter kent, denken we met burgerschapscompetenties, digitale geletterdheid, sociaal-emotionele ontwikkeling en zelfregulatie vier domeinen geselecteerd te hebben 1) die in de bredere literatuur als inhoudelijk relevant worden aangewezen; 2) waarvan op basis van een langere onderzoekstraditie substantiële uitspraken te doen zijn over de conceptuele uitwerking van de betreffende skills; 3) waarvan de formatie voor een relevant deel plaatsvindt tijdens het doorlopen van het primair en voortgezet onderwijs; 4) die betekenisvol beïnvloed kunnen worden door onderwijs; 5) waarvoor meerdere (typen) meetinstrumenten ontwikkeld zijn voor verschillende leeftijdsgroepen; en 6) die een brede dekking geven van de domeinen waar het om lijkt te gaan zodat vooralsnog een benadering kan worden gevolgd die

voldoende breed blijft om (bij de nog beperkte kennis) tot kennisgroei over meerdere lijnen te kunnen komen.

In elk van de vier voorgaande hoofdstukken gaven we een schets van de stand van inzichten in de conceptuele uitwerking, formatie en relevantie van de betreffende skill, alvorens een taxatie te geven van de meetinstrumenten die daarbij gebruikt worden. De taxaties hebben als doel een gefundeerd antwoord op de onderzoeksvraag te kunnen geven. Voor de selectie van meetinstrumenten hebben we gepoogd een zo goed mogelijk overzicht te geven van de meest gebruikte (typen) instrumenten voor het meten van de skills. Voorwaarde daarbij was dat er voldoende informatie beschikbaar is over de conceptuele uitwerking van het instrument en de resultaten van afnames in het onderwijs. Om deze reden zijn bijvoorbeeld instrumenten waarvan geen verantwoording beschikbaar is, die alleen in laboratoriumsettings afgenomen worden, of instrumenten die nog in ontwikkeling zijn, buiten beschouwing gelaten. In de verkenning zijn 7 instrumenten voor burgerschapscompetenties, 14 instrumenten voor digitale geletterdheid, 9 instrumenten voor sociaal-emotionele ontwikkeling en 6 instrumenten voor zelfregulatie meegenomen. Het overgrote deel van de geselecteerde instrumenten is toegespitst op de leeftijdsgroep bovenbouw primair onderwijs en onderbouw voortgezet onderwijs. Daarnaast is bij de selectie van instrumenten getracht om zowel instrumenten die gebruikt maken van directe als indirecte metingen van skills mee te nemen.

Resultaten van de verkenning

In deze sectie bespreken we de belangrijkste uitkomsten van het onderzoek.

Vervolgens gaan we in op de implicaties van de bevindingen voor toepassing in het NCO.

Een relevante uitkomst is dat in de literatuur voor elk van de skills die we hierboven bespreken kritisch gekeken wordt naar de conceptuele uitwerking van het begrip. Dit werd

hiervoor al gememoreerd voor overkoepelende termen als 21st century skills, maar blijkt ook voor de verschillende skills die daaronder geschaard worden een probleem. Zo wordt bijvoorbeeld voor zowel digitale geletterdheid, sociaal-emotionele ontwikkeling als zelfregulatie opgemerkt dat binnen het onderzoeksdomein verschillende begrippen gebruikt worden, en dat dezelfde begrippen soms verschillende betekenissen hebben, en verschillende begrippen soms naar dezelfde concepten verwijzen (bijv. Darling-Churchill & Lippman, 2016; Dinsmore et al., 2008; Voogt et al., 2019). Deze moeilijkheden rond de conceptuele uitwerking en begripsafbakening lijken illustratief voor de complexiteit van het meten van skills. Immer, wat moet precies gemeten worden?

Een gerelateerd conceptueel vraagstuk dat zich in het bespreken van de vier skills in deze verkenning opnieuw aandient is hoe de verschillende skills zich tot elkaar verhouden (Skills-platform, 2016; Van den Berge et al., 2014). Zo worden bijvoorbeeld zelfregulatie en sociaal-emotionele ontwikkeling vanuit beide onderzoeksdomeinen aan elkaar verbonden. Vanuit literatuur over zelfregulatie wordt aangegeven dat zelfregulatie ook een motivationele en affectieve component kent, waarvan bijvoorbeeld emotieregulatie onderdeel is (Panadero, 2017). Vanuit literatuur over sociaal-emotionele ontwikkeling kan zelfregulatie als component beschouwd worden naast sociale competentie, emotionele competentie en gedragsproblemen (Halle & Darling-Churchill, 2016). Een ander vraagstuk dat zich op het snijvlak van twee van de concepten lijkt te bevinden is bijvoorbeeld digitaal burgerschap. Deze ontwikkelingen lijken een integraal begrip van de ontwikkeling van verschillende skills verder te bemoeilijken en vragen om meer conceptuele helderheid. Om kip-of-ei vragen te voorkomen, hebben we gezien het doel van deze verkenning voor een pragmatische benadering gekozen en stellen vooralsnog (alleen) de eis dat het betreffende (deel)concept eenduidig geformuleerd en zinvol gemeten kan worden zodat de begripsvaliditeit niet voor al te grote problemen zorgt.

De instrumenten uit de inventarisatie hebben we geëvalueerd op basis van het ECD ontwerpmodel. Hierin worden het leerlingmodel, taakmodel, bewijsmodel en assemblagemodel onderscheiden (zie inleiding). Voor elk van deze componenten zijn een aantal bevindingen het bespreken waard.

Wat betreft het leermodel valt op dat veel instrumenten die skills van leerlingen meten weinig uitwerking geven aan een leerlingontwikkelingsmodel. Daaronder verstaan we een (onderbouwde) uitwerking van hoe verondersteld wordt dat leerlingen de betreffende skills (gedurende de levensloop) verwerven, en hoe het meetmoment zich tot die ontwikkeling verhoudt. Dat maakt dat instrumenten die voor verschillende doelgroepen ontwikkeld zijn moeilijk aan elkaar te koppelen zijn, omdat niet helder is hoe prestaties in beide stadia van ontwikkeling zich tot elkaar verhouden. Is het bij burgerschapscompetenties bijvoorbeeld zo dat – zoals het instrument Burgerschap meten veronderstelt – leerlingen van circa 10-16 jaar burgerschapscompetenties ontwikkelen in de sociaal-maatschappelijke situaties waar zij in de dagelijkse praktijk mee te maken krijgen, zoals het oplossen van een conflict tussen klasgenoten of het willen horen van verschillende standpunten in een discussie? Anders dan in dit voorbeeld expliciteren veel instrumenten echter niet welke mechanismen en ontwikkelingen verondersteld worden.

Voor wat betreft het taakmodel vinden we de grootste verschillen tussen directe en indirecte metingen. Voor alle vier de skills geldt dat traditioneel veel gebruik gemaakt wordt van indirecte metingen, zoals vragenlijsten voor leerlingen en docenten. Met name voor digitale geletterdheid zijn de laatste jaren verschillende instrumenten ontwikkeld die gebruik maken van directe meting van relevante vaardigheden. Ook voor de andere skills kwamen we instrumenten tegen die met behulp van digitale afnameomgevingen een directe meting beogen. Voor deze andere skills roept dat echter een belangrijke vraag op naar de mate waarin gedrag in digitale situaties representatief is voor gedrag daarbuiten. We hechten er

daarbij ook aan om games niet als panacee naar voren te schuiven voor het meten van skills. Niet alleen is het veld nog volop in ontwikkeling, maar ook in game-based assessment doen zich problemen voor (Redecker, 2013). Zo kunnen leerlingen het gewenste gedrag enkel laten zien om een hoge score te halen. Het verbergen van de taak waarop het gedrag van de leerling in de taaksituatie gescoord wordt biedt hier mogelijk deels een oplossing voor. Redecker (2013) beschouwt deze “embedded assessment” als de volgende stap in game-based assessment. Het ontwikkelen van dit type assessment is echter een intensief proces, zoals bijvoorbeeld de ruim tienjarige ontwikkeling van gStudy en nStudy voor het inzichtelijk maken van zelfregulatie laat zien (Nesbit et al., 2006; Winnie et al., 2019).

Wat betreft het bewijsmodel zien we verschillen tussen instrumenten in de mate waarin betrouwbaarheidsgegevens over scores van leerlingen op de doelvaardigheden beschikbaar zijn. Voor een derde van de instrumenten waren deze er niet. De instrumenten waarvoor wel betrouwbaarheidsgegevens beschikbaar waren, leverden doorgaans goede betrouwbaarheid op voor metingen op stelselniveau. De betrouwbaarheid was vaak minder geschikt om belangrijke beslissingen over leerlingen op te baseren. Ook instrumenten waar gebruik gemaakt wordt van meerdere beoordelaars scores matig tot voldoende op interbeoordelaarsbetrouwbaarheid. Het beeld is dus dat meetinstrumenten vooral geschikt zijn voor stelselmetingen, en dat ze (nog) niet inzetbaar zijn om bijvoorbeeld schoolcijfers voor leerlingen op te baseren. Waar sprake is van een normschaal is deze in vrijwel alle gevallen gebaseerd op relatieve scores van leerlingen, en vooral bij instrumenten voor sociaal-emotionele ontwikkeling gericht op het identificeren van leerlingen die risico lopen op problematische ontwikkeling.

Wat betreft het assemblagemodel wordt vaak matig tot voldoende uitgewerkt hoe de prestaties van leerlingen op de taken die geselecteerd zijn voor de meting, zich verhouden tot prestaties van leerlingen op het bredere domein van taken waaruit de selectie gemaakt is.

Deze constatering is ook gerelateerd aan de conceptuele problemen die hierboven beschreven worden: wanneer niet duidelijk uitgewerkt is wat de domeinvaardigheid inhoudt en welke taken en gedragingen van leerlingen hier onderdeel van uitmaken, is ook niet duidelijk vast te stellen in hoeverre dat domein door het instrument gedekt wordt. In meer technische termen gaat het hier om de criteriumvaliditeit: wat zegt een lage of een hoge score van een leerling op een taak over de mate van competent handelen in situaties buiten de test en in de toekomst?

Implicaties voor het meten van skills in het NCO

Een fundamenteel probleem dat in deze verkenning naar voren komt is de onduidelijkheid ten aanzien van de conceptuele uitwerkingen van de verschillende skills. Wat opvalt is dat in de verschillende bronnen binnen elk onderzoeksdomein vergelijkbare constateringen gedaan worden ten aanzien van wisselende en onduidelijke definities van kernbegrippen. Het ligt gezien de ontwikkelingsgeschiedenis niet in de rede te verwachten dat op korte termijn van meer conceptuele duidelijkheid of, meer precies, integratie van de uiteenlopende benaderingen sprake zal zijn. Voor realisering van skills-metingen op korte termijn is het niet alleen weinig realistisch daarop te wachten – dat is ook niet nodig. We stellen voor het accent te leggen op instrumenten die gebaseerd zijn op adequate conceptualisering en theoretische onderbouwing van de te meten skills, zodat gewaarborgd is dat helder is van welke definitie de meting uitgaat, en hoe deze precies geoperationaliseerd is. Op die manier wordt bereikt dat binnen afzienbare termijn metingen gerealiseerd kunnen worden. Dat is niet alleen zinvol vanwege het belang van een beter inzicht in de stand van zaken en ontwikkeling van bedoelde skills, maar ook omdat meer empirische kennis nodig is om tot verdere conceptuele ontwikkeling te kunnen komen.

In dat opzicht merken we op dat bij de verschillende skills goede voorbeelden te vinden zijn, zoals bijvoorbeeld de instrumenten van ICILS en NAP-ICT Literacy bij digitale geletterdheid. De bijlagen bij deze studie vormen dan ook het hart van onze verkenning en geven, samen met eerdere verkenningen, weer welke instrumenten op korte termijn in het kader van NCO gebruikt kunnen worden (zie Daas et al., 2019; Ledoux et al., 2013). Omdat de instrumenten op veel punten verschillen (denk bijvoorbeeld aan leeftijdsfase) is eerst een nadere bepaling nodig van de doelen en beoogde inzichten die de skills-modules in NCO zouden moeten realiseren. Pas dan is een nadere keuze zinvol. In het bestek van deze verkenning is het daarom niet mogelijk een specifiek advies over al dan niet te gebruiken instrumenten te formuleren. In plaats daarvan biedt de informatie in de bijlagen een goed overzicht van mogelijkheden (en minder wenselijke keuzes). Dat leidt tot de aanbeveling om nu een gecombineerde stap te gaan zetten, waarin meer specifieke wensen en doelen voor een skills-module worden geformuleerd. De geformuleerde wensen en doelen kunnen vervolgens gekoppeld worden aan concrete voorstellen omtrent te gebruiken (combinaties van) instrumenten op basis van de informatie in de bijlagen en de weging daarvan in dit hoofdstuk.

Veel instrumenten voor het meten van skills zijn toegespitst op afname in de bovenbouw primair onderwijs en onderbouw voortgezet onderwijs en – in het geval van sociaal-emotionele ontwikkeling – ook in de onderbouw van het primair onderwijs. De focus op deze leeftijdsgroepen komt vooral voort uit de veronderstelling dat deze leeftijdsperiode van belang is voor de formatie van de skills. Dit legt echter ook bloot waarom er weinig inzicht lijkt in de ontwikkeling van leerlingen die daaraan voorafgaat of op volgt. Deze inzichten in een leerlingontwikkelingsmodel ontbreken behalve in conceptueel opzicht dus ook grotendeels in empirisch opzicht. Om het al genoemde ‘kip-ei’-dilemma te voorkomen, is het raadzaam in beide te investeren. Voor NCO kan dan prioriteit worden gelegd bij verzameling van empirische kennis, eventueel gecombineerd met aanvullend conceptueel

ontwikkelwerk, bijvoorbeeld in een verdiepende ontwikkelingsmodule gericht op verdere theoretische ontwikkeling en nader onderbouwde conceptualisering in latere stappen. De roep om meer longitudinaal onderzoek rond de ontwikkeling van skills klinkt veelvuldig, en een eerdere Nederlandse simulatiestudie en het Zweedse Youth & Society onderzoek laten zien dat deze ook versneld verzameld kunnen worden (Amnå et al., 2009; Dijkstra et al., 2004). In beide studies naar de ontwikkeling van maatschappelijke en politieke oriëntaties is men gelijktijdig gaan meten onder verschillende leeftijdsgroepen in zowel het basisonderwijs, het voortgezet onderwijs en in de periode van (jong) volwassenheid. Door de ontwikkeling van deze groepen in beide studies in analytisch gekoppelde longitudinale designs te volgen zijn versneld inzichten ontwikkeld rond de formatie van deze competenties en skills gedurende de eerste fasen van de levensloop tot in de (jong) volwassenheid (Abdelzadeh & Lundberg, 2016; Russo & Statin, 2017).

Traditioneel wordt door de meeste instrumenten voor het meten van skills gebruik gemaakt van indirecte metingen. In deze verkenning hebben we daarom ook instrumenten meegenomen die directe meting van skills mogelijk maken. Deze instrumenten maken veelal gebruik van digitale, gesimuleerde taakomgevingen. Ook deze metingen zijn niet zonder nadelen. Voor digitale geletterdheid is een relevante taakomgeving relatief eenvoudig te realiseren (hoewel ook daar de contextuele relevantie van de leersituatie aandacht vraagt) maar voor de andere skills zal een vertaalslag gemaakt moeten worden naar ‘analoge’ situaties. Het ideale instrumentarium zal dus eerder bestaan uit een mix van directe en indirecte metingen. De nadruk in deze verkenning op directe metingen komt voort uit de (kwantitatieve) achterstand die dit type instrumenten hebben ten opzichte van indirecte metingen. Het lijkt niet te verwachten dat directe instrumenten op korte termijn zullen volstaan, en triangulatie lijkt in dat kader het meest vruchtbaar en de enige mogelijke optie.

Al met al leidt deze verkenning tot de volgende aanbevelingen:

1. *Start met empirisch onderzoek naar de ontwikkeling van skills. Wacht niet op verdere conceptuele ontwikkeling, maar werk daar wel aan.*

Een belangrijk aandachtspunt bij het meten van de vier skills in deze verkenning is dat voor elk van deze skills geldt dat er discussie bestaat over de conceptuele afbakening; wat is precies de inhoud van de skill? De onderzoeksliteratuur voor elk van de skills is opvallend vergelijkbaar in het uitspreken van de behoefte aan verheldering van het begrippenkader, en deze behoeften spelen vaak al langere tijd. Het is niet te verwachten dat dit op korte termijn tot eenduidige, breed gedragen definities zal leiden; verschillende instrumenten zullen blijven uitgaan van verschillende conceptuele kaders. Er is bovendien verschil in de mate waarin instrumenten op een heldere theoretische onderbouwing gebaseerd zijn. Deze onderbouwing is wel nodig voor het meten van skills in een NCO module. Het daarom verstandig te starten met empirische gegevensverzameling op basis van goed onderbouwde instrumenten, zodat hierop voortgebouwd kan worden. Verdere conceptuele uitwerking kan zich in eerste instantie richten op de meting in het kader van het NCO, waarna de resultaten van de meting gebruikt kunnen worden voor meer fundamentele conceptuele ontwikkeling rond het betreffende skills begrip.

Verschiede instrumenten bieden goede aanknopingspunten voor operationalisering en theoretische onderbouwing van de te meten skills, en zijn te gebruiken voor het ontwerp van de beoogde modules. De brede inventarisatie die voor deze studie uitgevoerd is, laat zien dat – hoewel elk van de instrumenten sterke en zwakke punten kent – diverse instrumenten geschikt zijn voor het meten van skills op stelselniveau. Daarbij gaat het niet om oordelen over het beheersingsniveau van individuele leerlingen, maar om inzichten in de mate van beheersing van groepen leerlingen. Het lijkt dus mogelijk om met beperkte investeringen deze instrumenten in te zetten voor doelgerichte gegevensverzameling. Voor de meting van de ontwikkeling van skills – waarvoor herhaalde metingen noodzakelijk zijn – is het nodig

tevens te investeren in doorontwikkeling van de gebruikte instrumenten zodat longitudinale metingen mogelijk zijn. Afhankelijk van de nadere specificatie van doelen en beoogde opbrengsten van NCO skills modules, zijn op basis van deze verkenning gerichte keuzes van instrumenten te maken waar, afhankelijk van de ambities, verder in te investeren valt.

2. *Betrek bij de keuze van instrumenten de doelen die met de metingen worden nagestreefd.*

De bruikbaarheid van instrumenten is onder meer afhankelijk van de doelen die met het afnemen ervan beoogd worden en de bijbehorende meetpretentie. Zo verschillen instrumenten in de reikwijdte van wat gemeten wordt, zoals een focus op een specifiek toepassingsdomein (bijvoorbeeld oordelen en argumenteren als onderdeel van actief burgerschap en sociale integratie) of de gehele skill (bijvoorbeeld een samengesteld instrument dat verschillende dimensies van burgerschapscompetenties meet). Daarnaast verschilt de inhoud naar gelang een integrale meting wordt nagestreefd van de kennis, houding en vaardigheden die onderdeel van de skill zijn of van een nadere selectie daarvan. Een ander relevant onderscheid is naar de functie die de gegevensverzameling vervult, bijvoorbeeld voor onderzoek gericht op beschrijvingen op stelselniveau, diagnostisering van (problematische) ontwikkeling, en op de praktijk gerichte interventiestudies. De keuze van passende instrumenten is dus afhankelijk van het doel van de meting. Waar het gaat om metingen ten behoeve van inzicht in de ontwikkeling van skills, is in ieder geval van belang dat het instrument gebaseerd is op een uitwerking die uitgaat van een ontwikkelbare, aanleerbare capaciteit die toegepast wordt in een ontwikkelingspsychologisch-relevante context. Hoewel de meeste onderzochte instrumenten hier wel van uitgaan, treffen we grote verschillen aan tussen instrumenten in de mate waarin deze veronderstellingen uitgewerkt worden. Met name de instrumenten voor diagnostisch of signalerend gebruik in deze verkenning geven hier beperkt expliciet uitwerking aan.

3. *Ga vooralsnog uit van combinaties van meetmethoden en investeer in ontwikkeling van metingen van skills op basis van directe taken.*

Metingen van skills zijn vaak opgebouwd uit separate metingen van kennis, houding en vaardigheden. Daarbij wordt kennis doorgaans gemeten aan de hand van een kennistoets met meerkeuzevragen. Houding en vaardigheden worden vaak gemeten door middel van een vragenlijst met Likert schalen. Metingen van vaardigheden komen daarmee doorgaans neer op de inschatting door leerlingen van het eigen kunnen (*self efficacy*). Dit betekent dat, meer nog dan voor kennis en houdingen, voldoende aandacht voor de validiteit van dergelijke metingen nodig is. Onderzoek laat dikwijls zien dat de beoordeling van de eigen vaardigheden van leerlingen tot andere beelden leidt dan waarnemingen van vaardigheden door anderen of directe metingen van daadwerkelijk getoond gedrag. Binnen metingen van digitale geletterdheid is een ontwikkeling zichtbaar naar directe meting via integratie van meetinstrumenten in digitale omgevingen. Ook voor sociaal-emotionele ontwikkeling en zelfregulatie zijn er instrumenten die in een digitale omgeving beogen relevant gedrag van leerlingen te monitoren om zo uitspraken te kunnen doen over hun beheersingsniveau. Deze ontwikkelingen zijn van belang met het oog op rechtstreekse meting van skills in relevante taaksituaties, en verdienen verdere aandacht. Mede vanwege conceptuele vragen over de interpretatie van verschillen tussen directe en indirecte metingen is triangulatie van metingen hierbij van belang.

4. *Investeer in de ontwikkeling van (empirisch onderbouwde) leerprogressiemodellen.*

Meetinstrumenten voor skills richten zich doorgaans op een specifieke leeftijdsgroep, vaak de bovenbouw van het basisonderwijs en/of de onderbouw van het voortgezet onderwijs. Doorgaans wordt daarbij wel geëxpliciteerd waarom die levensfase relevant is, maar zelden wordt gewerkt vanuit een conceptueel model van de ontwikkeling van de skill gedurende de levensloop. Omdat meetinstrumenten niet bij voorbaat geschikt zijn voor

verschillende leeftijdsgroepen, en niet duidelijk is welke ontwikkelingen voorwaardelijk dan wel opvolgend plaatsvinden, is het resultaat dat vooralsnog maar weinig inzicht bestaat in de ontwikkeling van skills. Vragen rond leerprogressie zijn daarom niet of nauwelijks te beantwoorden. Dat leidt ertoe dat metingen vaak op zichzelf staan en hun reikwijdte niet duidelijk is. Daardoor is het ook niet goed mogelijk om vragen te beantwoorden rond de inrichting van het onderwijsaanbod en interventies gericht op (effectieve) stimulering van de ontwikkeling van skills.

5. Realiseer een versnelde longitudinale opzet.

De vraag naar meer longitudinale gegevens – noodzakelijk voor inzicht in de ontwikkeling van skills gedurende de levensloop – klinkt veelvuldig. Dat vraagt veel investeringen en een lange adem. Het is echter ook denkbaar hiervoor een versnelde opzet te volgen. Door leerlingen in verschillende leeftijdsgroepen gelijktijdig te volgen, ontstaat versneld inzicht in de ontwikkelingen over deze groepen heen. Zo heeft men bijvoorbeeld in Zweden onderzoek gedaan naar de ontwikkeling van politieke houdingen van jongeren van 13, 16, 19 en 22 jaar. Door deze jongeren vijf jaar te volgen is men in staat geweest zowel de ontwikkeling per leeftijdsgroep, als de ontwikkeling over de groepen heen in kaart te brengen. Ook voor Nederland is dergelijk, op de verwerving van skills gericht onderzoek uitgevoerd. De resultaten daarvan laten zien dat zo'n benadering goed werkbaar is en – indien op grotere schaal toegepast – zeker voor de korte termijn relevante resultaten belooft. Inzet van NCO-modules zou hierbij goed denkbaar zijn, omdat het via het NCO mogelijk zou moeten zijn om groepen leerlingen over meerdere jaren te volgen.

Discussie: hoe verder?

De verkenning laat zien dat het beschikbare instrumentarium voor het meten van skills gemêleerd is; we zien verschillen tussen instrumenten op vele facetten. In deze sectie

schetsen we per skill enkele mogelijke aanknopingspunten voor verdere ontwikkeling van integratie van skills modules in het NCO. We beogen met deze slotopmerkingen een aanzet te geven tot discussie over effectieve investeringen om meer inzicht te krijgen in de ontwikkeling van skills.

Burgerschapscompetenties

Operationalisering van burgerschapscompetenties in meetinstrumenten gaat doorgaans uit van een uitsplitsing in kennis, houding en vaardigheden. Voor het meten van kennis wordt gebruik gemaakt van meerkeuzevragen. Voor het meten van houdingen en vaardigheden wordt gebruik gemaakt van zelfinschatting op basis van Likert schalen. Hoewel discussie over de inhoud van het begrip inherent lijkt aan wat burgerschap in een democratische samenleving betekent, baseren de meeste instrumenten zich op een gedeelde kern en lijken verschillen tussen instrumenten vooral een verschil in focus binnen het domein te reflecteren. Een van de sterke kanten is dat verschillende instrumenten gebaseerd zijn op leerlingontwikkelingsmodellen waarin deze componenten uitgewerkt worden in relevante contexten voor jongeren. Er is dus enige basis om herhaalde meting op te baseren. Dat lijkt vooral te vragen om uitbreiding van itembestanden zodat verschillende vergelijkbare afnames onder leerlingen mogelijk zijn. Er lijkt vooralsnog weinig materiaal voorhanden om tevens directe meting van deze skill in relevante (gesimuleerde) situaties te faciliteren.

Digitale geletterdheid

De inventarisatie die recent werd uitgevoerd voor het peilingsonderzoek digitale geletterdheid liet al zien dat internationaal verschillende instrumenten bestaan die taaksituaties simuleren waarin leerlingen digitale vaardigheden toepassen. Kennis wordt in dit domein vaak nog getoetst met meerkeuzevragen, en houdingen worden – wanneer ze gemeten worden – met behulp van Likert schalen bevraagd. Vanwege de aard van de skill leent digitale geletterdheid zich goed voor directe metingen in gesimuleerde situaties. Hoewel

de kwaliteit van instrumenten uiteenloopt, identificeren we in deze verkenning enkele goed onderbouwde en uitgewerkte instrumenten. Een van de voornaamste uitdagingen lijkt hier het definiëren en operationaliseren van het begrip digitale geletterdheid, mede vanwege de voortdurende ontwikkelingen in dit domein.

Sociaal-emotionele ontwikkeling

Sociaal-emotionele ontwikkeling is een breed concept. Op basis van de verkenning lijkt het treffender om van een reeks onderdelen te spreken, dan van één samengesteld, omvattend begrip. Sociaal-emotionele ontwikkeling wordt in instrumenten veelal gerelateerd aan (problematisch) gedrag. Instrumenten voor het meten van aspecten van sociaal-emotionele ontwikkeling zijn doorgaans niet gebaseerd op een leerlingontwikkelingsmodel waarin uitgewerkt wordt hoe de uitslag van de meting wordt uitgedrukt in een ontwikkelingsniveau. In plaats daarvan zijn metingen vooral gericht op het indexeren van gedrag. Vanwege de koppeling aan het NCO, zijn in de verkenning meerdere instrumenten meegenomen die onderdeel uitmaken van leerlingvolgsystemen die op scholen gebruikt worden. Deze instrumenten zijn gericht op het identificeren van leerlingen van wie het gedrag mogelijk problematisch is, zodat leraren daar vervolgstappen aan kunnen verbinden. De infrastructuur die door leerlingvolgsystemen geboden wordt biedt – mede vanwege de herhaaldelijke afnames – wel beloftevolle aanknopingspunten, maar de instrumenten die daar momenteel onderdeel van uitmaken zijn niet geschikt om inzichten in de ontwikkeling van deze skills te creëren.

Zelfregulatie

Over zelfregulatie bestaat relatief veel conceptuele onduidelijkheid. Dit lijkt verbonden met twee kenmerken die de literatuur rond dit thema typeren. Allereerst worden er verschillende begrippen gebruikt met verschillende, gerelateerde theoretische uitwerkingen: zelfregulatie, zelfregulerend leren, executieve functies en metacognitie. Verschillende

overzichtsstudies bieden een kader waarin deze begrippen geplaatst kunnen worden, maar dit heeft nog niet geleid tot een heldere begripsafbakening. Een daaraan verbonden moeilijkheid is dat vanuit verschillende onderzoekdisciplines gewerkt wordt, in het bijzonder (neuro)psychologische en pedagogische wetenschappen, waarbij de eerste vooral nadruk legt op cognitieve processen, en de tweede vooral op interactie met de omgeving. Het verenigen van deze conceptuele benaderingen lijkt op korte termijn niet haalbaar. Een tweede kenmerk van de discussie in dit skillsdomein is in hoeverre deze skills taakspecifiek of generiek ontwikkelbaar en meetbaar zijn. Binnen de gerelateerde onderzoeksdomeinen zijn verschillende instrumenten beschikbaar die potentieel in te zetten zijn in het NCO. De meeste instrumenten maken gebruik van zelfinschatting of observatieformulieren, waarbij niet wordt verwezen naar concreet gespecificeerde taaksituaties. Vooral voor executieve functies is ontwikkeling zichtbaar richting digitale toetsomgevingen waarin directe metingen plaatsvinden.

Conclusie

Bovenstaande bespreking laat zien dat er niet één skill het best gepositioneerd is voor verdere ontwikkeling in het NCO. De keuze zal dus mede ingegeven moeten worden vanuit het belang dat aan inzichten in de ontwikkeling van een specifieke skills gehecht wordt. We eindigen met mogelijke concrete vervolgstappen per skill. Voor burgerschapscompetenties kan gedacht worden aan uitbreiding van het instrumentarium voor herhaalde metingen op basis van bestaande aanzetten tot een leerlingontwikkelingsmodel. Voor digitale geletterdheid gaat het om het vertalen en uitbreiden van instrumenten voor directe metingen van vaardigheden, op basis van in de Nederlandse context relevante invullingen. Voor sociaal-emotionele ontwikkeling gaat het om het ontwerpen van instrumenten die binnen de bestaande structuren van leerlingvolgsystemen gebruikt kunnen worden om inzicht te krijgen in ontwikkelingen van leerlingen. Voor zelfregulatie gaat het om de keuze en

doorontwikkeling van instrumenten op basis van conceptuele afwegingen over welke van de verwante begrippen centraal staat en a) in welke contexten of vakgebieden dit gemeten zou moeten worden en b) in hoeverre uitspraken uit metingen kunnen worden gegeneraliseerd over toepassingscontexten en vakgebieden heen.

Referenties¹

- Abdelzadeh, A., & Lundberg, E. (2017). Solid or flexible? Social trust from early adolescence to young adulthood. *Scandinavian Political Studies*, 40(2), 207-227. <https://doi.org/10.1111/1467-9477.12080>.
- *ACARA. (2018a). *National assessment program - ICT literacy technical report*. Sydney, Australië: Australian Curriculum, Assessment and Reporting Authority.
- *ACARA. (2018b). *National assessment program - ICT literacy years 6 & 10 report 2018*. Sydney, Australië: Australian Curriculum, Assessment and Reporting Authority. https://www.nap.edu.au/docs/default-source/default-document-library/2017napictlreport_final.pdf?sfvrsn=2.
- *ACER. (2008). *National assessment program - ICT literacy technical report*. Camberwell, Australië: Australian Council for Educational Research.
- *Aesaert, K., Van Nijlen, D., Vanderlinde, R., & Van Braak, J. (2014). Direct measures of digital information processing and communication skills in primary education: Using item response theory for the development and validation of an ICT competence scale. *Computers & Education*, 76, 168–181. <https://doi.org/10.1016/j.compedu.2014.03.013>.
- *Aesaert, K., Van Braak, J., Van Nijlen, D., Vanderlinde, R. (2015). Primary school pupils' ICT competences: Extensive model and scale development. *Computers & Education*, 81, 326-344. <https://doi.org/10.1016/j.compedu.2014.10.021>.
- *Aesaert, K., Van Nijlen, D., Vanderlinde, R., Tondeur, J., Devlieger, I., & Van Braak, J. (2015). The contribution of pupil, classroom and school level characteristics to
-

¹ Referenties met een * zijn onderdeel van de inventarisatie van meetinstrumenten

- primary school pupils' ICT competences: A performance-based approach. *Computers & Education*, 87,55-69.<https://doi.org/10.1016/j.compedu.2015.03.014>.
- *Aesaert, K., Voogt, J., Kuiper, E., & Van Braak, J. (2017). Accuracy and bias of ICT self-efficacy: An empirical study into students' over- and underestimation of their ICT competences. *Computers in Human behavior*, 75, 92-102.
<https://doi.org/10.1016/j.chb.2017.05.010>.
- Alwin, D., & Krosnick, J. (1991). Aging, cohorts, and the stability of sociopolitical orientations over the life span. *American Journal of Sociology*, 97(1), 169-195.
<http://www.jstor.org/stable/2781642>.
- *Amná, E. (2012). How is civic engagement developed over time? Emerging answers from a multidisciplinary field. *Journal of Adolescence*, 35(3), 611-627.
<https://dx.doi.org/10.1016/j.adolescence.2012.04.011>.
- *Amná, E., Ekström, M., Kerr, M., & Stattin, H. (2009). Political socialization and human agency: The development of civic engagement from adolescence to adulthood. *Statsvetenskaplig Tidskrift*, 111(1), 27-40.
- *Amná, E., Ekström, M., Kerr, M., & Stattin, H. (2015). *Codebook: The political socialization program*. Örebro: Youth & Society at Örebro University, Sweden.
- *Avsec, S., & Jamsek, J. (2016). Technological literacy for students aged 6–18: A new method for holistic measuring of knowledge, capabilities, critical thinking and decision-making, *International Journal of Technology and Design Education*, 26(1), 43–60. <https://doi.org/10.1007/s10798-015-9299-y>.
- *Avsec, S., & Szewczyk-Zakrzewska, A. (2016). Predicting academic success and technological literacy in secondary education: A learning styles perspective. *International Journal of Technology and Design Education*, 27(2), 233-250.
<https://doi.org/10.1007/s10798-015-9344-x>.

- *Avsec, S., & Jamsek, J. (2018). A path model of factors affecting secondary school students' technological literacy. *International Journal of Technology and Design Education*, 28(1), 145-168. <https://doi.org/10.1007/s10798-016-9382-z>.
- *Bazalt. (2018). *Handleiding Kijk! webbased*. Rotterdam: Bazalt.
- Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology: An International Review*, 54(2), 199-231. <https://doi.org/10.1111/j.1464-0597.2005.00205.x>.
- *Broer, N.A., Haverhals, B., & de Bruin, H.L. (2012). *Verantwoording pedagogisch expertsysteem Zien! voor het primair onderwijs*. Gouda: Driestar onderwijsadvies.
- *Broer, N.A., Haverhals, B., Grootenhuis, A., & van der Klooster-Sturm, S.W. (2016a). *Verantwoording Zien! leerling 5-8: Leer- en leefklimaat. Expertsysteem Zien! voor het primair onderwijs*. Gouda: Driestar onderwijsadvies.
- *Broer, N.A., Haverhals, B., Grootenhuis, A., & van der Klooster-Sturm, S.W. (2016b). *Verantwoording Zien! leerling 5-8: Sociale vaardigheden. Expertsysteem Zien! voor het primair onderwijs*. Gouda: Driestar onderwijsadvies.
- *Cito. (2015). *WISEON 2.0. Handleiding*. Arnhem: Cito.
- *Craig, A. DeRosier, M., & Watanabe, Y. (2015). Differences between Japanese and U.S. children's performance on "Zoo U": A game-based social skills assessment. *Games for Health Journal*, 4(4), 285-294. <https://doi.org/10.1089/g4h.2014.0075>.
- *Credé, M., & Phillips, L. (2011). A meta-analytic review of the motivated strategies for learning questionnaire. *Learning and Individual Differences*, 21(4), 337-346. <https://doi.org/10.1016/j.lindif.2011.03.002>.
- Daas, R., & Dijkstra, A.B. (2016). Burgerschap. In Skills-platform, *Skills voor de toekomst: Een onderzoeksagenda* (pp. 31-38). Den Haag: Ministerie van Onderwijs, Cultuur & Wetenschap.

- Daas, R., Dijkstra, A.B., & de Kort, I. (2019). *Verkenning instrumenten opbrengstgericht burgerschapsonderwijs: Voorstudie naar ontsluiting van meetinstrumenten als randvoorwaarde voor opbrengstgericht burgerschapsonderwijs*. Amsterdam: Universiteit van Amsterdam.
- Daas, R., Dijkstra, A. B., & Karsten, S. (2019). Assessing young people's citizenship attitudes using rubrics. *Studies in Educational Evaluation*, 62, 118-128.
<https://doi.org/10.1016/j.stueduc.2019.05.002>.
- Daas, R., & Mennes, H. (2018). *Inhoudelijke vergelijking domeinbeschrijving burgerschap in het basisonderwijs en ICCS*. Amsterdam: Universiteit van Amsterdam.
- Daas, R., Ten Dam, G., & Dijkstra, A.B. (2016). Contemplating modes of assessing citizenship competences, *Studies in Educational Evaluation*, 51, 88-95.
<https://dx.doi.org/10.1016/j.stueduc.2016.10.003>.
- Darling-Churchill, K. E., & Lippman, L. (2016). Early childhood social and emotional development: Advancing the field of measurement. *Journal of Applied Developmental Psychology*, 45, 1-7. <https://doi.org/10.1016/j.appdev.2016.02.002>.
- Dent, A., & Koenha, A. (2016). The relation between self-regulated learning and academic achievement across childhood and adolescence: A meta-analysis. *Educational Psychology Review*, 28(3), 425-474. <https://doi.org/10.1007/s10648-015-9320-8>.
- *DeRosier, M. (Red.) (2014). *Social skills assessment through games: The new best practice*. Charleston, SC: Interlink Press.
- *DeRosier, M., Craig, A., & Sanchez, R. (2012). Zoo U: A stealth approach to social skills assessment in schools. *Advances in Human-Computer Interaction*.
<https://doi.org/10.1155/2012/654791>.
- *DeRosier, M., & Thomas, J. (2018). Establishing the criterion validity of Zoo U's game-based social emotional skills assessment for school-based outcomes. *Journal of*

Applied Developmental Psychology, 55, 52–61.

<https://doi.org/10.1016/j.appdev.2017.03.001>.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, 64, 135-168.

<https://doi.org/10.1146/annurev-psych-113011-143750>.

Dijkstra, A.B. (2012). *Sociale opbrengsten van onderwijs*. Amsterdam: AUP.

Dijkstra, A.B. (2015). *Moeilijk meetbare onderwijsresultaten op het sociale en maatschappelijke domein. Verkennende notitie voor de Onderwijsraad*. Amsterdam: Universiteit van Amsterdam.

Dijkstra, A.B., Hofstra, J., Oudenhoven, J., Peschar, J., & Van der Wal, M. (2004). *Oud gedaan, jong geleerd? Een studie naar de relaties tussen hechtingsstijlen, competenties, EVLN-intenties en sociale cohesie*. Amsterdam: Aksan.

Dijkstra, A.B., & Janssens, F.J.G. (Red.). (2012). *Om de kwaliteit van het onderwijs. Kwaliteitsbepaling en kwaliteitsbevordering*. Amsterdam: Boom Lemma.

Dinsmore, D., Alexander, P., & Loughlin, S. (2008). Focusing the conceptual lens on metacognition, self-regulation, and self-regulated learning. *Educational Psychology Review*, 20(4), 391-409. <https://doi.org/10.1007/s10648-008-9083-6>.

*Duncan, T. G., & McKeachie, W. C. (2005). The making of the motivated strategies for learning questionnaire. *Educational Psychologist*, 40(2), 117–128. https://doi.org/10.1207/s15326985ep4002_6.

Durlak, J., Weissberg, R., Dymnicki, A., Taylor, R., & Schellinger, K. (2011). The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82, 405-432. <https://doi.org/10.1111/j.1467-8624.2010.01564.x>.

- Eckstein, K., Noack, P., & Gniewosz, B. (2012). Attitudes toward political engagement and willingness to participate in politics. Trajectories throughout adolescence. *Journal of Adolescence*, 35, 485-495. <https://doi.org/10.1016/j.adolescence.2011.07.002>.
- Eidhof, B., Ten Dam, G., Dijkstra, A.B., Van De Werfhorst, H. (2016). Consensus and contested citizenship goals in Western Europe. *Education, Citizenship and Social Justice*, 11(2), 114-129. <https://doi.org/10.1177/1746197915626084>.
- Eidhof, B., Ten Dam, G., Dijkstra, A.B., & van de Werfhorst, H. (2017). Youth citizenship at the end of primary school: The role of language ability. *Research Papers in Education*, 32(2), 217-230. <https://dx.doi.org/10.1080/02671522.2016.1167235>.
- European Communities. (2007). *The key competences for lifelong learning: A European framework*. Luxemburg: Publications Office of the European Union. <https://hdl.voced.edu.au/10707/285153>.
- *Fernández-montalvo, J., Peñalva, A., Irazabal, I., & López-, J. J. (2017). Effectiveness of a digital literacy programme for primary education students / Efectividad de un programa de alfabetización digital para estudiantes de educación primaria. *Cultura y Educación*, 29(1), 1-30. <https://doi.org/10.1080/11356405.2016.1269501>.
- Ferrari, A. (2013). *DigComp: A Framework for Developing and Understanding Digital Competence in Europe*. Luxembourg. Geraadpleegd van <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC83167/lb-na-26035-enn.pdf>.
- *Fraillon, J., Ainley, J., Schulz, W., Duckworth, D., & Friedman, T. (2019). *International Computer and Information Literacy Study 2018 assessment framework*. Cham, Zwitserland: Springer. <https://doi.org/10.1007/978-3-030-19389-8>.
- *Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Gebhardt, E. (2014). *Preparing for life in a digital age: The IEA international computer and information literacy study*

- international report*. Cham, Zwitserland: Springer. <https://doi.org/10.1007/978-3-319-14222-7>.
- *Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Gebhardt, E. (2019). *Preparing for life in a digital world: IEA international computer and information literacy study 2018 international report*. Cham, Zwitserland: Springer.
- *Fraillon, J., Schulz, W., & Ainley, J. (2013). *International computer and information literacy study: assessment framework*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- *Fraillon, J., Schulz, W., Friedman, T., Ainley, J., & Gebhardt, E. (Red.) (2015). *International computer and information literacy study 2013 technical report*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA). <https://doi.org/10.15478/uuid:b9cdd888-6665-4e9f-a21e-61569845ed5b>.
- *Fraillon, J., Schulz, W., Friedman, T., & Meyer, S. (Red.). (in voorbereiding). *IEA international computer and information literacy study 2018 technical report*. Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement (IEA).
- *Garcia, E., Sulik, M., & Obradović, J. (2019). Teachers' perceptions of students' executive functions: Disparities by gender, ethnicity, and ELL status. *Journal of Educational Psychology, 111*(5), 918–931. <https://doi.org/10.1037/edu0000308>.
- *Geboers, E., Geijsel, F., Admiraal, W., Jorgensen, T., & Ten Dam, G. (2015). Citizenship development of adolescents during the lower grades of secondary education. *Journal of Adolescence, 45*, 89-97. <https://doi.org/10.1016/j.adolescence.2015.08.017>.
- Halle, T. G., & Darling-Churchill, K. E. (2016). Review of measures of social and emotional development. *Journal of Applied Developmental Psychology, 45*, 8-18. <https://doi.org/10.1016/j.appdev.2016.02.003>.

- *Hadwin, A. F., Nesbit, J. C., Code, J., Jamieson-Noel, D. L., & Winne, P. H. (2007). Examining trace data to explore self-regulated learning. *Metacognition and Learning*, 2(2-3), 107–124. <https://doi.org/10.1007/s11409-007-9016-7>.
- *Harrison, G., & Vallin, L. (2018). Evaluating the metacognitive awareness inventory using empirical factor-structure evidence. *Metacognition Learning* 13, 15-38. <https://doi.org/10.1007/s11409-017-9176-z>.
- *Hatlevik, O. E., Scherer, R., & Christophersen, K. (2017). Moving beyond the study of gender differences: An analysis of measurement invariance and differential item functioning of an ICT literacy scale. *Computers & Education*, 113, 280-293. <https://doi.org/10.1016/j.compedu.2017.06.003>.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.
- *Heitink, M. (2018). *Eliciting teachers' and students' technological competences: Assessing technological skills in practice* (proefschrift). <https://10.3990/1.9789036546850>.
- *Heirweg, S., De Smul, M., Devos, G., & Van Keer, H. (2019). Profiling upper primary school students' self-regulated learning through self-report questionnaires and think-aloud protocol analysis. *Learning and Individual Differences*, 70, 155–168. <https://doi.org/10.1016/j.lindif.2019.02.001>.
- Hofmann, W., Schmeichel, B., & Baddeley, A. (2012). Executive functions and self-regulation. *Trends in Cognitive Sciences*, 16(3), 174-180. <https://doi.org/10.1016/j.tics.2012.01.006>.
- *Hohlfeld, T. N., Ritzhaupt, A. D., & Barron, A. E. (2010). Development and validation of the student tool for technology literacy (ST2L). *Journal of Research on Technology in Education*, 42, 361-389. <https://doi.org/10.1080/15391523.2010.10782556>.

- *Hohlfeld, T. N., Ritzhaupt, A. D., & Barron, A. E. (2013). Are gender differences in perceived and demonstrated technology literacy significant? It depends on the model. *Educational Technology Research and Development*, 61(4), 639-663. <https://doi.org/10.1007/s11423-013-9304-7>.
- *Huggins, A. C., Ritzhaupt, A. D., & Dawson, K. (2014). Measuring information and communication technology literacy using a performance assessment: Validation of the student tool for technology literacy (ST2L). *Computers & Education*, 77, 1-12. <https://doi.org/10.1016/j.compedu.2014.04.005>.
- Inspectie van het Onderwijs. (2016). *Analyse en waardering van opbrengsten. Primair onderwijs*. Utrecht: Inspectie van het Onderwijs.
- *Joosten, F. (2006). *Een maat om op te bouwen: Sociale competentie meten voor het basisonderwijs* (proefschrift). Amsterdam: Universiteit van Amsterdam.
- *Joosten, F. (2007). *Verantwoording van de sociale competentie observatielijst (SCOL)*. Rotterdam: CED-Groep.
- *Joosten, F., ten Heggeler, M., & Mouissie, L. (2016). Handleiding sociale competentie observatie lijst: Primair onderwijs. Rotterdam: CED-Groep.
- *Joosten, F., van Bokkem, M., ten Heggeler, M., Pot, L., & van Dijk, L. (2009). Handleiding sociale competentie observatie lijst: Voortgezet onderwijs. Rotterdam: CED-Groep.
- *Jun, S., Han, S., Kim, H., & Lee, W. (2014). Assessing the computational literacy of elementary students on a national level in Korea. *Educational Assessment, Evaluation and Accountability*, 26(4), 319-332. <https://doi.org/10.1007/s11092-013-9185-7>.
- Kankaraš, M., Feron, E., & Renbarger, R. (2019). Assessing students' social and emotional skills through triangulation of assessment methods. *OECD Education Working Papers No. 208*. <https://doi.org/10.1787/717ad7f2-en>.

- *Kim, H. S., Ahn, S. H., & Kim, C. M. (2019). A New ICT Literacy Test for Elementary and Middle School Students in Republic of Korea. *Asia-Pacific Education Researcher*, 28(3), 203-212. <https://doi.org/10.1007/s40299-018-0428-8>.
- *Kim, H. S., Kil, H. J., & Shin, A. (2014). An analysis of variables affecting the ICT literacy level of Korean elementary school students. *Computers and Education*, 77, 29-38. <https://doi.org/10.1016/j.compedu.2014.04.009>.
- *Keating, A., Green, A., Janmaat, G. (2015). *Young Adults and Politics Today: disengaged and disaffected or engaged and enraged? The latest findings from the Citizenship Education Longitudinal Study (CELS)*.
- *Keating, A., Kerr, D., Benton, T., Mundy, E., & Lopes, J. (2010). *Citizenship education in England 2001–2010: Young people’s practices and prospects for the future: The eighth and final report from the Citizenship Education Longitudinal Study (CELS)*. Londen, Engeland: DfE.
- *Kranendonk, M., Mulder, L., Thijs, P., Wanders, F., Ten Dam, G., Van der Meer, T., & Van de Werfhorst, H. (2019). *Literatuurstudie: De ontwikkeling van democratische kernwaarden*. Amsterdam: Universiteit van Amsterdam.
- *Kuhlemeier, H., Van Boxtel, H., & Van Til, A. (2012). *Balans van de sociale opbrengsten in het basisonderwijs (PPON 48)*. Arnhem: Cito.
- *Kuhlemeier, H., Knoop, H., Boxtel, H. van, Papenburg, I., & Hollenberg, J. (2016). *Wetenschappelijke verantwoording VISEON 2.0. Volginstrument voor sociaal-emotioneel functioneren*. Arnhem: Cito.
- *Kranendonk, M., Mulder, L., Thijs, P., Wanders, F., Ten Dam, G., Van der Meer, T., & Van de Werfhorst, H. (2019). *Literatuurstudie. De ontwikkeling van democratische kernwaarden*. Amsterdam: Universiteit van Amsterdam.
- Lai, E. (2011). *Metacognition: A literature review*. Pearson Research Report.

- *Lazonder, A., Gijlers, H., Janssen, N., & Walraven, A. (2018). Klaar voor de toekomst? Een longitudinaal onderzoek naar de ontwikkeling van 21e-eeuwse vaardigheden in het primair onderwijs. Nijmegen/Enschede: Rosenmullers Communicatie & Organisatie.
<https://hdl.handle.net/2066/199143>.
- *Lazonder, A., Walraven, A., Gijlers, H., & Janssen, N. (2020). Longitudinal assessment of digital literacy in children: Findings from a large Dutch single-school study. *Computers and Education*, 143. <https://doi.org/10.1016/j.compedu.2019.103681>.
- *Ledoux, G., Geijsel, F., Reumerman, R., & Ten Dam, G. (2011). Burgerschapscompetenties van jongeren in Nederland. *Pedagogische Studiën*, 88(1), 3-22.
- Ledoux, G., Meijer, J., Van der Veen, I., & Breetvelt, I. (2013). *Meetinstrumenten voor sociale competenties, metacognitie en advanced skills: Een inventarisatie*. Amsterdam: Kohnstamm Instituut.
- *Malmberg, J. (2014). *Tracing the process of self-regulated learning: Students' strategic activity in g/nStudy learning environment* (proefschrift).
<https://urn.fi/urn:isbn:9789526204703>.
- *Maslowski, R., Van der Werf, M., Oonk, G., Naayer, H., & Isac, M. M. (2012). *Burgerschapscompetenties van leerlingen in de onderbouw van het voortgezet onderwijs: Eindrapport van de international civic and citizenship education study (ICCS) in Nederland*. Groningen: GION, Rijksuniversiteit Groningen.
- McKown, C. (2019). Challenges and opportunities in the applied assessment of student social and emotional learning. *Educational Psychologist*, 54(3), 205-221.
<https://doi.org/10.1080/00461520.2019.1614446>.
- Meelissen, M., Punter, R., & Drent, M. (2014). *Digitale geletterdheid van leerlingen in het tweede leerjaar van het voortgezet onderwijs: Nederlandse resultaten van ICILS-2013*. Enschede: Universiteit Twente.

- *Munniksma, A., Dijkstra, A.B., Van der Veen, I., Ledoux, G., Van de Werfhorst, H., & Ten Dam, G. (2017). *Burgerschap in het voortgezet onderwijs: Nederland in vergelijkend perspectief*. Amsterdam: Amsterdam University Press.
- *MCEETYA. (2007). *National assessment program - ICT literacy years 6 & 10 report 2005*. Carlton South, Australië: Ministerial Council on Education, Employment, Training and Youth Affairs.
- McCoy, D. (2019). Measuring young children's executive function and self-regulation in classrooms and other real-world settings. *Clinical Child and Family Psychology Review*, 22(1), 63-73. <https://doi.org/10.1007/s10567-019-00285-1>.
- *McKown, C. (2017). Social-emotional assessment, performance, and standards. *The Future of Children*, 27(1), 157–178. <https://doi.org/10.1353/foc.2017.0008>.
- *McKown, C. (2019). Reliability, factor structure, and measurement invariance of a web-based assessment of children's social-emotional comprehension. *Journal of Psychoeducational Assessment*, 37(4), 432-449. <https://doi.org/10.1177/0734282917749682>.
- *McKown, C., Russo-Ponsaran, N. M., Allen, A. A., Johnson, J., & Russo, J. (2016). Web-based direct assessment of children's social-emotional comprehension. *Journal of Psychoeducational Assessment*, 34(4), 322-338. <https://doi.org/10.1177/0734282915604564>.
- Mislevy, R. (2007a). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, 25(4), 6-20. <https://doi.org/10.1111/j.1745-3992.2006.00075.x>.
- Mislevy, R. (2007b). Validity by design. *Educational Researcher*, 36(8), 463-469. <https://doi.org/10.3102/0013189X07311660>.

- Mislevy, R., & Haertel, G. (2006). Implications for evidence-centered design for educational assessment. *Educational Measurement: Issues and Practice*, 25, 6–20.
- Mislevy, R., Almond, R., & Lukas, J. (2003). A brief introduction to evidence-centered design. *ETS Report Series*, 2003(1), i–29. <https://doi.org/10.1002/j.2333-8504.2003.tb01908.x>.
- *NCES. (2018). *Technology & engineering literacy framework for the 2018 national assessment of educational progress*. <http://files.eric.ed.gov/fulltext/ED594359.pdf>.
- *Nesbit, J., Winne, P., Jamieson-Noel, D., Code, J., Zhou, M., MacAllister, K., Bratt, S., Wang, W., & Hadwin, A. (2006). Using cognitive tools in gStudy to investigate how study activities covary with achievement goals. *Journal of Educational Computing Research*, 35(4), 339-358. <https://doi.org/10.2190/H3W1-8321-1260-1443>.
- *Obradović, J., Sulik, M., Finch, J., & Tirado-Strayer, N. (2018). Assessing students' executive functions in the classroom: Validating a scalable group-based procedure. *Journal of Applied Developmental Psychology*, 55, 4-13. <https://doi.org/10.1016/j.appdev.2017.03.003>.
- OECD. (2007). *Understanding the social outcomes of learning*. Paris: OECD.
- OECD. (2010). *Improving health and social cohesion through education*. Paris: OECD
- OECD. (2011). *Towards an OECD skills strategy*. Parijs: OECD.
- OECD. (2015). *Skills for social progress: The power of social and emotional skills*. Parijs: OECD.
- OECD. (2018). *The future of education and skills: Education 2030*. Parijs: OECD.
- OECD. (2019). *OECD skills strategy 2019: Skills to shape a better future*. Parijs: OECD. <https://doi.org/10.1787/9789264313835-en>.
- Onderwijsraad. (2003). *Onderwijs en burgerschap*. Den Haag: Onderwijsraad.

- Opdenakker, M.C., & Ledoux, G. (2016). Zelfregulatie. In Skills-platform, *Skills voor de toekomst: Een onderzoeksagenda* (pp. 23-29). Den Haag: Ministerie van Onderwijs, Cultuur & Wetenschap.
- Panadero, E. (2017). A review of self-regulated learning: Six models and four directions for research. *Frontiers in Psychology*, 8, 1-28. <https://doi.org/10.3389/fpsyg.2017.00422>.
- *ParnasSys Academie. (2019). *Gebruikshandleiding Zien!*. Gouda: ParnasSys Academie.
- *Pearson. (2016). *Developing achievement levels on the 2014 national assessment of educational progress in grade 8 technology and engineering literacy: Technical report*. Ioaw City, IA: Pearson.
- Pepper, D. (2011). Assessing key competences across the curriculum – and Europe. *European Journal of Education*, 46(3), 335-353. <https://doi.org/10.1111/j.1465-3435.2011.01484.x>.
- Peschar, J., Hooghoff, H., Dijkstra, A.B., & Ten Dam, G. (red.)(2010), *Scholen voor burgerschap. Naar een kennisbasis voor burgerschapsonderwijs*. Antwerpen: Garant.
- *Pintrich, P., & De Groot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33–40. <https://doi.org/10.1037/0022-0663.82.1.33>.
- *Pintrich, P., Smith, D., García, T., & McKeachie, W. (1991). *A manual for the use of the motivated strategies for learning questionnaire (MSLQ)*. Ann Arbor: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.
- Platform Onderwijs2032. (2016). *Ons onderwijs2032: Eindadvies*. Den Haag: Platform Onderwijs2032.

- *Porat, E., Blau, I., & Barak, A. (2018). Measuring digital literacies: Junior high-school students' perceived competencies versus actual performance. *Computers & Education*, 126, 23-36. <https://doi.org/10.1016/j.compedu.2018.06.030>.
- Rademacher, A., & Koglin, U. (2019). The concept of self-regulation and preschoolers' social-emotional development: A systematic review. *Early Child Development and Care*, 189(14), 2299-2317. <https://doi.org/10.1080/03004430.2018.1450251>.
- Redecker, C. (2013). *The use of ICT for the assessment of key competences*. Luxembourg, Luxembourg: Publications Office of the European Union.
<https://doi.org/10.2791/87007>.
- Roelofs, E., & Sluifster, C. (2016). Meetinstrumenten. In Skills-platform, *Skills voor de toekomst: Een onderzoeksagenda* (pp. 63-72). Den Haag: Ministerie van Onderwijs, Cultuur & Wetenschap.
- *Russo, S., & Stattin, H. (2017). Stability and change in youths' political interest. *Social Indicators Research*, 132(2), 643–658. <https://doi.org/10.1007/s11205-016-1302-9>.
- Rychen, D.S., & Salganik, L.H. (2003). *Key competences for a successful life and a well-functioning society*. Cambridge, MA: Hogrefe & Huber Publishers.
- *Schulz, W., Ainley, J., Fraillon, J., Kerr, D., & Losito, B. (2010). *ICCS 2009 international report: Civic knowledge, attitudes, and engagement among lower-secondary school students in 38 countries*. Amsterdam: IEA.
- *Schulz, W., Ainley, J., Fraillon, J., Losito, B., & Agrusti, G. (2016). *IEA international civic and citizenship education study 2016: Assessment framework*. Amsterdam: IEA.
- *Schulz, W., Ainley, J., Fraillon, J., Losito, B., Agrusti, G., & Friedman, T. (2017). *Becoming citizens in a changing world: ICCS 2016 international report*. Amsterdam: IEA.

- *Schulz, W., Carstens, R., Losito, B., & Fraillon, J. (2018)(Red.). International civic and citizenship education study 2016: Technical report. Amsterdam: IEA.
- *Schraw, G., & Dennison, R. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*(4), 460-475. <https://doi.org/10.1006/ceps.1994.1033>.
- Sears, D. O., & Levy, S. (2003). Childhood and adult political development. In D. O. Sears, L. Huddy, & R. Jervis (Red.), *Oxford handbook of political psychology* (pp. 60–109). Oxford: Oxford University Press.
- Siddiq, F., Hatlevik, O. E., Olsen, R. V., 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. <https://doi.org/10.1016/j.edurev.2016.05.002>.
- Skills-platform. (2016). *Skills voor de toekomst: Een onderzoeksagenda*. Den Haag: Ministerie van Onderwijs, Cultuur & Wetenschap.
- *Sperling, R. A., Howard, B. C., Miller, L. A., & Murphy, C. (2002). Measures of children's knowledge and regulation of cognition. *Contemporary Educational Psychology, 27*(1), 51-79. <https://doi.org/10.1006/ceps.2001.1091>.
- *Sperling, R., Richmond, A., Ramsay, C., & Klapp, M. (2012). The Measurement and predictive ability of metacognition in middle school learners. *Journal of Educational Research, 105*(1), 1-7. <https://doi.org/10.1080/00220671.2010.514690>.
- *Ten Dam, G., Geijsel, F., Ledoux, G., Reumerman, R., Keunen, N., & Visser, A. (2014). *Burgerschap Meten: Handleiding primair onderwijs versie 1.2*. Rovict BV / Universiteit van Amsterdam.
- *Ten Dam, G., Geijsel, F., Ledoux, G., Reumerman, R., Keunen, N., & Visser, A. (2014). *Burgerschap Meten: Handleiding voortgezet onderwijs versie 1.2*. Rovict BV / Universiteit van Amsterdam.

- *Ten Dam, G., Geijssel, F., Reumerman, R., & Ledoux, G. (2010).
Burgerschapscompetenties: De ontwikkeling van een meetinstrument. *Pedagogische Studiën*, 87(5), 313-333.
- Ten Dam, G., Volman, M., Westerbeek, K., Wolfgram, P., Ledoux, G., & Peschar, J. (2003).
Sociale competentie langs de meetlat: Het bestrijden en voorkomen van onderwijsachterstand. Den Haag: Transferpunt Onderwijsachterstanden.
- *Thijs, P., Kranendonk, M., Mulder, L., Wanders, F., Ten Dam, G., Van der Meer, T., & Van de Werfhorst, H. (2019). *Democratische kernwaarden in het voortgezet onderwijs: Adolescentenpanel democratische kernwaarden en schoolloopbanen jaar 1 – 2018/2019*. Amsterdam: Universiteit van Amsterdam.
- *Treffers, Ph. D. A., Goedhart, A. W., Van den Bergh, B. R. H., Veerman, J. W., Ackaert, L., & De Rycke, L. (2002). *Competentiebelevingsschaal voor adolescenten (CBSA): Handleiding*. Lisse: Swets & Zeitlinger.
- Van de Werfhorst, H. (2014). Changing societies and four tasks of schooling: Challenges for strongly differentiated educational systems. *International Review of Education*, 60(1), 123-44. <https://10.1007/s11159-014-9410-8>.
- Van den Berge, W., Daas, R., Dijkstra, A.B., Ooms, T., & Ter Weel, B. (2014). *Investeren in skills en competenties*. Amsterdam/Den Haag: Universiteit van Amsterdam/Centraal Planbureau.
- *Van den Bosch, A., Groenewold, E., & Dogger-Stigter, A. M. (2009a). *Kijk! Op sociale competentie groep 3-4*. Vlissingen: Bazalt.
- *Van den Bosch, A., Groenewold, E., & Dogger-Stigter, A. M. (2009b). *Kijk! Op sociale competentie groep 5-6*. Vlissingen: Bazalt.
- *Van den Bosch, A., Groenewold, E., & Dogger-Stigter, A. M. (2009c). *Kijk! Op sociale competentie groep 7-8*. Vlissingen: Bazalt.

- Van Hout-Wolters, B. (2009). Leerstrategieën meten. Soorten meetmethoden en hun bruikbaarheid in onderwijs en onderzoek. *Pedagogische Studiën*, 86(2), 110–129.
- Van Laar, E., van Deursen, A., van Dijk, J., & de Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior*, 72, 577-588. <https://dx.doi.org/10.1016/j.chb.2017.03.010>.
- *Vandeveldde, S., Van Keer, H., & Rosseel, Y. (2013). Measuring the complexity of upper primary school children's self-regulated learning: A multi-component approach. *Contemporary Educational Psychology*, 38(4), 407-425. <https://doi.org/10.1016/j.cedpsych.2013.09.002>.
- Veenman, M., & Van Cleef, D. (2019). Measuring metacognitive skills for mathematics: students' self-reports versus on-line assessment methods. *ZDM International Journal of Mathematics Education*, 51(4), 691-701. <https://doi.org/10.1007/s11858-018-1006-5>.
- Visser, A. (2018). *Burgerschap in het basisonderwijs. Domeinbeschrijving ten behoeve van peilingsonderzoek*. Enschede: SLO Research.
- *Vliek, L. (2015). *Effects of Kanjertraining (Topper Training) on Emotional Problems, Behavioural Problems and Classroom Climate* (Proefschrift). Utrecht: Digitale Klerken.
- *Vliek, L., Riet, B., & Weide, G. (2012). *Verantwoording van de Kanjervragenlijst*. Almere: Instituut voor Kanjertrainingen.
- *Vliek, L., Weide, G., & Houtman, B. (2012). *Kanjervragenlijst. Handleiding*. Almere: Instituut voor Kanjertrainingen.
- Voogt, J., Godaert, E., Aesaert, K., & van Braak, J. (2019). *Review digitale geletterdheid*. Zwolle / Gent: Hogeschool Windesheim / Universiteit Gent

- Voogt, J., & Pareja Roblin, N. (2012). A comparative analysis of international frameworks for 21st century competences: Implications for national curriculum policies. *Journal of Curriculum Studies*, 44(3), 299-321.
- *Vorst, H. (2018). *Handleiding svl-i: Vlaamse schoolvragenlijst-internet lager en secundair onderwijs*. Antwerpen: VorstMulder.
- *Vorst, H., & Smits, J. (2004). *Schoolvragenlijst voor basisonderwijs en voortgezet onderwijs: Handleiding en verantwoording* (4^e druk). Amsterdam: Harcourt Assessment.
- *Wagenaar, H., Keune, K., & Van Weerden, J. (2012). *Balans Oordelen en argumenteren als onderdeel van actief burgerschap en sociale integratie. Uitkomsten van de peiling in 2009 (PPON 49)*. Arnhem: Cito.
- *Wagenaar, H., Van der Schoot, F., & Hemker, B. (2011). *Balans Actief burgerschap en sociale integratie. Uitkomsten van de peiling in 2009 (PPON 45)*. Arnhem: Cito.
- *Whiteley, P. (2005) *Citizenship Education Longitudinal Study - Second Literature Review* *Citizenship Education: The Political Science Perspective*. Slough: NFER.
- Wigelsworth, M., Humphrey, N., Kalambouka, A., & Lendrum, A. (2010). A review of key issues in the measurement of children's social and emotional skills. *Educational Psychology in Practice*, 26(2), 173-186. <https://doi.org/10.1080/02667361003768526>.
- *Winne, P., & Hadwin, A. (2013). nStudy: Tracing and supporting self-regulated learning in the internet. In R. Azevedo & V. Aleven (Red.), *International handbook of metacognition and learning technologies*. New York: Springer. https://doi.org/10.1007/978-1-4419-5546-3_20.
- *Winne, P., Nesbit, J., & Popowich, F. (2017). nStudy: A system for researching information problem solving. *Technology, Knowledge and Learning*, 22(3), 369-376. <https://doi.org/10.1007/s10758-017-9327-y>.

- Winne, P., & Perry, N. (2000). Measuring self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Red.), *Handbook of self-regulation* (pp. 531-566). Orlando, FL: Academic Press.
- *Winne, P., Teng, K., Chang, D., Lin, M., Marzouk, Z., Nesbit, J., Patzak, A., Raković, M., Samadi, D., & Vytasek, J. (2019). nStudy: Software for learning analytics about processes for self-regulated learning. *Journal of Learning Analytics*, 6(2), 95–106. <https://doi.org/10.18608/jla.2019.62.7>.
- Wools, S., Sanders, P., Eggen, T., Baartman, L., & Roelofs, E. (2011). Evaluatie van een beoordelingssysteem voor de kwaliteit van competentie-assessments. *Pedagogische Studiën*, 88(1), 23-40.

Bijlage 1. Resultaten instrumentanalyses

Tabel 1. Auteurs en doelgroep

#	Skill	Instrument	Auteur(s)	Jaar	Land	VVE	PO onder- bouw	VO boven- bouw	MBO	HO
1.1	Burgerschap	ADKS (Adolescentenpanel Democratische Kernwaarden en Schoolloopbanen)	Thijs et al.	2013	Nederland			X		
1.2	Burgerschap	Burgerschap meten	Ten Dam et al.	2010	Nederland		X	X		
1.3	Burgerschap	ICCS (International Civic and Citizenship Education Study)	Schulz et al.	2017	Internationaal			X		
1.4	Burgerschap	PPON 45: Actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)	Wagenaar et al.	2011	Nederland		X			
1.5	Burgerschap	PPON 49: Oordelen en argumenteren als onderdeel van actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)	Wagenaar et al.	2012	Nederland		X			
1.6	Burgerschap	YeS (Youth and Society)	Amna et al.	2015	Zweden			X	X	X X
1.7	Burgerschap	CELS (Citizenship Education Longitudinal Study)	Keating et al.	2011	Verenigd Koninkrijk		X	X	X	X X
2.1	Digitale geletterdheid	ST ² L (Student Tool for Technology Literacy)	Hohlfeld et al.	2014	Verenigde Staten			X		
2.2	Digitale geletterdheid	Performance based test ICT competences	Aesaert et al.	2014	Vlaanderen, België		X			
2.3	Digitale geletterdheid	KERIS 2014 (Korea Education Research and Information Service, K-12 test)	Jun et al.	2014	Zuid-Korea		X	X		
2.4	Digitale geletterdheid	ICILS (International Computer and Information Literacy Study)	Frailon et al.	2014	Internationaal			X		
2.5	Digitale geletterdheid	ICT literacy test	Kim et al.	2014	Zuid-Korea		X			
2.6	Digitale geletterdheid	New ICT Literacy test	Kim et al.	2019	Zuid-Korea		X	X		

#	Skill	Instrument	Auteur(s)	Jaar	Land	VVE	PO		VO		MBO	HO
							onder- bouw	boven- bouw	onder- bouw	boven- bouw		
2.7	Digitale geletterdheid	MMTL (Methods for Measuring Technological Literacy)	Avsec & Jamsek	2016	Slovenië					X		
2.8	Digitale geletterdheid	ICT literacy assessment	Hatlevik et al.	2017	Noorwegen					X		
2.9	Digitale geletterdheid	Effectiveness study: digital literacy program	Fernandez-Montalvo et al.	2017	Navarra, Spanje			X				
2.10	Digitale geletterdheid	NAP-ICT Literacy (National Assessment Program: ICT Literacy)	ACARA	2018	Australië			X	X			
2.11	Digitale geletterdheid	Measurement of actual digital literacy	Porat et al.	2018	Israël					X		
2.12	Digitale geletterdheid	TEL Framework (Technology and Engineering Literacy)	NCES	2018	Verenigde Staten					X		
2.13	Digitale geletterdheid	Authentic assessment online information literacy	Heitink	2018	Nederland			X	X			
2.14	Digitale geletterdheid	Testen digitale geletterdheid	Lazonder et al.	2018	Nederland			X				
3.1	Sociaal-emotionele ontwikkeling	Zien!	ParnasSys Academie	2016	Nederland			X	X			
3.2	Sociaal-emotionele ontwikkeling	SCOL (Sociale Competentie Observatie Lijst)	CED-Groep	2016	Nederland			X	X	X		
3.3	Sociaal-emotionele ontwikkeling	Viseon 2.0	Cito	2016	Nederland			X	X			
3.4	Sociaal-emotionele ontwikkeling	CBSA (Competentie belevingsschaal voor adolescenten)	Pearson	2002	Nederland					X	X	
3.5	Sociaal-emotionele ontwikkeling	Kanjervragenlijst	Instituut voor Kanjertrainingen	2012	Nederland				X			
3.6	Sociaal-emotionele ontwikkeling	SAQI (School Attitude Questionnaire Internet)	VorstMulder	2004	Nederland				X	X		
3.7	Sociaal-emotionele ontwikkeling	Kijk! op sociale competentie	Bazalt	2009	Nederland			X	X			
3.8	Sociaal-emotionele ontwikkeling	SELweb	xSEL Labs	2016	Verenigde Staten			X				

#	Skill	Instrument	Auteur(s)	Jaar	Land	VVE	PO		VO		MBO	HO
							onder- bouw	boven- bouw	onder- bouw	boven- bouw		
3.9	Sociaal-emotionele ontwikkeling	Zoo U	DeRosier	2014	Verenigde Staten		X					
4.1	Zelfregulatie	MSLQ (Motivated Strategies for Learning Questionnaire)	Pintrich	1991	Verenigde Staten				X	X	X	X
4.2	Zelfregulatie	CP-SRLI (Children's Perceived use of Self-Regulated Learning Inventory)	Vandevelde	2013	Vlaanderen, België			X				
4.3	Zelfregulatie	MAI (Metacognitive Awareness Inventory)	Schraw & Dennison	1994	Verenigde Staten							X
4.4	Zelfregulatie	jr.MAI	Sperling et al.	2002	Verenigde Staten			X	X			
4.5	Zelfregulatie	gStudy / nStudy	Winne et al.	2019	Canada	X	X	X	X	X	X	X
4.6	Zelfregulatie	AMES (Assessment of Motivation, Effort, and Self-regulation)	Obradović et al.	2018	Verenigde Staten			X				

Tabel 2. Uitwerking leerlingmodel

#	Instrument	Conceptueel helder onderbouwd leerlingmodel	Onderbouwing aanleerbaarheid competentie	Gedefinieerd in termen van een set mentale taakprocessen	Specificering taken en situaties	Ontwikkelingsstadia leerling	Criteriumvaliditeit	Gericht op competentie als geheel
1.1	ADKS (Adolescentenpanel Democratische Kernwaarden en Schoolloopbanen)	voldoende	voldoende	voldoende	voldoende	matig	matig	nee
1.2	Burgerschap meten	goed	goed	voldoende	voldoende	matig	matig	nee
1.3	ICCS (International Civic and Citizenship Education Study)	goed	voldoende	voldoende	voldoende	matig	matig	nee
1.4	PPON 45: Actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)	voldoende	voldoende	onvoldoende	voldoende	matig	matig	nee
1.5	PPON 49: Oordelen en argumenteren als onderdeel van actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)	voldoende	voldoende	voldoende	goed	matig	voldoende	nee
1.6	YeS (Youth and Society)	goed	voldoende	matig	voldoende	voldoende	voldoende	nee
1.7	CELS (Citizenship Education Longitudinal Study)	voldoende	voldoende	onvoldoende	voldoende	voldoende	matig	ja
2.1	ST ² L (Student Tool for Technology Literacy)	voldoende	voldoende	matig	matig	onvoldoende	voldoende	nee
2.2	Performance based test ICT competences	voldoende	voldoende	voldoende	goed	onvoldoende	matig	ja
2.3	KERIS 2014 (Korea Education Research and Information Service, K-12 test)	matig	matig	voldoende	matig	matig	matig	nee
2.4	ICILS (International Computer and Information Literacy Study)	voldoende	voldoende	goed	goed	matig	matig	ja
2.5	ICT literacy test	matig	matig	matig	matig	matig	matig	nee

#	Instrument	Conceptueel helder onderbouwd leerlingmodel	Onderbouwing aanleerbaarheid competentie	Gedefinieerd in termen van een set mentale taakprocessen	Specificering taken en situaties	Ontwikkelingsstadia leerling	Criteriumvaliditeit	Gericht op competentie als geheel
2.6	New ICT Literacy test	matig	matig	matig	matig	matig	matig	nee
2.7	MMTL (Methods for Measuring Technological Literacy)	voldoende	voldoende	goed	voldoende	matig	matig	nee
2.8	ICT literacy assessment	voldoende	matig	matig	matig	onvoldoende	matig	nee
2.9	Effectiveness study: digital literacy program	matig	matig	matig	matig	onvoldoende	matig	nee
2.10	NAP-ICT Literacy (National Assessment Program: ICT Literacy)	goed	voldoende	goed	voldoende	voldoende	voldoende	ja
2.11	Measurement of actual digital literacy	voldoende	voldoende	goed	voldoende	matig	matig	ja
2.12	TEL Framework (Technology and Engineering Literacy)	goed	voldoende	goed	matig	goed	matig	ja
2.13	Authentic assessment online information literacy	voldoende	matig	goed	voldoende	onvoldoende	voldoende	ja
2.14	Testen digitale geletterdheid	goed	voldoende	voldoende	voldoende	voldoende	voldoende	nee
3.1	Zien!	goed	voldoende	voldoende	voldoende	matig	matig	ja
3.2	SCOL (Sociale Competentie Observatie Lijst)	voldoende	voldoende	matig	matig	matig	matig	ja
3.3	Viseon 2.0	voldoende	matig	voldoende	voldoende	matig	matig	nee
3.4	CBSA (Competentie belevingsschaal voor adolescenten)	voldoende	matig	matig	matig	matig	matig	nee
3.5	Kanjervragenlijst	voldoende	matig	matig	matig	matig	voldoende	ja
3.6	SAQI (School Attitude Questionnaire Internet)	voldoende	matig	matig	matig	onvoldoende	matig	nee

#	Instrument	Conceptueel helder onderbouwd leerlingmodel	Onderbouwing aanleerbaarheid competentie	Gedefinieerd in termen van een set mentale taakprocessen	Specificering taken en situaties	Ontwikkelingsstadia leerling	Criteriumvaliditeit	Gericht op competentie als geheel
3.7	Kijk! op sociale competentie	voldoende	voldoende	matig	matig	matig	matig	ja
3.8	SELweb	voldoende	voldoende	goed	voldoende	matig	voldoende	ja
3.9	Zoo U	goed	voldoende	goed	voldoende	matig	voldoende	ja
4.1	MSLQ (Motivated Strategies for Learning Questionnaire)	voldoende	voldoende	voldoende	matig	matig	goed	ja
4.2	CP-SRLI (Children's Perceived use of Self-Regulated Learning Inventory)	matig	matig	voldoende	matig	matig	voldoende	ja
4.3	MAI (Metacognitive Awareness Inventory)	matig	matig	voldoende	matig	onvoldoende	voldoende	ja
4.4	jr.MAI	matig	matig	voldoende	matig	onvoldoende	voldoende	ja
4.5	gStudy / nStudy	voldoende	voldoende	goed	voldoende	matig	voldoende	ja
4.6	AMES (Assessment of Motivation, Effort, and Self-regulation)	voldoende	matig	voldoende	matig	matig	voldoende	ja

Tabel 3. Uitwerking taakmodel: algemeen

#	Instrument	Direct (taken) of indirect (vragenlijsten)	De (onderdelen van de) doelvaardigheid zijn uitgewerkt in relevant gedrag	De (onderdelen van de) doelvaardigheid zijn uitgewerkt in relevante situaties	De prestatiecondities zijn welomschreven	Er is beschreven is op welke manier variatie is aangebracht in complexiteit van taken	Responseformat doet recht aan de aard van de te meten doelvaardigheid - zelf construeren (open)	Responseformat doet recht aan de aard van de te meten doelvaardigheid - kiezen uit opties (gesloten)
1.1	ADKS (Adolescentenpanel Democratische Kernwaarden en Schoolloopbanen)	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig	voldoende	onvoldoende		voldoende
1.2	Burgerschap meten	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig	goed	matig		voldoende
1.3	ICCS (International Civic and Citizenship Education Study)	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig	goed	goed	voldoende	voldoende
1.4	PPON 45: Actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig	goed	voldoende		voldoende
1.5	PPON 49: Oordelen en argumenteren als onderdeel van actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)	directe taken, waarvan kandidaat zich bewust is (testtaken)	voldoende	voldoende	goed	matig	goed	
1.6	YeS (Youth and Society)	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig	goed	onvoldoende		voldoende
1.7	CELS (Citizenship Education Longitudinal Study)	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig	voldoende	onvoldoende		voldoende
2.1	ST ² L (Student Tool for Technology Literacy)	directe taken, waarvan kandidaat zich bewust is (testtaken)	goed	goed	voldoende	voldoende	goed	voldoende
2.2	Performance based test ICT competences	directe taken, waarvan kandidaat zich bewust is (testtaken)	goed	goed	matig	goed	goed	

#	Instrument	Direct (taken) of indirect (vragenlijsten)	De (onderdelen van de) doelvaardigheid zijn uitgewerkt in relevant gedrag	De (onderdelen van de) doelvaardigheid zijn uitgewerkt in relevante situaties	De prestatiecondities zijn welomschreven	Er is beschreven is op welke manier variatie is aangebracht in complexiteit van taken	Responseformat doet recht aan de aard van de te meten doelvaardigheid - zelf construeren (open)	Responseformat doet recht aan de aard van de te meten doelvaardigheid - kiezen uit opties (gesloten)
2.3	KERIS 2014 (Korea Education Research and Information Service, K-12 test)	directe taken, waarvan kandidaat zich bewust is (testtaken)	voldoende	voldoende	matig	matig		matig
2.4	ICILS (International Computer and Information Literacy Study)	directe taken, waarvan kandidaat zich bewust is (testtaken)	goed	goed	voldoende	goed	goed	voldoende
2.5	ICT literacy test	directe taken, waarvan kandidaat zich bewust is (testtaken)	matig	matig	onvoldoende	matig		
2.6	New ICT Literacy test	directe taken, waarvan kandidaat zich bewust is (testtaken)	matig	matig	onvoldoende	matig		
2.7	MMTL (Methods for Measuring Technological Literacy)	directe taken, waarvan kandidaat zich bewust is (testtaken)	voldoende	voldoende	voldoende	voldoende		voldoende
2.8	ICT literacy assessment	directe taken, waarvan kandidaat zich bewust is (testtaken)	matig	matig	voldoende	voldoende		voldoende
2.9	Effectiveness study: digital literacy program	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig	voldoende	onvoldoende		matig
2.10	NAP-ICT Literacy (National Assessment Program: ICT Literacy)	directe taken, waarvan kandidaat zich bewust is (testtaken)	goed	goed	goed	goed	goed	voldoende
2.11	Measurement of actual digital literacy	directe taken, waarvan kandidaat zich bewust is (testtaken)	voldoende	voldoende	voldoende	matig	goed	voldoende
2.12	TEL Framework (Technology and Engineering Literacy)	directe taken, waarvan kandidaat zich bewust is (testtaken)	matig	matig	matig	voldoende		
2.13	Authentic assessment online information literacy	directe taken, maar verborgen in een andere taak (testtaken)	voldoende	voldoende	voldoende	matig	goed	
2.14	Testen digitale geletterdheid	directe taken, waarvan kandidaat zich bewust is (testtaken)	voldoende	voldoende	goed	voldoende	goed	voldoende
3.1	Zien!	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig				

#	Instrument	Direct (taken) of indirect (vragenlijsten)	De (onderdelen van de) doelvaardigheid zijn uitgewerkt in relevant gedrag	De (onderdelen van de) doelvaardigheid zijn uitgewerkt in relevante situaties	De prestatiecondities zijn welomschreven	Er is beschreven is op welke manier variatie is aangebracht in complexiteit van taken	Responseformat doet recht aan de aard van de te meten doelvaardigheid - zelf construeren (open)	Responseformat doet recht aan de aard van de te meten doelvaardigheid - kiezen uit opties (gesloten)
3.2	SCOL (Sociale Competentie Observatie Lijst)	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig				
3.3	Viseon 2.0	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig				
3.4	CBSA (Competentie belevingsschaal voor adolescenten)	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig				
3.5	Kanjervragenlijst	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig				
3.6	SAQI (School Attitude Questionnaire Internet)	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig				
3.7	Kijk! op sociale competentie	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig				
3.8	SELweb	directe taken, waarvan kandidaat zich bewust is (testtaken)	matig	matig	goed	goed	goed	voldoende
3.9	Zoo U	directe taken, maar verborgen in een andere taak (testtaken)	voldoende	voldoende	goed	matig		voldoende
4.1	MSLQ (Motivated Strategies for Learning Questionnaire)	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig	goed	matig		matig
4.2	CP-SRLI (Children's Perceived use of Self-Regulated Learning Inventory)	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	matig	matig	goed	matig		matig

#	Instrument	Direct (taken) of indirect (vragenlijsten)	De (onderdelen van de) doelvaardigheid zijn uitgewerkt in relevant gedrag	De (onderdelen van de) doelvaardigheid zijn uitgewerkt in relevante situaties	De prestatiecondities zijn welomschreven	Er is beschreven is op welke manier variatie is aangebracht in complexiteit van taken	Responseformat doet recht aan de aard van de te meten doelvaardigheid - zelf construeren (open)	Responseformat doet recht aan de aard van de te meten doelvaardigheid - kiezen uit opties (gesloten)
4.3	MAI (Metacognitive Awareness Inventory)	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	onvoldoende	onvoldoende	goed	onvoldoende		matig
4.4	jr.MAI	geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten)	onvoldoende	onvoldoende	goed	onvoldoende		matig
4.5	gStudy / nStudy	directe taken, maar verborgen in een andere taak (testtaken)	voldoende	voldoende	voldoende	matig	goed	
4.6	AMES (Assessment of Motivation, Effort, and Self-regulation)	directe taken, waarvan kandidaat zich bewust is (testtaken)	matig	matig	goed	matig		voldoende

Tabel 4. Uitwerking taakmodel: directe of indirecte meting

#	Instrument	Samengesteld direct instrument				Direct onderdeel 1				Direct onderdeel 2				Samengesteld indirect instrument				Indirect (vragenlijst) onderdeel 1				Indirect (vragenlijst) onderdeel 2			
		G	K	V	A	G	K	V	A	G	K	V	A	G	K	V	A	G	K	V	A	G	K	V	A
1.1	ADKS (Adolescentenpanel Democratische Kernwaarden en Schoolloopbanen)					X																		X	X
1.2	Burgerschap meten					X																		X	X
1.3	ICCS (International Civic and Citizenship Education Study)					X																		X	X
1.4	PPON 45: Actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)																								
1.5	PPON 49: Oordelen en argumenteren als onderdeel van actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)					X																			
1.6	YeS (Youth and Society)					X																		X	
1.7	CELS (Citizenship Education Longitudinal Study)					X																		X	
2.1	ST ² L (Student Tool for Technology Literacy)												X												
2.2	Performance based test ICT competences	X																							
2.3	KERIS 2014 (Korea Education Research and Information Service, K-12 test)																							X	X
2.4	ICILS (International Computer and Information Literacy Study)					X																			
2.5	ICT literacy test																							X	
2.6	New ICT Literacy test																							X	

#	Instrument	Samengesteld direct instrument				Direct onderdeel 1				Direct onderdeel 2				Samengesteld indirect instrument				Indirect (vragenlijst) onderdeel 1				Indirect (vragenlijst) onderdeel 2			
		G	K	V	A	G	K	V	A	G	K	V	A	G	K	V	A	G	K	V	A	G	K	V	A
2.7	MMTL (Methods for Measuring Technological Literacy)		X																						
2.8	ICT literacy assessment		X																						
2.9	Effectiveness study: digital literacy program													X	X	X									
2.10	NAP-ICT Literacy (National Assessment Program: ICT Literacy)					X																	X	X	
2.11	Measurement of actual digital literacy					X																		X	
2.12	TEL Framework (Technology and Engineering Literacy)	X																							
2.13	Authentic assessment online information literacy							X																X	
2.14	Testen digitale geletterdheid		X	X																					
3.1	Zien!																	X					X		
3.2	SCOL (Sociale Competentie Observatie Lijst)																	X					X		
3.3	Viseon 2.0																		X	X			X		
3.4	CBSA (Competentie belevingsschaal voor adolescenten)															X									
3.5	Kanjervragenlijst												X												
3.6	SAQI (School Attitude Questionnaire Internet)														X										
3.7	Kijk! op sociale competentie																	X					X		
3.8	SELweb					X				X															

#	Instrument	Samengesteld direct instrument				Direct onderdeel 1				Direct onderdeel 2				Samengesteld indirect instrument				Indirect (vragenlijst) onderdeel 1				Indirect (vragenlijst) onderdeel 2			
		G	K	V	A	G	K	V	A	G	K	V	A	G	K	V	A	G	K	V	A	G	K	V	A
3.9	Zoo U	X																							
4.1	MSLQ (Motivated Strategies for Learning Questionnaire)												X												
4.2	CP-SRLI (Children's Perceived use of Self-Regulated Learning Inventory)												X												
4.3	MAI (Metacognitive Awareness Inventory)												X												
4.4	jr.MAI												X												
4.5	gStudy / nStudy	X																							
4.6	AMES (Assessment of Motivation, Effort, and Self-regulation)					X				X															

Noot: G = kennis, vaardigheden en attitude geïntegreerd; K = kennis; V = vaardigheden; A = attitudes

Tabel 5. Uitwerking bewijsmodel: score en normering

#	Instrument	Onderbouwde rationale voor combineren scores tot totaalscore voor doelvaardigheid		Interbeoordelaarsbetrouwbaarheid		Normen / standaarden beschikbaar	Betekenis en de beperkingen van de normschaal worden duidelijk gemaakt	Type normschaal is in overeenstemming met het doel van het instrument
		direct	indirect	direct	indirect			
1.1	ADKS (Adolescentenpanel Democratische Kernwaarden en Schoolloopbanen)	onvoldoende	voldoende			nee		
1.2	Burgerschap meten	voldoende	voldoende			ja	onvoldoende	voldoende
1.3	ICCS (International Civic and Citizenship Education Study)	voldoende	voldoende			ja	voldoende	goed
1.4	PPON 45: Actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)	voldoende				ja	voldoende	voldoende
1.5	PPON 49: Oordelen en argumenteren als onderdeel van actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)	matig		matig		nee		
1.6	YeS (Youth and Society)	voldoende	voldoende			nee		
1.7	CELS (Citizenship Education Longitudinal Study)	matig	matig			nee		
2.1	ST ² L (Student Tool for Technology Literacy)	voldoende				nee		
2.2	Performance based test ICT competences	goed	goed			nee		
2.3	KERIS 2014 (Korea Education Research and Information Service, K-12 test)	voldoende	voldoende			ja	onvoldoende	matig
2.4	ICILS (International Computer and Information Literacy Study)	goed	goed			ja	voldoende	goed
2.5	ICT literacy test	matig	matig			ja	matig	voldoende

#	Instrument	Onderbouwde rationale voor combineren scores tot totaalscore voor doelvaardigheid		Interbeoordelaarsbetrouwbaarheid		Normen / standaarden beschikbaar	Betekenis en de beperkingen van de normschaal worden duidelijk gemaakt	Type normschaal is in overeenstemming met het doel van het instrument
		direct	indirect	direct	indirect			
2.6	New ICT Literacy test	matig	matig			ja	matig	voldoende
2.7	MMTL (Methods for Measuring Technological Literacy)	goed	goed			nee		
2.8	ICT literacy assessment	voldoende	voldoende			nee		
2.9	Effectiveness study: digital literacy program	onvoldoende	onvoldoende			nee		
2.10	NAP-ICT Literacy (National Assessment Program: ICT Literacy)	goed	goed	voldoende		ja	goed	goed
2.11	Measurement of actual digital literacy	goed	goed		voldoende	nee		
2.12	TEL Framework (Technology and Engineering Literacy)	voldoende	voldoende	voldoende		ja	matig	voldoende
2.13	Authentic assessment online information literacy	onvoldoende	onvoldoende	voldoende	voldoende	nee		
2.14	Testen digitale geletterdheid	voldoende	voldoende		voldoende	nee		
3.1	Zien!			matig		ja	voldoende	goed
3.2	SCOL (Sociale Competentie Observatie Lijst)	voldoende	voldoende	matig		ja	matig	goed
3.3	Viseon 2.0	goed	goed	matig		ja	voldoende	goed
3.4	CBSA (Competentie belevingsschaal voor adolescenten)					ja		
3.5	Kanjervragenlijst		voldoende			ja	voldoende	goed

#	Instrument	Onderbouwde rationale voor combineren scores tot totaalscore voor doelvaardigheid		Interbeoordelaarsbetrouwbaarheid		Normen / standaarden beschikbaar	Betekenis en de beperkingen van de normschaal worden duidelijk gemaakt	Type normschaal is in overeenstemming met het doel van het instrument
		direct	indirect	direct	indirect			
		3.6	SAQI (School Attitude Questionnaire Internet)		voldoende			
3.7	Kijk! op sociale competentie	onvoldoende	onvoldoende			nee		
3.8	SELweb	voldoende		voldoende		nee		
3.9	Zoo U	goed				ja	voldoende	goed
4.1	MSLQ (Motivated Strategies for Learning Questionnaire)					nee		
4.2	CP-SRLI (Children's Perceived use of Self-Regulated Learning Inventory)					nee		
4.3	MAI (Metacognitive Awareness Inventory)					nee		
4.4	jr.MAI		matig			nee		
4.5	gStudy / nStudy					nee		
4.6	AMES (Assessment of Motivation, Effort, and Self-regulation)	matig				nee		

Tabel 6. Uitwerking bewijsmodel: betrouwbaarheid

#	Instrument	Betrouwbaarheidsgegevens beschikbaar	Betrouwbaarheidsgegevens berekend voor doelpopulatie	Betrouwbaarheid voldoende voor een individuele, <i>belangrijke</i> beslissing (zoals een examen)				Betrouwbaarheid voldoende voor een individuele, <i>minder</i> belangrijke beslissing (zoals een formatieve toets, deelttoets)				Betrouwbaarheid voldoende voor een conclusie/beslissing op stelselniveau (zoals een landelijke peiling)			
				direct		indirect		direct		indirect		direct		indirect	
				Hoofd-schalen	Sub-schalen	Hoofd-schalen	Sub-schalen	Hoofd-schalen	Sub-schalen	Hoofd-schalen	Sub-schalen	Hoofd-schalen	Sub-schalen	Hoofd-schalen	Sub-schalen
1.1	ADKS (Adolescentenpanel Democratische Kernwaarden en Schoolloopbanen)	nee													
1.2	Burgerschap meten	ja	goed	voldoende	onvoldoende	goed	matig	goed	matig	goed	voldoende	goed	voldoende	goed	goed
1.3	ICCS (International Civic and Citizenship Education Study)	ja	goed				matig				voldoende	goed			goed
1.4	PPON 45: Actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)	nee													
1.5	PPON 49: Oordelen en argumenteren als onderdeel van actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)	nee													
1.6	YeS (Youth and Society)	nee													
1.7	CELS (Citizenship Education Longitudinal Study)	nee													
2.1	ST ² L (Student Tool for Technology Literacy)	ja	voldoende	goed	voldoende			goed	goed			goed	goed		
2.2	Performance based test ICT competences	ja	goed	goed				goed				goed			
2.3	KERIS 2014 (Korea Education Research and Information Service, K-12 test)	nee													
2.4	ICILS (International Computer and Information Literacy Study)	ja	goed	goed				goed				goed			

#	Instrument	Betrouwbaarheidsgegevens beschikbaar	Betrouwbaarheidsgegevens berekend voor doelpopulatie	Betrouwbaarheid voldoende voor een individuele, <i>belangrijke</i> beslissing (zoals een examen)				Betrouwbaarheid voldoende voor een individuele, <i>minder</i> belangrijke beslissing (zoals een formatieve toets, deoltoets)				Betrouwbaarheid voldoende voor een conclusie/beslissing op stelselniveau (zoals een landelijke peiling)			
				direct	indirect	direct	indirect	direct	indirect	direct	indirect	direct	indirect	direct	indirect
				Hoofd-schalen	Sub-schalen	Hoofd-schalen	Sub-schalen	Hoofd-schalen	Sub-schalen	Hoofd-schalen	Sub-schalen	Hoofd-schalen	Sub-schalen	Hoofd-schalen	Sub-schalen
2.5	ICT literacy test	nee													
2.6	New ICT Literacy test	ja	matig	matig				matig				voldoende			
2.7	MMTL (Methods for Measuring Technological Literacy)	ja	voldoende	voldoende				voldoende				goed			
2.8	ICT literacy assessment	ja	goed	goed				goed				goed			
2.9	Effectiveness study: digital literacy program	nee													
2.10	NAP-ICT Literacy (National Assessment Program: ICT Literacy)	ja	goed	goed				goed				goed			
2.11	Measurement of actual digital literacy	ja	voldoende	voldoende				voldoende				goed			
2.12	TEL Framework (Technology and Engineering Literacy)	nee													
2.13	Authentic assessment online information literacy	nee													
2.14	Testen digitale geletterdheid	ja	voldoende	voldoende				voldoende				goed			
3.1	Zien!	ja	goed		voldoende		matig		goed		voldoende		goed		goed
3.2	SCOL (Sociale Competentie Observatie Lijst)	ja	goed	goed	matig	matig	onvoldoende	goed	voldoende	voldoende	onvoldoende	goed	voldoende	voldoende	onvoldoende
3.3	Viseon 2.0	ja	goed		voldoende		voldoende		goed		goed		goed		goed
3.4	CBSA (Competentie belevingsschaal voor adolescenten)	ja	voldoende			matig				voldoende			goed		

#	Instrument	Betrouwbaarheidsgegevens beschikbaar	Betrouwbaarheidsgegevens berekend voor doelpopulatie	Betrouwbaarheid voldoende voor een individuele, <i>belangrijke</i> beslissing (zoals een examen)				Betrouwbaarheid voldoende voor een individuele, <i>minder</i> belangrijke beslissing (zoals een formatieve toets, deelttoets)				Betrouwbaarheid voldoende voor een conclusie/beslissing op stelselniveau (zoals een landelijke peiling)			
				direct		indirect		direct		indirect		direct		indirect	
				Hoofd-schalen	Sub-schalen	Hoofd-schalen	Sub-schalen	Hoofd-schalen	Sub-schalen	Hoofd-schalen	Sub-schalen	Hoofd-schalen	Sub-schalen	Hoofd-schalen	Sub-schalen
3.5	Kanjervragenlijst	ja	goed			voldoende				goed				goed	
3.6	SAQI (School Attitude Questionnaire Internet)	ja	goed			goed	voldoende			goed	goed			goed	goed
3.7	Kijk! op sociale competentie	nee													
3.8	SELweb	ja	goed	voldoende	matig			goed	voldoende			goed	goed		
3.9	Zoo U	ja	goed	voldoende				goed				goed			
4.1	MSLQ (Motivated Strategies for Learning Questionnaire)	ja	goed				matig				voldoende				goed
4.2	CP-SRLI (Children's Perceived use of Self-Regulated Learning Inventory)	ja	voldoende				matig				voldoende				goed
4.3	MAI (Metacognitive Awareness Inventory)	ja	voldoende			goed				goed				goed	
4.4	jr.MAI	ja	voldoende			voldoende				goed				goed	
4.5	gStudy / nStudy	nee													
4.6	AMES (Assessment of Motivation, Effort, and Self-regulation)	ja	voldoende	onvoldoende	goed			matig	goed			voldoende	goed		

Tabel 7. Uitwerking bewijsmodel: beoordelaar

#	Instrument	Leerling (self-assessment)		Docent		Medeleerling (peer-assessment)		Specialist (zoals orthopedagoog, testpsycholoog)		Onderzoeker		Software-algoritme	
		direct	indirect	direct	indirect	direct	indirect	direct	indirect	direct	indirect	direct	indirect
1.1	ADKS (Adolescentenpanel Democratische Kernwaarden en Schoolloopbanen)		X							X			
1.2	Burgerschap meten		X							X			
1.3	ICCS (International Civic and Citizenship Education Study)		X							X			
1.4	PPON 45: Actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)									X			
1.5	PPON 49: Oordelen en argumenteren als onderdeel van actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)									X			
1.6	YeS (Youth and Society)		X							X			
1.7	CELS (Citizenship Education Longitudinal Study)		X										
2.1	ST ² L (Student Tool for Technology Literacy)									X	X	X	X
2.2	Performance based test ICT competences							X	X			X	X
2.3	KERIS 2014 (Korea Education Research and Information Service, K-12 test)												
2.4	ICILS (International Computer and Information Literacy Study)									X	X	X	X
2.5	ICT literacy test												
2.6	New ICT Literacy test												

#	Instrument	Leerling (self-assessment)		Docent		Medeleerling (peer-assessment)		Specialist (zoals orthopedagoog, testpsycholoog)		Onderzoeker		Software-algoritme	
		direct	indirect	direct	indirect	direct	indirect	direct	indirect	direct	indirect	direct	indirect
2.7	MMTL (Methods for Measuring Technological Literacy)									X	X		
2.8	ICT literacy assessment											X	X
2.9	Effectiveness study: digital literacy program									X	X		
2.10	NAP-ICT Literacy (National Assessment Program: ICT Literacy)									X	X		
2.11	Measurement of actual digital literacy									X	X		
2.12	TEL Framework (Technology and Engineering Literacy)									X	X		
2.13	Authentic assessment online information literacy									X	X	X	X
2.14	Testen digitale geletterdheid									X	X		
3.1	Zien!		X	X									
3.2	SCOL (Sociale Competentie Observatie Lijst)		X	X									
3.3	Viseon 2.0		X	X									
3.4	CBSA (Competentie belevingsschaal voor adolescenten)		X										
3.5	Kanjervragenlijst		X										
3.6	SAQI (School Attitude Questionnaire Internet)		X										
3.7	Kijk! op sociale competentie		X	X									

#	Instrument	Leerling (self-assessment)		Docent		Medeleerling (peer-assessment)		Specialist (zoals orthopedagoog, testpsycholoog)		Onderzoeker		Software-algoritme	
		direct	indirect	direct	indirect	direct	indirect	direct	indirect	direct	indirect	direct	indirect
3.8	SELweb									X			X
3.9	Zoo U												X
4.1	MSLQ (Motivated Strategies for Learning Questionnaire)		X										
4.2	CP-SRLI (Children's Perceived use of Self-Regulated Learning Inventory)		X										
4.3	MAI (Metacognitive Awareness Inventory)		X										
4.4	jr.MAI		X										
4.5	gStudy / nStudy									X			X
4.6	AMES (Assessment of Motivation, Effort, and Self-regulation)												X

Tabel 8. Uitwerking assemblagemodel

#	Instrument	Kenmerken van taaksituaties waarin (onderdelen van) de vaardigheid moet worden getoond zijn voldoende gespecificeerd		Scores van de leerling op doelvaardigheid zijn gebaseerd op bewijs uit taken waarnaar wordt gegeneraliseerd		Handleiding beschikbaar?
		direct	indirect	direct	indirect	
1.1	ADKS (Adolescentenpanel Democratische Kernwaarden en Schoolloopbanen)	matig	voldoende	voldoende	voldoende	geen handleiding
1.2	Burgerschap meten	voldoende	voldoende	voldoende	voldoende	gecombineerde technische en inhoudelijke handleiding
1.3	ICCS (International Civic and Citizenship Education Study)	voldoende	voldoende	voldoende	voldoende	technische handleiding
1.4	PPON 45: Actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)	voldoende		voldoende		geen handleiding
1.5	PPON 49: Oordelen en argumenteren als onderdeel van actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)	goed		matig		geen handleiding
1.6	YeS (Youth and Society)	matig	voldoende	matig	voldoende	
1.7	CELS (Citizenship Education Longitudinal Study)	matig	matig	voldoende	voldoende	geen handleiding
2.1	ST ² L (Student Tool for Technology Literacy)	matig		voldoende		geen handleiding
2.2	Performance based test ICT competences	voldoende	voldoende	voldoende	voldoende	geen handleiding
2.3	KERIS 2014 (Korea Education Research and Information Service, K-12 test)	matig	matig	matig	matig	geen handleiding
2.4	ICILS (International Computer and Information Literacy Study)	voldoende	voldoende	voldoende	voldoende	inhoudelijke handleiding
2.5	ICT literacy test	matig	matig	matig	matig	
2.6	New ICT Literacy test	matig	matig	matig	matig	geen handleiding

#	Instrument	Kenmerken van taaksituaties waarin (onderdelen van) de vaardigheid moet worden getoond zijn voldoende gespecificeerd		Scores van de leerling op doelvaardigheid zijn gebaseerd op bewijs uit taken waarnaar wordt gegeneraliseerd		Handleiding beschikbaar?
		direct	indirect	direct	indirect	
		2.7	MMTL (Methods for Measuring Technological Literacy)	voldoende	voldoende	
2.8	ICT literacy assessment	voldoende	voldoende	matig	matig	geen handleiding
2.9	Effectiveness study: digital literacy program	matig	matig	matig	matig	geen handleiding
2.10	NAP-ICT Literacy (National Assessment Program: ICT Literacy)	goed	goed	voldoende	voldoende	inhoudelijke handleiding
2.11	Measurement of actual digital literacy	voldoende	voldoende	voldoende	voldoende	geen handleiding
2.12	TEL Framework (Technology and Engineering Literacy)	matig	matig	matig	matig	geen handleiding
2.13	Authentic assessment online information literacy	matig	matig	voldoende	voldoende	geen handleiding
2.14	Testen digitale geletterdheid	voldoende	voldoende	voldoende	voldoende	geen handleiding
3.1	Zien!	voldoende	matig	goed	voldoende	inhoudelijke handleiding
3.2	SCOL (Sociale Competentie Observatie Lijst)	matig	matig	voldoende	voldoende	inhoudelijke handleiding
3.3	Viseon 2.0	voldoende	voldoende	goed	goed	gecombineerde technische en inhoudelijke handleiding
3.4	CBSA (Competentie belevingsschaal voor adolescenten)		matig		voldoende	inhoudelijke handleiding
3.5	Kanjervragenlijst		matig		voldoende	gecombineerde technische en inhoudelijke handleiding
3.6	SAQI (School Attitude Questionnaire Internet)		matig		voldoende	gecombineerde technische en inhoudelijke handleiding
3.7	Kijk! op sociale competentie	matig	matig	voldoende	voldoende	inhoudelijke handleiding

#	Instrument	Kenmerken van taaksituaties waarin (onderdelen van) de vaardigheid moet worden getoond zijn voldoende gespecificeerd		Scores van de leerling op doelvaardigheid zijn gebaseerd op bewijs uit taken waarnaar wordt gegeneraliseerd		Handleiding beschikbaar?
		direct	indirect	direct	indirect	
		3.8	SELweb	voldoende		
3.9	Zoo U	goed		voldoende		geen handleiding
4.1	MSLQ (Motivated Strategies for Learning Questionnaire)		matig		matig	inhoudelijke handleiding
4.2	CP-SRLI (Children's Perceived use of Self-Regulated Learning Inventory)		matig		voldoende	geen handleiding
4.3	MAI (Metacognitive Awareness Inventory)		onvoldoende		voldoende	geen handleiding
4.4	jr.MAI		onvoldoende		voldoende	geen handleiding
4.5	gStudy / nStudy	voldoende		voldoende		geen handleiding
4.6	AMES (Assessment of Motivation, Effort, and Self-regulation)	voldoende		matig		geen handleiding

Tabel 9. Uitwerking presentatiemodel

#	Instrument	Waar vindt de taakuitvoering door de leerling ter meting van de					Waar vindt de data-opslag plaats?			
		Niet gespecificeerd	Fysieke situatie binnen school	Fysieke situatie buiten school	Computer-based (webapp)	Papier en pen	Niet gespecificeerd	Op lokale computers van scholen of onderzoeksinstellingen	Vrij toegankelijke web-omgeving van derden	Afgeschermd web-omgeving van derden
1.1	ADKS (Adolescentenpanel Democratische Kernwaarden en Schoolloopbanen)				X			X		
1.2	Burgerschap meten				X	X	X			
1.3	ICCS (International Civic and Citizenship Education Study)					X		X		
1.4	PPON 45: Actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)						X			
1.5	PPON 49: Oordelen en argumenteren als onderdeel van actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)					X	X			
1.6	YeS (Youth and Society)					X	X			
1.7	CELS (Citizenship Education Longitudinal Study)					X	X			
2.1	ST ² L (Student Tool for Technology Literacy)				X		X			
2.2	Performance based test ICT competences				X			X	X	
2.3	KERIS 2014 (Korea Education Research and Information Service, K-12 test)	X					X			
2.4	ICILS (International Computer and Information Literacy Study)				X			X		
2.5	ICT literacy test	X					X			

#	Instrument	Waar vindt de taakuitvoering door de leerling ter meting van de					Waar vindt de data-opslag plaats?			
		Niet gespecificeerd	Fysieke situatie binnen school	Fysieke situatie buiten school	Computer-based (webapp)	Papier en pen	Niet gespecificeerd	Op lokale computers van scholen of onderzoeksinstellingen	Vrij toegankelijke web-omgeving van derden	Afgeschermd web-omgeving van derden
2.6	New ICT Literacy test				X		X			
2.7	MMTL (Methods for Measuring Technological Literacy)					X		X		
2.8	ICT literacy assessment				X				X	
2.9	Effectiveness study: digital literacy program					X	X			
2.10	NAP-ICT Literacy (National Assessment Program: ICT Literacy)				X			X	X	
2.11	Measurement of actual digital literacy				X			X		
2.12	TEL Framework (Technology and Engineering Literacy)				X		X			
2.13	Authentic assessment online information literacy				X			X	X	
2.14	Testen digitale geletterdheid				X	X		X		
3.1	Zien!				X				X	
3.2	SCOL (Sociale Competentie Observatie Lijst)				X				X	
3.3	Viseon 2.0				X				X	
3.4	CBSA (Competentie belevingsschaal voor adolescenten)					X	X			
3.5	Kanjervragenlijst				X		X			

#	Instrument	Waar vindt de taakuitvoering door de leerling ter meting van de					Waar vindt de data-opslag plaats?			
		Niet gespecificeerd	Fysieke situatie binnen school	Fysieke situatie buiten school	Computer-based (webapp)	Papier en pen	Niet gespecificeerd	Op lokale computers van scholen of onderzoeksinstellingen	Vrij toegankelijke web-omgeving van derden	Afgeschermd web-omgeving van derden
3.6	SAQI (School Attitude Questionnaire Internet)				X		X			
3.7	Kijk! op sociale competentie				X		X			
3.8	SELweb				X		X			
3.9	Zoo U				X		X			
4.1	MSLQ (Motivated Strategies for Learning Questionnaire)					X		X		
4.2	CP-SRLI (Children's Perceived use of Self-Regulated Learning Inventory)					X	X			
4.3	MAI (Metacognitive Awareness Inventory)					X	X			
4.4	jr.MAI					X	X			
4.5	gStudy / nStudy				X		X			
4.6	AMES (Assessment of Motivation, Effort, and Self-regulation)				X		X			

Tabel 10. Bronnen en websites

#	Instrument	Publicaties						Websites
		1	2	3	4	5	6	
1.1	ADKS (Adolescentenpanel Democratische Kernwaarden en Schoolloopbanen)	Kranendonk, Mulder, Thijs, Wanders, Ten Dam, Van der Meer, & Van de Werfhorst (2019)	Thijs, Kranendonk, Mulder, Wanders, Ten Dam, Van der Meer, & Van de Werfhorst (2019)					https://adks.nl/
1.2	Burgerschap meten	Geboers, Geijsel, Admiraal, Jorgensen, & Ten Dam (2015)	Kuhlemeier, Van Boxtel, & Van Til (2012)	Ledoux, Geijsel, Reumerman, & Ten Dam (2011)	Ten Dam, Geijsel, Ledoux, Reumerman, Keunen, & Visser (2014)	Ten Dam, Geijsel, Ledoux, Reumerman, Keunen, & Visser (2014)	Ten Dam, Geijsel, Reumerman, & Ledoux (2010)	http://www.cool5-18.nl/
1.3	ICCS (International Civic and Citizenship Education Study)	Maslowski, Van der Werf, Oonk, Naayer, & Isac (2012)	Munniksmas, Dijkstra, Van der Veen, Ledoux, Van de Werfhorst, & Ten Dam (2017)	Schulz, Ainley, Fraillon, Kerr, & Losito (2010)	Schulz, Ainley, Fraillon, Losito, & Agrusti (2016)	Schulz, Ainley, Fraillon, Losito, Agrusti, & Friedman (2017)	Schulz, Carstens, Losito, & Fraillon (2018)	https://www.iccsnederland.nl/
1.4	PPON 45: Actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)	Wagenaar, Van der Schoot, & Hemker (2011)						
1.5	PPON 49: Oordelen en argumenteren als onderdeel van actief burgerschap en sociale integratie (Periodieke Peiling Onderwijsniveau Nederland)	Wagenaar, Keune, & Van Weerden (2012)						
1.6	YeS (Youth and Society)	Amnå (2012)	Amnå, Ekström, Kerr, & Stattin (2009)	Amnå, Ekström, Kerr, & Stattin (2015)	Russo & Stattin (2017)			https://www.oru.se/english/research/research-environments/hs/yes/
1.7	CELS (Citizenship Education Longitudinal Study)	Keating, Green, Janmaat (2015)	Keating, Kerr, Benton, Mundy, & Lopes (2010)	Whiteley (2005)				

#	Instrument	Publicities						Websites
		1	2	3	4	5	6	
2.1	ST ² L (Student Tool for Technology Literacy)	Hohlfeld, Ritzhaupt, & Barron (2010)	Hohlfeld, Ritzhaupt, & Barron (2013)	Huggins, Ritzhaupt, & Dawson (2014)				
2.2	Performance based test ICT competences	Aesaert, Van Nijlen, Vanderlinde, & Van Braak (2014)	Aesaert, Van Braak, Van Nijlen, Vanderlinde (2015)	Aesaert, Van Nijlen, Vanderlinde, Tondeur, Devlieger, & Van Braak (2015)	Aesaert, Voogt, Kuiper, & Van Braak (2017)			
2.3	KERIS 2014 (Korea Education Research and Information Service, K-12 test)	Jun, Han, Kim, & Lee (2014)						
2.4	ICILS (International Computer and Information Literacy Study)	Fraillon, Ainley, Schulz, Duckworth, & Friedman (2019)	Fraillon, Ainley, Schulz, Friedman, & Gebhardt (2014)	Fraillon, Ainley, Schulz, Friedman, & Gebhardt (2019)	Fraillon, Schulz, Friedman, & Ainley (2013)	Fraillon, Schulz, Friedman, & Meyer (i.v.)		
2.5	ICT literacy test	Kim, Kil, & Shin (2014)						
2.6	New ICT Literacy test	Kim, Ahn, & Kim (2019)						
2.7	MMTL (Methods for Measuring Technological Literacy)	Avsec & Jamsek (2016)	Avsec & Szewczyk-Zakrzewska (2016)	Avsec & Jamsek (2018)				
2.8	ICT literacy assessment	Hatlevik, Scherer, & Christophersen (2017)						
2.9	Effectiveness study: digital literacy program	Fernández-Montalvo, Peñalva, Irazabal, & López (2017)						

#	Instrument	Publicaties						Websites
		1	2	3	4	5	6	
2.10	NAP-ICT Literacy (National Assessment Program: ICT Literacy)	ACARA (2018a)	ACARA (2018b)	ACER (2008)	MCEETYA (2007)			
2.11	Measurement of actual digital literacy	Porat, Blau, & Barak (2018)						
2.12	TEL Framework (Technology and Engineering Literacy)	NCES (2018)	Pearson (2016)					
2.13	Authentic assessment online information literacy	Heitink (2018)						
2.14	Testen digitale geletterdheid	Lazonde, Gijlers, Janssen, & Walraven (2018)	Lazonder, Walraven, Gijlers, & Janssen (2020)					
3.1	Zien!	Broer, Haverhals, & de Bruin (2012)	Broer, Haverhals, Grootenhuis, & van der Klooster-Sturm (2016a)	Broer, Haverhals, Grootenhuis, & van der Klooster-Sturm (2016b)	ParnasSys Academie (2019)			https://www.parnassys.nl/producten/modules/zien
3.2	SCOL (Sociale Competentie Observatie Lijst)	Joosten (2006)	Joosten (2007)	Joosten, ten Heggeler, & Mouissie (2016)	Joosten, van Bokkem, ten Heggeler, Pot, & van Dijk (2009)			https://www.scol.nl/
3.3	Viseon 2.0	Cito (2015)	Kuhlemeier, Van Boxtel, & Van Til (2012)	Kuhlemeier, Knoop, Van Boxtel, Papenburg, & Hollenberg (2016)				https://www.cito.nl/onderwijs/primair-onderwijs/lvs-toetsen/toetsen/viseon

#	Instrument	Publicaties						Websites
		1	2	3	4	5	6	
3.4	CBSA (Competentie belevingsschaal voor adolescenten)	Treffers, Goedhart, Van den Bergh, Veerman, Ackaert, & De Rycke (2002)						https://www.pearsonclinical.nl/cbsa-competentie-belevingsschaal-voor-adolescenten
3.5	Kanjervragenlijst	Vliek (2015)	Vliek, Riet, & Weide (2012)	Vliek, Weide, & Houtman (2012)				https://www.kanjertraining.nl/
3.6	SAQI (School Attitude Questionnaire Internet)	Vorst (2018)	Vorst & Smits (2004)					http://www.svlsaqi.nl/
3.7	Kijk! op sociale competentie	Bazalt (2018)	Van den Bosch, Groenewold, & Dogger-Stigter (2009a)	Van den Bosch, Groenewold, & Dogger-Stigter (2009b)	Van den Bosch, Groenewold, & Dogger-Stigter (2009c)			https://www.bazalt.nl/expertise-kijk/kijk-webbased
3.8	SELweb	McKown (2017)	McKown (2019)	McKown, Russo-Ponsaran, Allen, Johnson, & Russo (2016)				https://xsel-labs.com/
3.9	Zoo U	Craig, DeRosier, & Watanabe (2015)	DeRosier (2014)	DeRosier, Craig, & Sanchez (2012)	DeRosier & Thomas (2018)			https://www.centervention.com/social-skills-game/
4.1	MSLQ (Motivated Strategies for Learning Questionnaire)	Credé & Phillips (2011)	Duncan & McKeachie (2005)	Pintrich & De Groot (1990)	Pintrich, Smith, García, & McKeachie (1991)			
4.2	CP-SRLI (Children's Perceived use of Self-Regulated Learning Inventory)	Heirweg, De Smul, Devos, & Van Keer (2019)	Vandeveld, Van Keer, & Rosseel (2013)					
4.3	MAI (Metacognitive Awareness Inventory)	Harrison & Vallin (2018)	Schraw & Dennison (1994)					

#	Instrument	Publicities						Websites
		1	2	3	4	5	6	
4.4	jr.MAI	Sperling, Howard, Miller, & Murphy, (2002)	Sperling, Richmond, Ramsay, & Klapp (2012)					
4.5	gStudy / nStudy	Malmberg (2014)	Hadwin, Nesbit, Code, Jamieson-Noel, & Winne (2007)	Nesbit, Winne, Jamieson-Noel, Code, Zhou, MacAllister, Bratt, Wang, & Hadwin (2006)	Winne & Hadwin (2013)	Winne, Nesbit, & Popowich (2017)	Winne, Teng, Chang, Lin, Marzouk, Nesbit, Patzak, Raković, Samadi, & Vytasek (2019)	https://www.sfu.ca/edpsychlab/nstudy.html
4.6	AMES (Assessment of Motivation, Effort, and Self-regulation)	Garcia, Sulik, & Obradović (2019)	Obradović, Sulik, Finch, & Tirado-Strayer (2018)					https://gse-it.stanford.edu/ames

Bijlage 2. Gehanteerd screeningsinstrument voor skills-instrumenten, gebaseerd op het beoordelingssysteem voor de kwaliteit van competentie-assessments van Wools et al. (2011) en aangevuld met een uitwerking van principes van Evidence-Centered Design (Roelofs & Sluijter, 2016)

Onderdeel	Toelichting	Antwoordmogelijkheden
0. Algemene vragen over het soort instrument		
1. Naam	Naam van het instrument	
2. Auteurs	Auteur(s)	
3. Jaar	Jaar meest recente uitgifte-verantwoording instrument	
4. Land	Land van ontwikkeling-uitgifte	
5. Skill	Welke (advanced) skill staat centraal in het instrument?	1) Burgerschap; 2) Digitale geletterdheid; 3) Sociaal-emotionele ontwikkeling; 4) Zelfregulatie
6. Type	Soort instrument	1) Onderzoeksinstrument t.b.v fundamenteel onderzoek; 2) Onderzoeksinstrument t.b.v peilingsonderzoek; 3) Instrument t.b.v gebruik in de praktijk; 4) Anders, namelijk:
1. Leerlingmodel		
11. Conceptueel helder onderbouwd leerlingmodel	De te beoordelen doelvaardigheid en onderdelen daarvan zijn gebaseerd op een conceptueel helder onderbouwd leerlingmodel	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
12. Aanleerbaarheid	In de onderbouwing worden een of meer bewijzen aangeleverd dat de doelvaardigheid (onderdelen) aanleerbaar is.	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
13. Criteriumvaliditeit	Het wordt voldoende duidelijk hoe presteren op (onderdelen van) de vaardigheid samenhangt met de veronderstelde prestaties in specifieke leeropbrengsten of prestaties in praktijkdomeinen, al dan niet in latere levensfasen	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
14. Taakprocessen	De doelvaardigheid wordt gedefinieerd in termen van een set mentale taakprocessen die worden ingezet bij taakuitvoering.	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
15. Ontwikkelingsstadia	Stadia in verwerving van de doelvaardigheid (leerprogressies) worden specifiek toegelicht	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed

Onderdeel	Toelichting	Antwoordmogelijkheden
16. Deel of geheelbenadering?	6. Deel of geheelbenadering: Het instrument is gericht op de volgende onderdelen (K, A, V), dan wel een integratieve meting van de skills als geheel (meerdere antwoorden mogelijk)	1) Integratieve meting skill (competentie) als geheel; 2) Kennis; 3) vaardigheid; 4) attitude
17. Specificering taken en situaties	In de theoretische onderbouwing worden (typerende) taken en taaksituaties vermeld waarin de kundigheid of bereidheid van de leerling zich manifesteert (zoals tekstgenres en leesdoelen bij geletterdheid)	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
18. Doelgroep: breedte leeftijdsrange en schooltype	Geef de doelgroepen van het instrument aan op de volgende leeftijdsgroep en schooltypen.	Vroeg- en voorschoolse educatie (VVE; 3-5 jaar); Basisschool onderbouw (6-9); Basisschool bovenbouw (10-12); Voortgezet onderwijs onderbouw: 12-15 jaar; Voortgezet onderwijs bovenbouw: 15-18 jaar; 15-18 jaar: MBO; 18-24 jaar: MBO; 18-24 jaar: HO
2. Taakmodel		
21. Operationalisatie in taken	De (onderdelen van de) doelvaardigheid is (zijn) uitgewerkt in assessmenttaken, waarop de leerling kundigheid of bereidheid kan demonstreren.	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
21a. Operationalisatie in taakrelevant gedrag	De (onderdelen van de) doelvaardigheid zijn uitgewerkt in relevant gedrag.	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
21b. Operationalisatie situatiekenmerken	Bij de operationalisatie in taken of vragen worden kenmerken van situaties onderscheiden, waarin de doelvaardigheid toegepast dient te worden	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
21c. Basis van score of uitslag	Geef aan wat de basis is van een score of uitslag op een instrument is (K, V, A afzonderlijk of integratie daarvan). Indien nodig uitsplitsen naar instrumentonderdelen	Samengestelde vragenlijst (niet gesplitst in onderdelen); indirect (vragenlijst); Samengestelde directe test (niet gesplitst in onderdelen); direct (toets of test)
21d. Directheid van taken	Directheid van taken: Geef aan in hoeverre bij onderdelen van het instrument prestaties over leerlingen worden beoordeeld met behulp van directe taken ofwel zonder gebruik van directe taken.	1) geen directe taken, algemeen retrospectief oordeel (vragenlijsten); 2) geen directe taken, retrospectief oordeel over aangeduide situaties (contextrijke vragenlijsten); 3) directe taken, waarvan kandidaat zich bewust is (testtaken); 4) directe taken, maar verborgen in een andere taak (testtaken); 5) Andere directe taakvorm, namelijk
22. Prestatiecondities	De prestatiecondities zijn welomschreven: mate van voorstructurering van de taak, te hanteren hulpmiddelen, eisen aan de uitwerking van antwoorden of producten.	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed

Onderdeel	Toelichting	Antwoordmogelijkheden
23. Taakcomplexiteitsfactoren	Er is in voldoende mate beschreven is op welke manier variatie is aangebracht in complexiteit van taken waardoor de vaardigheid lichter of zwaarder wordt belast (de opgeroepen mentale of affectieve taaklast)	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
24. Responseformat	Zelf te construeren of keuze uit alternatieven	1) zelf construeren 2) kiezen uit opties; 3) combinaties
25. Geschiktheid responsekanaal	Het gekozen responsekanaal doet recht aan de aard van de te meten doelvaardigheid. Beoordeel dit voor de kanalen die van toepassing zijn.	Verbaal (geschreven, gesproken), visueel (uitgebeeld), fysiek (handelend, bedienend), symbolisch (getallen, tekens, symbolen)
3. Bewijsmodel - beoordelaar en aard responsedata		
31a. Beoordeling van indirecte prestatie	Duid aan wie betrokken is bij de BEOORDELING (niet het doorlopen als kandidaat) van de skill(onderdelen) en of geef aan of dat een indirecte of directe meting betreft. Meerdere betrokkenen zijn mogelijk.	Leerling; Docent; Peer (zoals medeleerling);specialist (zoals orthopedagoog, testpsycholoog);Onderzoeker; Software (algoritme)
31b. Beoordeling van directe prestatie	Duid aan wie betrokken is bij de BEOORDELING (niet het doorlopen als kandidaat) van de skill(onderdelen) en of geef aan of dat een indirecte of directe meting betreft. Meerdere betrokkenen zijn mogelijk.	Leerling; Docent; Peer (zoals medeleerling);specialist (zoals orthopedagoog, testpsycholoog);Onderzoeker; Software (algoritme)
32a1. Data-invoer registratie uitvoering	Hoe worden registratie van taakuitvoering, aanpak, proces en strategie vastgelegd?	1) Handmatig op papier (later omzetten en invoeren); 2) Handmatig, via elektronische invoermethode; 3) Geautomatiseerd (zonder tussenkomst beoordelaar)
32a2. Data-invoer scoring aanpak	Hoe worden scoring van de aspecten aanpak, proces en strategie vastgelegd?	1) Handmatig op papier (later omzetten en invoeren); 2) Handmatig, via elektronische invoermethode; 3) Geautomatiseerd (zonder tussenkomst beoordelaar)
32a3. Data-invoer registratie antwoord	Hoe wordt registratie van antwoord of oplossing vastgelegd?	1) Handmatig op papier (later omzetten en invoeren); 2) Handmatig, via elektronische invoermethode; 3) Geautomatiseerd (zonder tussenkomst beoordelaar)
32a4. Data-invoer scoring open antwoord	Hoe wordt scoring van open antwoord of oplossing vastgelegd?	1) Handmatig op papier (later omzetten en invoeren); 2) Handmatig, via elektronische invoermethode; 3) Geautomatiseerd (zonder tussenkomst beoordelaar)
32a5. Data-invoer scoring gesloten antwoord	Hoe wordt scoring van gesloten antwoord of oplossing vastgelegd?	1) Handmatig op papier (later omzetten en invoeren); 2) Handmatig, via elektronische invoermethode; 3) Geautomatiseerd (zonder tussenkomst beoordelaar)
32b. Berekening totaalscores	Hoe worden totaalscores op de skill(onderdelen) vastgelegd?	1) Handmatig invoeren; 2) Geautomatiseerd (zonder tussenkomst beoordelaar); 3) Overige (geef nadere toelichting)

Onderdeel	Toelichting	Antwoordmogelijkheden
33. Beoordelaarstraining	Beoordelaars zijn bewezen geïnstrueerd in het gebruik van het instrument, zoals het achterhalen en verifiëren van bewijs voor scoring op een beoordelingslijst en het nauwkeurig scoren van indicatoren op grond van bewijs	Apart beoordelen voor instrumenten met directe taken of instrumenten o.b.v. vragenlijsten (indirect): 1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
34. Beoordelingsmodel of scoremodel	Het instrument maakt gebruik van een onderbouwd beoordelings- of observatieschema voor de taken/opdrachten/bewijssituaties (hetgeen de kans op verschillende beoordelingen van beoordelaars verkleint).	Apart beoordelen voor instrumenten met directe taken of instrumenten o.b.v. vragenlijsten (indirect): 1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
35 Expliciete beoordelingsregels	Er zijn scoringsregels of beoordelingsregels beschikbaar, waarin expliciet wordt aangegeven hoe getoond gedrag in taken of taaksituaties moet worden gescoord op indicatoren voor de doelvaardigheid.	Apart beoordelen voor instrumenten met directe taken of instrumenten o.b.v. vragenlijsten (indirect): 1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
36. Beoordelaars-betrouwbaarheid	Er wordt een bevredigend niveau van interbeoordelaarsbetrouwbaarheid (>.90) resp. beoordelaarsovereenstemming (>.90) bereikt. Kappa>.80	Apart beoordelen voor instrumenten met directe taken of instrumenten o.b.v. vragenlijsten (indirect): 1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
37. Beoordelingsalgoritme	Bij gebruik van een scoringsalgoritme, is dit algoritme gebaseerd op beoordelingen van bewezen hoge kwaliteit geijkt op een breed spectrum van prestaties en een hoge beoordelaarsovereenstemming)	Apart beoordelen voor instrumenten met directe taken of instrumenten o.b.v. vragenlijsten (indirect): 1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
38. Onderbouwing samenstelling toetsscores	Er is een onderbouwde rationale achter de manier waarop scores op afzonderlijke onderdelen worden gecombineerd tot een totaalscore voor de doelvaardigheid	Apart beoordelen voor instrumenten met directe taken of instrumenten o.b.v. vragenlijsten (indirect): 1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
4. Bewijsmodel - gebruikte normen en standaarden		
41. Validiteitsbewijzen	Zijn er voor de validatie van de standaard-bepalingsprocedure validiteitsbewijzen verstrekt? Zoals verschillen tussen experts en beginners?	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
42. Representativiteit normgroep	Is (Zijn) de gebruikte normgroep(en) representatief?	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
43. Toereikende steekproefgrootte	Is de steekproefgrootte (per normgroep) toereikend?	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
44. Gegevens over scoreverdeling	Worden gemiddelden, standaardafwijkingen en gegevens over de scoreverdeling vermeld?	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
45. Betekenis normschaal	De betekenis en de beperkingen van de normschaal worden duidelijk gemaakt.	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
46. Normschaal en gebruiksdoel	Het type normschaal is in overeenstemming met het doel van het instrument.	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed

Onderdeel	Toelichting	Antwoordmogelijkheden
47. Gebruikte normen en standaarden	Is de standaardbepalingsmethode op een verantwoorde manier geselecteerd en is deze geschikt voor het te beoordelen instrument?	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
48. Correctheid standaardbepalingsmethode	Is de standaardbepalingsprocedure voldoende omschreven en is deze zorgvuldig uitgevoerd?	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
5. Bewijsmodel - betrouwbaarheid		
51. Gerichtheid op doelpopulatie	Zijn de betrouwbaarheidsgegevens berekend voor de doelpopulatie waarvoor het instrument gebruikt wordt?	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
52a. Betrouwbaar voor individueel belangrijke beslissing	Zijn de resultaten inzake betrouwbaarheid voldoende voor individueel belangrijke beslissing die met behulp van het instrument wordt genomen?	Apart beantwoorden voor hoofd- en subschalen, en voor directe en indirecte instrumenten. 1) onvoldoende (alpha lager dan .70); 2) matig (alpha .70-.79); 3) voldoende (alpha .80 - .89); 4) goed (alpha .90 of hoger)
52b. Betrouwbaar voor individueel minder belangrijke beslissing	Zijn de resultaten inzake betrouwbaarheid voldoende voor individueel minder belangrijke beslissing die met behulp van het instrument wordt genomen?	Apart beantwoorden voor hoofd- en subschalen, en voor directe en indirecte instrumenten. 1) onvoldoende (alpha lager dan .60); 2) Matig (alpha .60-.69); 3) Voldoende (alpha .70 - .79); 4) goed (alpha .80 of hoger)
52c. Betrouwbaar voor conclusies op stelselniveau	Zijn de resultaten inzake betrouwbaarheid voldoende voor conclusies op stelselniveau die met behulp van het instrument wordt genomen?	Apart beantwoorden voor hoofd- en subschalen, en voor directe en indirecte instrumenten. 1) onvoldoende (alpha lager dan .50); 2) matig (alpha .51-.59); 3) voldoende (alpha .60 - .69); 4) goed (alpha .70 of hoger)
6. Assemblagemodel		
61. Specificatie matrijs-domein naar kenmerken van te dekken taaksituaties	De kenmerken van taaksituaties waarin (onderdelen van) de vaardigheid moet worden getoond zijn voldoende gespecificeerd (zoals leesdoelen, tekstgenres, onderwerp bij geletterdheid).	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
62. Representativiteit samenstelling naar taaksituaties en omstandigheden	Scores op verschillende kwaliteiten van de leerling die ten grondslag liggen aan de doelvaardigheid zijn gebaseerd op bewijs uit die taken en observatieomstandigheden die redelijkerwijs aan de orde kunnen komen in situaties waarnaar wordt gegeneraliseerd	1) onvoldoende; 2 matig); 3) voldoende ; 4) goed
7. Presentatiemodel		

Onderdeel	Toelichting	Antwoordmogelijkheden
71. Locatie taakuitvoering	Waar vindt de taakuitvoering door de leerling ter meting van de doelvaardigheden plaats?	1) Niet gespecificeerd; 2) Fysieke situatie (mensen-apparaten) binnen school; 3) Fysieke situatie (mensen-apparaten) buiten school; 4) Computer-based (webapp); 5) Papier en pen
72. Locatie vastleggen assessmentdata	72. Locatie data-opslag: Waar vindt de data-verzameling plaats?	1) Niet gespecificeerd; 2) Op lokale computers van scholen of onderzoeksinstellingen; 3) Op een speciaal ingerichte vrij toegankelijke web-omgeving van derden (uitgevers, stichtingen); 4) Op een speciaal ingerichte afgeschermd web-omgeving van derden (uitgevers, stichtingen); 5) Anders, namelijk:
8. Handleiding		
80. Beschikbaarheid handleiding	Is er een handleiding beschikbaar bij het instrument?	1) Ja, maar alleen een technische handleiding (vooral gericht op bediening zoals bij een webapplicatie); 2) Ja, een inhoudelijke handleiding (inhoud, afname, scoring en interpretatie van resultaten van het instrument); 3) Ja, een gecombineerde technische en inhoudelijke handleiding; 4) Geen handleiding