

Redeneerfouten in onderzoeksvaardigheden bij bovenbouwleerlingen

Datum

15-10-2023

Student

A.J.C. Breur
s1083046

Begeleiding

Ard Lazonder
Sebastiaan de Klerk (Cito)

Aantal woorden

8840 woorden

Samenvatting

In dit scriptieonderzoek is onderzocht welke redeneerfouten in onderzoeksvaardigheden bovenbouwleerlingen in het basisonderwijs maken en in hoeverre het aantal redeneerfouten samenhangt met hun scores op begrijpend lezen en rekenvaardigheid. De onderzoeksvaardigheden die in dit scriptieonderzoek centraal stonden waren hypothesen opstellen, experimenten opzetten en uitvoeren, en bewijs evalueren. Bij 22 leerlingen uit groep 6, 7 en 8 van één school in Utrecht is individueel een meerkeuzetest voor deze onderzoeksvaardigheden afgenomen. Leerlingen beantwoordden de vragen hardop denkend, om zo hun eventuele redeneerfouten te kunnen achterhalen. De redeneringen werden gecodeerd volgens Kranz et al. (2022). Begrijpend lezen en rekenvaardigheid zijn gemeten met IEP-voortgangstoetsen. De leerlingen maakten in totaal 143 redeneerfouten, die onevenwichtig waren verdeeld over 11 categorieën: bij meer dan de helft van de redeneerfouten werd namelijk afgegaan op voorkennis/opvattingen. Daarnaast bleek een juist gekozen antwoordoptie niet garant te staan voor een juiste bijpassende redenering. Het totale aantal redeneerfouten hing niet samen met de vaardigheden in begrijpend lezen en rekenen, en verschillen tussen leerlingen wat betreft het aantal redeneerfouten konden niet worden verklaard door de gezamenlijke vaardigheden in begrijpend lezen en rekenen. Desondanks suggereert het hoge aantal redeneerfouten het belang van goed W&T-onderwijs. De resultaten suggereren dat leerkrachten goed op de hoogte moeten zijn van mogelijke redeneerfouten in onderzoeksvaardigheden. Zij moeten niet alleen bij foute, maar juist ook bij goede antwoorden van leerlingen de achterliggende redenering controleren. Toetsontwikkelaars kunnen hier inhoudelijk rekening mee houden, waardoor leerkrachten die formatieve informatie kunnen gebruiken om hun W&T-onderwijs, en dus de onderzoeksvaardigheden van leerlingen, te verbeteren.

Sleutelwoorden: onderzoeksvaardigheden, redeneerfouten, formatieve beoordeling, leerlingkenmerken, basisonderwijs

Inleiding

De afgelopen jaren is er in Nederland steeds meer aandacht voor Wetenschap en Technologie (W&T) in het basisonderwijs (Djoyoadhiningrat-Hol & Klein Tank, 2022). Zo zijn basisscholen sinds 2020 verplicht om W&T-onderwijs op te nemen in het curriculum (Djoyoadhiningrat-Hol & Klein Tank, 2022; Rijksoverheid, 2013). Veel W&T-leslessen zijn geschikt voor onderzoekend leren, een instructiemethode waarbij leerlingen onder begeleiding van de leerkracht zelf een onderzoek plannen en uitvoeren (Pedaste et al., 2015). Hierdoor kunnen leerlingen iets leren over het onderwerp dat zij onderzoeken én hun onderzoeksvaardigheden verbeteren (Lazonder & Harmsen, 2016; Pedaste et al., 2021). Voorbeelden van deze onderzoeksvaardigheden zijn hypothesen opstellen, experimenten uitvoeren, resultaten interpreteren en conclusies trekken (Pedaste et al., 2015; Zimmerman, 2007). De ontwikkeling van onderzoeksvaardigheden sluit aan bij kerndoel 42 (Djoyoadhiningrat-Hol & Klein Tank, 2022) en is daarnaast van belang om leerlingen als autonome, kritische denker te kunnen laten deelnemen aan de kennismaatschappij (Fischer et al., 2014), waarin onderzoeksvaardigheden ten grondslag liggen aan alledaagse denkprocessen als causaal redeneren en probleemoplossen (Dunbar & Klahr, 2012).

Hoewel onderzoek heeft aangetoond dat onderzoeksvaardigheden zich kunnen ontwikkelen zonder uitleg of hulp door een leerkracht, gebeurt dit alleen als leerlingen gedurende langere tijd intensief met onderzoek bezig zijn (Dean & Kuhn, 2007). Dit is echter in het Nederlandse onderwijs vaak onmogelijk, gezien de vaak geringe tijd die aan W&T wordt besteed: wekelijks gemiddeld drie kwartier in de onderbouw tot een uur in de bovenbouw (Mommers et al., 2016). Om onderzoekend leren - en dus het verwerven van onderzoeksvaardigheden - ook effectief te laten zijn, is begeleiding door de leerkracht hierbij dus belangrijk (Aditomo & Klieme, 2020; D'Angelo et al., 2014; Furtak et al., 2012; Hmelo-Silver et al., 2007). Onderzoekend leren onder (passende) begeleiding zorgt voor effectievere leerresultaten van leerlingen vergeleken met leerlingen die geen of minimale begeleiding krijgen (Kirschner et al., 2006; Klahr & Nigam, 2004) of wanneer leerlingen alleen observeren wat de leerkracht voordoet (Alfieri et al., 2011).

Om de begeleiding bij onderzoekend leren te laten aansluiten bij het niveau en de wensen van leerlingen, is formatieve beoordeling nodig. Hierbij gebruiken leerkrachten of (mede)leerlingen informatie over het leren van leerlingen om te beslissen over volgende stappen in het onderwijsleerproces (Black & Wiliam, 2009). Het geeft hen inzicht in de leervoortgang, om zo het leren en de instructie te verbeteren (Boston, 2002; Winger, 2005). Naast bijvoorbeeld het observeren van leerlingen en het houden van klassendiscussies, is ook

het analyseren van leerlingwerk zoals toetsen een vorm van formatieve beoordeling (Boston, 2002).

Voor formatieve beoordeling van één of meer onderzoeksvaardigheden zijn verschillende betrouwbare en valide meetinstrumenten ontwikkeld voor het basisonderwijs, zoals performance assessments (zie bijvoorbeeld Kruit et al., 2018; Lazonder & Janssen, 2021) en testen op papier (zie bijvoorbeeld Koerber & Osterhaus, 2019; Kruit et al., 2018; Mayer et al., 2014; Osterhaus et al., 2020). In onderzoeken met deze meetinstrumenten is echter voornamelijk gekeken naar hoe goed leerlingen bepaalde onderzoeksvaardigheden beheersen (Kruit et al., 2018; Mayer et al., 2014; Schlatter et al., 2021) en blijven de redeneerfouten die leerlingen maken onderbelicht. In recent onderzoek is wel naar deze fouten gekeken (Baur, 2021, 2023), maar ook hier blijft onduidelijk waarom leerlingen deze fouten maken. Juist die informatie is belangrijk voor formatieve beoordeling, omdat het leerkrachten aanknopingspunten geeft voor volgende lesmomenten waarin onderzoeksvaardigheden centraal staan. In dit scriptieonderzoek wordt daarom geprobeerd te achterhalen welke redeneerfouten bovenbouwleerlingen maken bij het uitvoeren van onderzoeksvaardigheden en in hoeverre het aantal gemaakte redeneerfouten samenhangt met leerlingkenmerken.

Onderzoeksvaardigheden

In de literatuur worden verschillende fases onderscheiden in het onderzoeksproces. Het aantal fases en hun definities verschillen echter. Zo worden door Harlen en Jelly (1990), Wu en Hsieh (2006) en Van Graft et al. (2014) zeven fases voorgesteld, die inhoudelijk gezien niet volledig overeenkomen. Om de overeenkomsten en verschillen tussen dit soort beschrijvingen te analyseren, hebben Pedaste et al. (2015) een systematisch literatuuronderzoek uitgevoerd. Zij concludeerden dat vijf algemene onderzoeksfases in de meeste modellen van onderzoekend leren voorkomen: Oriëntatie, Conceptualisatie, Onderzoek, Conclusie en Discussie. In het Scientific Discovery as Dual Search (SDDS) model (Klahr & Dunbar, 1988) worden de fases conceptualisatie, onderzoek en conclusie verder uitgewerkt aan de hand van de respectievelijke vaardigheden hypotheses opstellen, experimenten bedenken en uitvoeren, en de resultaten daarvan evalueren. Omdat het SDDS-model beter aansluit bij wat er in een enkele W&T-les gebeurt dan de fases van Pedaste et al. (2015), worden de bijbehorende drie vaardigheden in dit onderzoek als uitgangspunt genomen.

Hypotheses opstellen

Deze vaardigheid vindt vaak aan het begin van het onderzoeksproces plaats en omvat het bedenken van ideeën over mogelijke uitkomsten van een onderzoek of oplossingen van

een probleem of vraag (Van Joolingen & De Jong, 1991; Wenham, 1993). Wenham (1993) maakt onderscheid tussen beschrijvende, verklarende en voorspellende hypothesen. Het proces van hypothesevorming beslaat drie subprocessen: het identificeren en selecteren van variabelen en het bepalen van een relatie die verondersteld wordt te bestaan tussen deze variabelen (Van Joolingen & De Jong, 1991). Een hypothese is correct geformuleerd als deze kan worden getest door middel van experimenten of observaties (Pedaste et al., 2012).

Onderzoek opzetten en uitvoeren

Experimenteren betreft de vaardigheid om systematische vergelijkingen te ontwerpen en uit te voeren die leiden tot interpreteerbare resultaten, om daarna hypothesen te testen of te onderbouwen (Chen & Klahr, 1999; Pedaste et al., 2015; Piekny & Maehler, 2013; Zimmerman, 2000). Een belangrijke stap in het experimenteren is dan ook het bedenken van een onderzoeksopzet die, ongeacht de uitkomst, valide data oplevert (Crocker & Buchanan, 2011). Dit kan door de Control-of-Variables Strategy (CVS) te gebruiken (Chen & Klahr, 1999; Tschirgi, 1980). Dit is een domein-algemene strategie voor het ontwerpen of herkennen van gecontroleerde experimenten met meerdere variabelen. Hierbij wordt de onafhankelijke variabele gemanipuleerd terwijl alle andere variabelen constant worden gehouden. Zo kan een valide uitspraak worden gedaan over de afhankelijke variabele. Wordt de CVS niet gebruikt, dan is er sprake van een verstoord experiment met een derde, verstorende variabele waarvoor niet is gecontroleerd (Gott & Duggan, 2003).

Bewijs evalueren

Bewijs evalueren kan worden onderverdeeld in achtereenvolgens het interpreteren en evalueren van data en het trekken van conclusies (Schlatter et al., 2021). Het interpreteren van data draait om het herkennen van patronen of regelmatigigheden in één of meerdere reeksen numerieke of visuele representaties van experimentele uitkomsten, zoals tabellen en grafieken (Kanari & Millar, 2004). Ook gaat het vaak om het interpreteren van covariatiegegevens in relatie tot concurrerende hypothesen (Piekny & Maehler, 2013). Het uiteindelijke resultaat van deze vaardigheid is dus een interpretatie van de gegevens die het mogelijk maakt om terug te keren naar de oorspronkelijke onderzoeksvraag of hypothese en hierover een conclusie te trekken (Pedaste et al., 2015).

Bij de evaluatie van data zijn meerdere aspecten betrokken. Bij het aspect relevantie vraag je je af of je het juiste bewijs hebt om een bewering te beoordelen (Sandoval et al., 2014). Bij toereikendheid vraag je je daarnaast af of je voldoende bewijs hebt om de bewering te beoordelen (Masnick & Morris, 2008; Sandoval et al., 2014). Verder moet rekening gehouden worden met de variabiliteit tussen en binnen de datasets (Masnick & Morris, 2008).

Ook is het belangrijk om beslissingen te nemen op basis van gegevens en niet op basis van voorkennis of theorieën over het fenomeen in kwestie. Tot slot is het noodzakelijk om het bewijs te onderscheiden van de hypothese die het ondersteunt (Piekny & Maehler, 2013).

Als laatste het concluderen, het proces waarin conclusies worden getrokken uit de gegevens. Het bewijs uit de experimenten moet correct worden geëvalueerd om te bepalen of het voldoende is om de hypothesen te verwerpen of te accepteren (Klahr, 2000; Klahr & Dunbar, 1988). Er zijn verschillende soorten bewijs. Sluitend bewijs is bewijs dat onweerlegbaar wijst op één enkele conclusie. Gedeeltelijk bewijs geeft een suggestie voor de oorzaak van het effect, maar is niet doorslaggevend. Bewijs dat niet eenduidig is, is bewijs waaruit geen conclusie kan worden getrokken, aangezien de ene helft van het bewijs de andere helft tegenspreekt (Van der Graaf et al., 2016). In de laatste twee gevallen moet terug worden gegaan naar de fase van experimenteren (Klahr & Dunbar, 1988).

Ontwikkeling van en fouten in onderzoeksvaardigheden

Bovenstaande onderzoeksvaardigheden ontwikkelen zich in de basisschoolleeftijd. Dit gebeurt echter niet op hetzelfde moment en in hetzelfde tempo, waardoor regelmatig fouten worden gemaakt. Kranz et al. (2022) kwamen in hun recente review - waarin zij eerdere reviews meenamen - tot 43 uitdagingen in de vorm van fouten en specifieke benaderingen van leerlingen die zich kunnen voordoen tijdens het uitvoeren van onderzoeksvaardigheden. Hieronder worden per onderzoeksvaardigheid de ontwikkeling en enkele fouten beschreven.

Hypotheses opstellen

De vaardigheid om hypothesen te formuleren of voorspellingen te doen, ontwikkelt zich tijdens de basisschoolleeftijd (Piekny & Maehler, 2013). Hoewel de prestaties toenemen rond 7 jaar, worden pas significant betere hypothesen opgesteld rond 11 jaar (Piekny & Maehler, 2013). Leerlingen hebben dan echter nog steeds de neiging om plausibele hypothesen te bedenken en proberen de juistheid hiervan aan te tonen. Ze vermijden hypothesen die een grotere kans hebben om verworpen te worden en blijven zich vasthouden aan één hypothese. Oudere leerlingen daarentegen overwegen meerdere hypothesen en bedenken vaker ongeloofwaardige hypothesen (Klahr et al., 1993). Tot slot gaan leerlingen bij het opstellen van hypothesen vaak alleen uit van een positieve covariantie tussen de afhankelijke en onafhankelijke variabelen (Kanari & Millar, 2004).

Onderzoek opzetten en uitvoeren

De meeste kleuters kunnen al enigszins systematisch testen (Köksal-Tuncer & Sodian, 2018; Van der Graaf et al., 2015) en begrijpen eenvoudige experimenten (Piekny & Maehler, 2013). De meeste leerlingen uit groep 3 (5-8 jaar) kiezen al een geschikte test om een

hypothese te testen en maakt onderscheid tussen het testen van hypothesen en het produceren van effecten (Köksal-Tuncer & Sodian, 2018; Piekny et al., 2014; Sodian et al., 1991). Vanaf 8 jaar verkiezen leerlingen een gecontroleerd experiment boven een verstoord experiment en verantwoorden ze dit ook op de juiste manier. Ze passen dan echter nog niet uit zichzelf de CVS toe (Bullock & Ziegler, 1999).

Vanaf 10 jaar kunnen de meeste leerlingen een gecontroleerd onderzoek uitvoeren naar de relatie tussen duidelijk covariërende variabelen (Kanari & Millar, 2004; Schauble et al., 1995). Volgens onderzoek van Kuhn en Dean (2005) kunnen 11-12-jarigen zonder hulp enigszins gecontroleerde experimenten ontwerpen; met hulp om te focussen op één variabele verbetert dit echter significant (Schauble et al., 1995). Richting het eind van de basisschool gaat het opzetten van een gecontroleerd experiment dan ook beter (Bullock & Ziegler, 1999).

Desondanks worden ook bij deze vaardigheid fouten gemaakt. Leerlingen voeren bijvoorbeeld maar één experimenteel onderzoek of enkel een controle-onderzoek uit en kunnen dus de invloed van de onafhankelijke variabele(n) niet vergelijken (Baur, 2021, 2023; Hammann et al., 2008; Kanari & Millar, 2004). Ook komt het voor dat ze hetzelfde experiment meerdere keren uitvoeren (Baur, 2021, 2023; Kuhn et al., 1992) en controleren en manipuleren ze de variabelen vaak onjuist (Chen & Klahr, 1999; Hammann et al., 2008; Kanari & Millar, 2004; Kuhn et al., 1992). Verder zoeken sommige leerlingen naar bewijs dat hun huidige hypothese ondersteunt. Dit kan hen ervan weerhouden een alternatieve hypothese te formuleren, zelfs als het bewijs inconsistent is (Dunbar, 1993).

Bewijs evalueren

Het vermogen om bewijsmateriaal te evalueren is al goed ontwikkeld op vierjarige leeftijd en neemt in de loop van de tijd toe bij een perfect patroon van covariantie of non-covariantie (Piekny & Maehler, 2013; Piekny et al., 2014; Van der Graaf et al., 2018). Leerlingen van alle leeftijden hebben echter meer moeite met het trekken van conclusies als deze gebaseerd zijn op non-covariantie (Kanari & Millar, 2004; Koerber et al., 2005) en imperfect bewijs (Piekny et al., 2014). Deze laatste vaardigheid is pas volgroeid tegen het einde van de basisschool (Piekny & Maehler, 2013). Verder blijkt dat leerlingen vaak conclusies trekken op basis van één of een invalide experiment (Hammann et al., 2008; Kuhn et al., 1992) en verwarren ze het resultaat van een experiment met een conclusie (Boaventura et al., 2013). Daarnaast hebben leerlingen van 8-12 jaar de neiging om gegevens te veranderen of zelfs te negeren, wanneer bewijs (van covariantie) niet overeenkomt met hun eerdere overtuigingen of kennis (Amsel & Brock, 1996; Chinn & Brewer, 1993; Dunbar & Klahr, 1989; Kuhn et al., 1988). Ook maken ze nauwelijks gebruik van oordelen gebaseerd op

bewijs, maar gaan ze in plaats daarvan af op hun voorkennis (Amsel & Brock, 1996; Masnick & Morris, 2008).

Tienjarigen kunnen visuele en numerieke data interpreteren (Masnick et al., 2017) en dit bewijs gebruiken als argument om een hypothese te evalueren (Krummenauer & Kuntze, 2019). Leerlingen kiezen bij herhaalde metingen echter vaak de eerste, laatste of tussenliggende waarde in plaats van het gemiddelde te berekenen (Masnick & Klahr, 2003). Negenjarigen hebben al een intuïtief bewustzijn van het aantal en de variabiliteit van observaties als ze gegevens analyseren, maar ze vinden het lastig om conclusies te trekken op basis van meerdere waarnemingen. Ook letten zij nauwelijks op de variantie binnen en tussen deze metingen. Leerlingen van 11-12 jaar houden daarentegen iets beter rekening met de steekproefomvang, maar ook zij besteden weinig aandacht aan variantie tussen en binnen groepen (Masnick & Morris, 2008).

Onderzoeksvaardigheden en leerlingkenmerken

In eerdere onderzoeken naar onderzoeksvaardigheden bleken substantiële verschillen in prestaties tussen leerlingen van dezelfde leeftijd (Pieknny & Maehler, 2013; Schlatter et al., 2021). Deze variatie bleek gedeeltelijk te verklaren door (cognitieve) leerlingkenmerken, of er bleek samenhang te zijn tussen leerlingkenmerken en de prestaties van leerlingen. Hieronder worden enkele leerlingkenmerken uiteengezet.

Ten eerste basisvaardigheden in taal. Uit onderzoek van Schlatter et al. (2021) bleek dat de variantie in gemiddelde scores op een onderzoeksvaardighedentest tussen 11-jarige leerlingen gedeeltelijk werd verklaard door hun begrijpend leesvaardigheden. Begrijpend lezen had daarnaast een significante correlatie met experimenteren, data evalueren en conclusies trekken, maar niet met hypothesen opstellen. Uit onderzoek van Van de Sande et al. (2019) bleek daarentegen dat begrijpend lezen significant correleerde met experimenteren en conclusies trekken, en variantie verklaarde in alle onderzoeksvaardigheden. Specifieker gekeken bleek uit onderzoek van Siler et al. (2010) dat begrijpend lezen een verklarend effect had bij het opzetten van gecontroleerde experimenten met behulp van de CVS. Tot slot bleek uit meerdere onderzoeken dat begrijpend lezen significant correleerde met de gemiddelde scores op een onderzoeksvaardighedentest bij 7-10-jarigen en/of een significant deel van de variantie hierin verklaarde (Koerber et al., 2015; Lazonder & Janssen, 2021; Mayer et al., 2014; Schiefer et al., 2019).

Voor rekenvaardigheid zijn eerdere onderzoeksresultaten minder eenduidig. Zo bleek uit onderzoek van Mayer et al. (2014) dat ruimtelijk inzicht significant correleerde met gemiddelde scores op een onderzoeksvaardighedentest. Uit onderzoek van Lazonder en

Janssen (2021) bij 7-10-jarigen bleek dat rekenvaardigheid significant correleerde met en variantie verklaarde in gemiddelde scores op een onderzoeksvaardighedentest. Ook bij jongere leerlingen bleek dit het geval (Koerber & Osterhaus, 2019). Schlatter et al. (2021) vonden daarentegen juist geen verklarend vermogen van rekenvaardigheid op gemiddelde scores op een onderzoeksvaardighedentest, enkel een significante samenhang met het evalueren van bewijs.

Ook intelligentie is een cognitieve factor die regelmatig is meegenomen in onderzoek. In onderzoeken van Koerber et al. (2015), Mayer et al. (2014) en Schiefer et al. (2019) bij leerlingen van 8-10 jaar bleek intelligentie significant samen te hangen met en/of invloed te hebben op gemiddelde scores op een onderzoeksvaardighedentest. Ook bij jongere leerlingen bleek dit het geval (Koerber & Osterhaus, 2019).

Tot slot zijn er nog enkele kenmerken die minder vaak onderzocht zijn. Wat betreft probleemoplossingsvaardigheden bleek uit onderzoek van Mayer et al. (2014) bij tienjarigen dat deze significant correleerden met gemiddelde scores op een onderzoeksvaardighedentest, terwijl Schlatter et al. (2021) juist geen verklarend effect en significante correlatie vonden. Van de Sande et al. (2019) vonden daarentegen dat probleemoplossingsvaardigheden variantie verklaarden in hypothesen opstellen en experimenteren, maar niet in conclusies trekken. Verder vonden Koerber et al. (2015) bij leerlingen van 8-10 jaar ook een significante invloed van het opleidingsniveau van de ouders op gemiddelde scores op een onderzoeksvaardighedentest.

Samengevat blijken verschillende leerlingkenmerken van invloed te zijn op de verschillen in prestaties tussen leerlingen, waarbij het meeste bewijs is voor een positieve samenhang tussen onderzoeksvaardigheden en taal- en rekenvaardigheden. Deze twee kenmerken worden dan ook meegenomen in huidig scriptieonderzoek.

Huidig onderzoek

Eerder onderzoek naar onderzoeksvaardigheden heeft zich voornamelijk gericht op de leerlingprestaties en de samenhang daarvan met leerlingkenmerken als basisvaardigheden in taal en rekenen (Mayer et al., 2014; Schlatter et al., 2021; Van de Sande et al., 2019). Hoewel fouten wel werden benoemd in onderzoek naar onderzoeksvaardigheden en enkele reviewstudies naar fouten zijn gedaan (De Jong & Van Joolingen, 1998; Kranz et al., 2022; Zimmerman, 2007), is er weinig onderzoek dat zich specifiek richtte op het identificeren en verklaren van redeneerfouten in onderzoeksvaardigheden (maar zie Baur 2021, 2023). Bovendien is nog niet eerder onderzocht wat de samenhang is tussen redeneerfouten bij onderzoeksvaardigheden en leerlingkenmerken. Dit is opvallend, aangezien dit juist

waardevolle, formatieve informatie kan opleveren voor leerkrachten en de inrichting van hun W&T-onderwijs over onderzoeksvaardigheden.

De onderzoeksvraag die in dit onderzoek centraal staat is dan ook als volgt: Welke redeneerfouten maken leerlingen uit groep 6, 7 en 8 op een geschreven test van onderzoeksvaardigheden en in hoeverre hangt het aantal gemaakte redeneerfouten samen met de leerlingkenmerken rekenvaardigheid en begrijpend lezen?

Verwacht wordt dat een lagere score op begrijpend lezen samenhangt met meer redeneerfouten op alle onderzoeksvaardigheden afzonderlijk (Van de Sande et al., 2019) en dus ook met de totaalscore op een onderzoeksvaardigheidentest (Koerber et al., 2015; Lazonder & Janssen, 2021; Mayer et al., 2014; Schiefer et al., 2019). Een lagere score voor rekenvaardigheid wordt verwacht samen te hangen met meer redeneerfouten in totaal en met fouten op de deelvaardigheid ‘bewijs evalueren’ (Lazonder & Janssen, 2021; Mayer et al., 2014; Schlatter et al., 2021).

Methode

Participanten

In totaal hebben 23 leerlingen (10 meisjes en 13 jongens) uit groep 6, 7 en 8 van één school uit de provincie Utrecht deelgenomen aan dit scriptieonderzoek. Eén leerling is uit de steekproef verwijderd, omdat deze de Nederlandse taal onvoldoende beheerste. De gemiddelde leeftijd van de overgebleven 22 leerlingen was 11 jaar en 2 maanden ($SD = 11$ maanden). Zie Tabel 1 voor een overzicht van de kenmerken van de deelnemers. Het onderzoek vond plaats tussen 22 en 31 mei 2023 nadat de ouders van de deelnemende leerlingen toestemming hadden gegeven door een informed consent formulier te ondertekenen. Het onderzoeksvoorstel is getoetst door de Ethiek Commissie Sociale Wetenschappen (ECSW-LT-2023-5-14-60414) en er was geen bezwaar tegen dit onderzoek.

Tabel 1

Kenmerken Deelnemers Uitgesplitst per Groep

	Groep 6 (n = 6)	Groep 7 (n = 8)	Groep 8 (n = 8)	Totaal (n = 22)
Leeftijd	<i>M</i> = 10 jaar, 3 maanden <i>SD</i> = 5 maanden	<i>M</i> = 11 jaar, 2 maanden <i>SD</i> = 6 maanden	<i>M</i> = 12 jaar, 0 maanden <i>SD</i> = 8 maanden	<i>M</i> = 11 jaar, 2 maanden <i>SD</i> = 11 maanden
Geslacht (%)				
Jongens	4 (67%)	3 (38%)	6 (75%)	14 (64%)
Meisjes	2 (33%)	5 (62%)	2 (25%)	8 (36%)

Instrumenten

Onderzoeksvaardigheden

Voor het meten van de onderzoeksvaardigheden hypothesen opstellen, experimenten opstellen en uitvoeren, en bewijs evalueren zijn vragen uit bestaande meerkeuzetests voor onderzoeksvaardigheden gebruikt (Kruit et al., 2018; Osterhaus et al., 2020). Tabel 2 geeft een overzicht van de gebruikte vragen per onderzoeksvaardigheid. Elke vraag bestond uit een korte, verhalende introductie met vervolgens de vraag en drie, vier of vijf antwoordopties. Er was telkens maar één antwoordoptie juist. Twee vragen zijn inhoudelijk aangepast om de eventuele fouten beter naar voren te laten komen. Enkele andere vragen zijn woordelijk aangepast om de vraag duidelijker te maken. De vragen uit het onderzoek van Osterhaus et al. (2020) zijn naar het Nederlands vertaald. In Bijlage A is de uiteindelijk samengestelde test met 11 vragen te vinden. In Bijlage B is het antwoordmodel opgenomen.

Tabel 2

Overzicht Vragen

Onderzoeks-vaardigheid	Vraagnummer	Deelvaardigheid	Afkomstig uit	Mogelijke fouten
Hypothesen opstellen	1 (kamerplant)	Hypothese opstellen	Kruit et al. (2018) (aangepast)	H0, M2
	6 (bonenzaden)	Hypothese opstellen	Kruit et al. (2018) (aangepast)	H0, H1, M2
Experimenten opzetten en uitvoeren	3 (bloeddruk)	Correct observeren/meten	Kruit et al. (2018)	E10
	4 (touwtjes)	CVS	Kruit et al. (2018)	E3, E5, M4
	9 (aardvarken)	Onderscheid maken tussen produceren van effect en testen van hypothese (falsificatie)	Osterhaus et al. (2020)	E7
	10 (school)	Experimenteel ontwerp begrijpen (error herkennen)	Osterhaus et al. (2020)	E2
	11 (waterbak)	CVS	Osterhaus et al. (2020)	E3, M4
Bewijs evalueren	2 (mieren)	Conclusie trekken	Kruit et al. (2018)	B1, B10
	5 (kaars)	Gevolgtrekkingen maken op basis van bewijs	Kruit et al. (2018)	B5, B8, B10, B13, M2
	7 (Simanezen)	Data-interpretatie: Omgaan met non-covariantie	Osterhaus et al. (2020)	B6, B13
	8 (kruiden)	Data-interpretatie: Herkennen van verstoorde aard van de data	Osterhaus et al. (2020)	B12

Foutencodering

Voor de foutencodering is gebruikgemaakt van de reviewstudie van Kranz et al. (2022). Zij kwamen tot 43 uitdagingen (fouten) die kunnen voorkomen tijdens het uitvoeren van de onderzoeksvaardigheden ‘onderzoeksvragen stellen’, ‘hypothesen genereren’, ‘een experiment plannen en uitvoeren’, ‘gegevens analyseren’ en ‘conclusies trekken’.

Onderzoeksvragen stellen is in deze scriptie buiten beschouwing gelaten, waardoor nog 41

fouten overbleven. Tien hiervan zijn door Kranz et al. (2022) geclassificeerd als specifieke benaderingen die het onderzoeksproces kunnen bemoeilijken of de ontwikkeling van onderzoeksvaardigheden belemmeren. Deze 10 benaderingen zijn niet meegenomen in de foutencodering, waardoor 31 soorten fouten overbleven. Vervolgens is gekeken welke van deze fouten zouden kunnen voorkomen in de vragen van Kruit et al. (2018) en Osterhaus et al. (2020). Hierdoor bleven uiteindelijk 15 fouten over, verdeeld over 10 vragen. Om het aantal fouten bij hypothesen opstellen uit te breiden, is er buiten de fouten uit Kranz et al. (2022) nog een eventueel voor te komen fout meegenomen uit Penner en Klahr (1996). Hiermee komt het aantal vragen op 11 en het aantal mogelijk te maken fouten meegenomen in dit scriptieonderzoek op 16 (zie Tabel 3).

Tabel 3

Foutcodes en Omschrijvingen uit Kranz et al. (2022)

Foutcode	Foutomschrijving
H0 ^a	Voorspelling doen over de uitkomst van een experiment in plaats van expliciete hypothese stellen.
H1	Niet weten wat een hypothese is.
E2	Het experiment beperken tot slechts een paar onafhankelijke variabelen.
E3	Variabelen niet controleren.
E5	Werken zonder controleconditie.
E7	Technische benadering; Leerlingen creëren een effect als uitkomst ('engineering approach') of voeren gewoon leuke experimenten uit.
E10	Gebrek aan herhaling van metingen.
B1	Verwarring tussen observatie en interpretatie van gegevens.
B5	Moeilijkheden bij het lezen van grafieken.
B6	Conclusies trekken wanneer het resultaat geen covariatie laat zien, is moeilijk.
B8	Het negeren van afwijkende gegevens.
B10	Verkeerde conclusies trekken.
B12	Leerlingen trekken geen uitsluitingsconclusies (exclusion inferences).
B13	Leerlingen geven minder geavanceerde strategieën niet op bij het interpreteren van resultaten.
M2	Variabelen niet kunnen identificeren.
M4	Uitproberen; Leerlingen volgen geen systematische aanpak bij het experimenteren en gaan daarom niet op een wetenschappelijke manier te werk.

Noot. H = hypothese, E = experimenten opzetten en uitvoeren, B = bewijs evalueren, M = valt onder meerdere onderzoeksfasen.

^a Afkomstig uit Penner en Klahr (1996).

Begrijpend lezen

Begrijpend lezen is gemeten door middel van de IEP LVS Lezen, ontwikkeld door Bureau ICE. Er waren verschillende versies voor verschillende jaargroepen, allemaal van de

IEP 2022-I. Elke versie bestond uit 2 sets van 20 opgaven en werd digitaal afgenomen, waarbij de leerling een tekstboekje met de corresponderende teksten uit de toets ook geprint op tafel had liggen. De toets bestond uit verschillende tekstsoorten met bijbehorende vragen. Dit waren bijvoorbeeld zakelijke teksten zoals een handleiding of nieuwsbericht, maar ook fictionele, literaire en narratieve teksten (bijvoorbeeld een blog of een gedicht). De tekstlengte en moeilijkheidsgraad waren niveau-afhankelijk. De vragen varieerden tussen meerkeuzevragen (1 antwoord goed) en multikeuzevragen (2 antwoorden goed). Er moest bijvoorbeeld de juiste woordbetekenis worden gekozen of een juiste afleiding uit de tekst worden gemaakt. Dyslectische leerlingen met een leesachterstand van 2 jaar kregen deze toets voorgelezen. De toets werd automatisch nagekeken, waarbij elk volledig correct antwoord een ruwe score van 1 werd toegekend en elk (deels) fout antwoord een ruwe score van 0. Vervolgens is de totale ruwe score automatisch omgezet in een ontwikkelscore. Deze toetsen zijn afgenomen op het voor de school standaardmoment, in november 2022 (groep 8) en januari 2023 (groepen 6 en 7). De gemiddelde ontwikkelscores voor deze leerlingpopulatie zijn voor de groepen 6, 7 en 8 respectievelijk 61, 71 en 79.

Rekenvaardigheid

Rekenvaardigheid is gemeten door middel van de IEP LVS Rekenen, ontwikkeld door Bureau ICE. Er waren verschillende versies voor verschillende jaargroepen, allemaal van de IEP 2022-I. Elke versie bestond uit 2 sets van 20 opgaven en werd digitaal afgenomen. De domeinen Getallen, Meten en meetkunde, Verhoudingen en Verbanden kwamen aan bod, met daarin zowel kale sommen als contextsommen en zowel open als gesloten vragen. Elk kind had kladpapier om denkstappen te kunnen opschrijven. Bij alle rekentoetsen was audio-ondersteuning beschikbaar voor alle leerlingen. De afname en scoring was identiek aan de IEP LVS Lezen. De gemiddelde ontwikkelscores voor de groepen 6, 7 en 8 zijn respectievelijk 53, 63 en 71.

Procedure

Voorafgaand aan het onderzoek is in elke deelnemende groep een korte introductie gegeven en zijn de toestemmingsformulieren voor de ouders meegegeven. Vervolgens zijn de deelnemende leerlingen, op een met de betreffende leerkracht overlegd moment, één voor één meegenomen voor het maken van de test. Deze is door de onderzoeker individueel afgenomen in een stille, afgesloten ruimte in de school tijdens reguliere schooluren. De afname duurde 25 tot 50 minuten. Voorafgaand aan de test werd een mondelinge instructie gegeven over de opbouw en de afnamemethode volgens een instructietekst (zie Bijlage C). Er was één voorbeeldvraag waarin het hardop denken werd voorgedaan en vervolgens een oefenvraag.

Ook dit waren vragen uit een eerdere test (Kruit et al., 2018), maar bij een onderzoeksvaardigheid die in huidig onderzoek niet is meegenomen (onderzoeksvraag opstellen). Tijdens deze oefenvraag was enige hulp en feedback toegestaan, maar tijdens de eigenlijke taken werd geen feedback gegeven over de juistheid van de antwoorden. Vervolgens beantwoordde de leerling de 11 vragen hardop denkend. Hiervoor is gekozen omdat het gebruik van hardopdenkprotocollen gedetailleerde informatie geeft over de cognitieve processen die worden doorlopen tijdens het oplossen van een probleem (Van Someren et al., 1994). De onderzoeker noteerde ondertussen per vraag de redenering van de leerling bij de volgens hen juiste antwoordoptie. Waar mogelijk werd hier direct een foutcode toegevoegd. Ook wanneer de antwoordoptie juist was, maar de redenering fout, werd deze toegevoegd aan het codeerschema. Soms werd de leerling gevraagd even te stoppen, zodat de onderzoeker iets uitgebreider kon noteren. Aan de hand van de instructietekst is consistent gereageerd op bijvoorbeeld het stilvallen van de leerling (zie Bijlage C).

Analyse

Na afloop van de dataverzameling zijn de foutieve rederingen van de leerlingen op een deductieve manier gecodeerd. Hierbij is aanvankelijk gekeken naar de foutcodes uit Tabel 3. Onvoorziene fouten waarvoor geen geschikte code bestond, zijn aan het codeerschema toegevoegd. Om de betrouwbaarheid van de codering te waarborgen, heeft een tweede beoordelaar 44 fouten van 4 willekeurige leerlingen gecodeerd. Cohen's Kappa is berekend met deze data, waaruit een interbeoordelaarsbetrouwbaarheid bleek van $\kappa = .870$. Dit is een hoog niveau van overeenkomst tussen de beoordelaars (Landis & Koch, 1977).

Om te beantwoorden in hoeverre het aantal gemaakte fouten op de test voor onderzoeksvaardigheden samenhang met de scores voor rekenvaardigheid en begrijpend lezen, is een correlatie en multiële regressie uitgevoerd. Bij deze laatste analyse zijn rekenvaardigheid en begrijpend lezen in één model opgenomen. Er is niet gecontroleerd voor leeftijd, aangezien de IEP-scores op een schaal worden uitgedrukt en door de leerjaren heen lopen. Voorafgaand aan de multiële regressie zijn de assumpties gecontroleerd (Allen et al., 2014). De gegevens zijn geanalyseerd met behulp van IBM SPSS 28.

Resultaten

Onderzoeksvaardigheden

Ten eerste is gekeken naar het aantal goed beantwoorde vragen, ongeacht de redenering. Tabel 4 laat zien dat groep 7 gemiddeld het hoogst scoorde en groep 6 het laagst. De scores voor de groepen 6, 7 en 8 liepen uiteen van respectievelijk 2-8, 4-10 en 3-8 punten. Over de totale groep gezien was de helft van de gegeven antwoordopties goed. Verder bleek

dat bij vragen 3 en 5 het vaakst de juiste antwoordoptie werd gekozen, met respectievelijk 18 en 15 goede antwoordopties. Bij vragen 9 en 10 werd het minst vaak de juiste antwoordoptie gekozen, met respectievelijk 6 en 5 goede antwoordopties.

Tabel 4

Testscores Uitgesplitst naar Groep

Testscores	Groep 6 (n = 6)		Groep 7 (n = 8)		Groep 8 (n = 8)		Totaal (n = 22)	
	M	SD	M	SD	M	SD	M	SD
IEP-score Rekenen	60.33	7.45	67.88	12.71	76.38	13.62	68.91	13.10
IEP-score Lezen	67.50	4.59	73.88	16.80	81.25	13.23	74.82	13.74
Aantal goed	4.67	2.34	6.25	1.83	5.38	1.85	5.50	1.99
Aantal fout	7.50	2.07	5.75	5.25	6.50	1.60	6.50	2.02

Noot. Aantal goed = aantal goed gegeven antwoordopties (ongeacht redenering), Aantal fout = aantal foute redeneringen (ongeacht antwoordoptie).

Vervolgens is gekeken naar het aantal vragen waarbij een foute redenering werd gegeven, ongeacht de gekozen antwoordoptie. In Tabel 4 is te zien dat groep 6 gemiddeld de meeste redeneerfouten had en groep 7 de minste. Totaal gezien werd bij meer dan de helft van de vragen een onjuiste redenering gegeven. Het minimum aantal redeneerfouten was 2, het maximum 10. Verder is te zien dat het aantal goede antwoordopties en het aantal foute redeneringen niet complementair aan elkaar waren; het geven van een juiste antwoordoptie betekende niet dat de bijbehorende redenering ook juist was, en andersom. Er zijn door alle leerlingen gezamenlijk 143 redeneerfouten gemaakt, die konden worden gegroepeerd in 11 soorten (zie Tabel 5).

Correlatiecoëfficiënten en voorspellende waarde IEP-scores op aantal foute redeneringen

Een bivariate Pearson's r is gebruikt om de grootte en richting van de lineaire relatie tussen de variabelen vast te stellen. Voor de totale groep was de correlatie tussen de IEP-score Rekenen en de IEP-score Lezen significant en sterk, $r(20) = .664$, $p < .001$. De IEP-score Rekenen hing niet samen met het aantal redeneerfouten, $r(20) = -.330$, $p = .134$, en hetzelfde gold voor de IEP-score Lezen, $r(20) = -.323$, $p = .143$. De negatieve correlaties houden in dat een hogere IEP-score samenhangt met minder redeneerfouten (en dus een hoger aantal goed beredeneerde antwoorden). Gezien het grote aantal redeneerfouten gebaseerd op voorkennis is

ook hier een correlatie voor uitgevoerd tussen de IEP-scores en het aantal redeneerfouten. De bivariate relatie tussen IEP-score Lezen en het aantal redeneerfouten in voorkennis was niet-significant, $r(20) = -.284$, $p = .200$. De bivariate relatie tussen IEP-score Rekenen en het aantal redeneerfouten in voorkennis was evenmin significant, $r(20) = -.188$, $p = .401$.

Tabel 5*Soorten Redeneerfouten en hun Kenmerken*

	Foutcode	Foutomschrijving	Aantal keer voorgekomen	% van totaal	Ratio
Afkomstig uit Kranz et al. (2022)	1 - E	Variabelen niet controleren	8	6.0%	18.2%
	4 - B	Moelijkheden bij het lezen van grafieken en/of tabellen	3	2.1%	6.8%
	5 - B	Causale interpretatie van noncovariatie	3	2.1%	13.6%
	6 - B	Moelijkheden bij confounded data → Verkeerde interpretatie; het negeren van verwarrende aard van de data, wel referentie naar data	5	3.5%	22.7%
	7 - M	Variabelen niet kunnen identificeren / afgaan op verkeerde variabele	6	4.2%	5.5%
	8 - M	Uitproberen → Leerlingen volgen geen systematische aanpak bij het experimenteren en gaan daarom niet op een wetenschappelijke manier te werk	5	3.5%	11.4%
Nieuwe codes	2 - E	Niet-sluitend experiment kiezen bij twee tegenstrijdige hypothesen (verschil tussen testen hypothese en produceren van een effect)	15	10.5%	68.2%
	3 - E	Niet de juiste ontwerpfout ontdekken → Een verklaring die een ontwerpkenmerk bevat dat minder cruciaal is voor de validiteit van het experiment	9	6.3%	40.9%
	9 - O	Afgaan op voorkennis / opvattingen om de vraag te beantwoorden	76	53.1%	31.4%
	10 - O	Zoeken naar (woord)associaties tussen vraag en antwoordopties	7	4.9%	2.9%
	11 - O	Geen inhoudelijke reden geven	6	4.2%	2.5%
	Totaal		143	100%	

Noot. E = experimenten opzetten en uitvoeren, B = bewijs evalueren, M = valt onder meerdere onderzoeksfases. O = overige fouten. Ratio = aantal keer dat de fout kan voorkomen ÷ het aantal keer dat de fout is voorgekomen.

Aangezien de grootte van de correlaties behoorlijk was, ondanks dat deze niet significant waren, is een standaard multipele regressie uitgevoerd om de relaties tussen de voorspellende variabelen IEP-score Rekenen en IEP-score Lezen en redeneerfouten verder te onderzoeken. Voorafgaand zijn de assumpties van normaliteit, outliers, multicollineariteit, en de normaliteit, lineariteit en homoscedasticiteit van residuen gecontroleerd (Allen et al., 2014). Aan alle assumpties werd voldaan. Tot slot is gekeken naar de verhouding tussen het aantal deelnemers (N) en het aantal predictoren (k). Voor het testen van een volledig regressiemodel is een respondentenaantal van $50 + 8(k)$ nodig, wat in huidig onderzoek neerkomt op 33 deelnemers. Aan deze assumptie is niet voldaan, gezien $n = 22$. Deze analyse

is echter wel uitgevoerd in het belang van het onderzoek en er is dus verder gewerkt met deze data. Hierdoor moeten de resultaten met voorzichtigheid bekeken worden en zijn deze vooral richtinggevend.

Uit de multiële regressieanalyse bleek dat de twee voorspellers samen een niet-significante 12.8% van de variantie verklaarden in het aantal redeneerfouten, $R^2 = .13$, adjusted $R^2 = .04$, $F(2, 19) = 1.40$, $p = .272$. Cohen's f^2 was .15, dit is een gemiddeld effect. De ongestandaardiseerde (B) en gestandaardiseerde (β) regressiecoëfficiënten en de squared semi-partial correlaties (sr^2) voor elke voorspeller in het regressiemodel zijn gerapporteerd in Tabel 6. Deze tabel 6 laat verder zien dat geen enkele variabele significant voorspellend was. Alle variabelen waren negatief voorspellend, wat inhoudt dat een lagere IEP-score zorgt voor een hoger aantal redeneerfouten.

Tabel 6

Ongestandaardiseerde (B) en Gestandaardiseerde (β) Regressiecoëfficiënten en Squared Semi-Partial Correlaties (sr^2) voor Elke Voorspeller in een Multiële Regressie Model dat Redeneerfouten op Onderzoeksvaardigheden Voorspelt

Variabele	B [95% CI]	β	sr^2	t	p
Constant	10.73 [-12.69, 12.03]			4.17	<.001
IEP-score Lezen	-.03 [-.12, .06]	-.19	.02	-.65	.524
IEP-score Rekenen	-.03 [-.12, .06]	-.21	.02	-.72	.480

Noot. $N = 22$. CI = confidence interval.

Soorten redeneerfouten

In totaal zijn na het coderen 11 verschillende soorten fouten naar voren gekomen. De vijf redeneerfouten uit Tabel 5 met de hoogste ratio-score worden hieronder aan de hand van enkele voorbeelden besproken.

Meer dan de helft van de gemaakte redeneerfouten was gebaseerd op verkeerde voorkennis/opvattingen, samengevoegd onder fout 9. Bij vraag 10 bijvoorbeeld denkt de directeur, na één andere school met grotere klaslokalen en betere cijfers te hebben bezocht, dat de leerlingen van zijn school slechte cijfers halen omdat de klaslokalen te klein zijn en de leerlingen zich niet goed kunnen concentreren. De vraag is of hij gelijk heeft. Een leerling gaf hier (het onjuiste) antwoord 2 ('Ja, hij heeft gelijk. Leerlingen kunnen zich niet concentreren in kleine klaslokalen.'). "want als leerlingen in kleine klaslokalen zitten, hebben ze minder ruimte en zitten ze dicht bij elkaar en dan heb je minder ruimte om je te concentreren". De leerling gaat hier af op eigen ideeën, in plaats van op te merken dat ook naar andere

variabelen gekeken moet worden. Bij vraag 2 wordt de snelheid van mieren onderzocht, waarbij wordt gedacht dat mieren sneller lopen bij een hogere buitentemperatuur. Bijgevoegd is een tabel met resultaten met de vraag welke conclusie je hieruit kan trekken. Een leerling koos hier voor (het onjuiste) antwoord B ('De warmte heeft invloed op de snelheid van het lopen.'). "want als je rent is er heel veel zon en dan is het vermoeiender om te lopen voor de mieren". Ook deze leerling gebruikt voorkennis om de vraag te beantwoorden, in plaats van de informatie uit de tabel. Bij vraag 5 staat een grafiek met resultaten van de brandtijd van een kaars ten opzichte van flessen van verschillende grootte. De volgende zin moet afgemaakt worden: *De resultaten van dit experiment laten zien dat hoe groter de fles, hoe...* Een leerling koos hier voor (het juiste) antwoord B (... langer de kaars brandde), "want met een grotere fles kan je meer rook opvangen. Want als je meer rook opvangt dan blijft hij [de kaars] langer branden". Alle voorgaande leerlingen gingen dus af op informatie anders dan die vermeld in de vraag/antwoorden. Deze foutsoort kwam daarnaast regelmatig voor bij vragen 2 en 5.

Verder valt in de tabel op dat redeneerfouten 1, 2, 3 en 6 naar verhouding vaker voorkwamen dan de andere fouten. Een voorbeeld van fout 1 is te vinden bij vraag 11. Hierin wil Sandra onderzoeken waardoor het waterpeil in een waterbak stijgt als je er een blokje in stopt. Ze heeft twee bakken met water en blokjes van verschillende grootte en gewicht. De vraag is hoe Sandra kan laten zien dat grote blokjes zorgen voor een grotere stijging van het waterpeil dan kleine blokjes. Een leerling koos hier voor (het onjuiste) antwoord 1 ('Ze kan een groot en een klein blokje gebruiken die een verschillend gewicht hebben.'). "want het zware blokje is groter en het lichte blokje is kleiner en ze wil testen of het waterpeil hoger stijgt met een groter blokje". De leerling controleert hier de variabelen niet, aangezien deze niet opmerkt dat het kleinere blokje een ander gewicht heeft dan het grotere blokje.

Fout 2 kwam enkel en veelvuldig voor bij vraag 9. Hier zijn verschillende ideeën te zien over of een aardvarken wel of niet goed kan ruiken, met als onderzoeksvoorwerpen een doos, sterk geurende kaas en niet-geurend brood. De vraag was waar de kaas of het brood neergelegd moet worden om te onderzoeken of het aardvarken wel of niet goed kan ruiken. Een leerling gaf hier (het onjuiste) antwoord 2 ('kaas in de doos'), "want op de doos kunnen ze het [de kaas] goed ruiken en als je het in de doos doet, kunnen ze het minder goed ruiken. En als ze het dan wel ruiken, dan weet je dat het aardvarken goed kan ruiken". Deze redenering werd meerdere keren in soortgelijke woorden gegeven. Deze is fout, want de redenering bij het gekozen antwoord gaat maar af op één van de twee hypothesen, namelijk dat het aardvarken goed kan ruiken. Er wordt dus gekozen voor een niet-sluitend experiment, waardoor de andere hypothese niet wordt gefalsificeerd.

Fout 3 kwam enkel en veelvuldig voor bij vraag 10. Een leerling gaf hier (het onjuiste) antwoord 1 ('Nee. De directeur had meer dan één andere school moeten vragen. '), " ... want het kon zijn dat dit een rijke school is en dus meer kan betalen en bij sommige scholen zijn de klaslokalen klein. Niet bij alle scholen zijn de klaslokalen heel groot of heel klein". Een andere leerling koos hier dezelfde antwoordoptie, "want dan kan hij het vergelijken met andere scholen die wel goede cijfers halen en als het dan allemaal op één plek valt dan weten ze het". Hoewel de redeneringen bij de gegeven antwoordoptie passen, is dit niet de ontwerpfout waar het om draait. De andere variabelen worden genegeerd, dat is de juiste ontwerpfout. Als er geen andere variabelen waren geweest, waren deze redeneringen wel juist geweest.

Fout 6 kwam enkele keren voor bij vraag 8. Een leerling koos hier voor antwoord 3 ("De verse kruiden hebben de inwoners genezen, omdat iedereen die het drankje met verse kruiden heeft gedronken geen buikpijn meer heeft."), "want de inwoners die zijn genezen, hebben op het plaatje wel verse kruiden en de zieke inwoners niet". Dit is een verkeerde interpretatie van de data waarbij er wel een verwijzing naar de data is, maar de verwarrende aard van de data wordt genegeerd. Er wordt enkel verwezen naar één van de twee verschillen tussen de gezonde en de zieke inwoners; ook de wortels verschilden namelijk tussen de groepen.

Discussie

In dit onderzoek is gekeken naar de redeneerfouten die bovenbouwleerlingen maken op een geschreven test van onderzoeksvaardigheden en de samenhang tussen het aantal gemaakte redeneerfouten en rekenvaardigheid en begrijpend lezen. Om dit te onderzoeken is bij 22 leerlingen uit de bovenbouw van het basisonderwijs een schriftelijke test voor onderzoeksvaardigheden afgenomen en zijn de IEP-scores van begrijpend lezen en rekenvaardigheid verzameld. Er zijn in totaal elf soorten redeneerfouten geobserveerd, die in verschillende mate voorkwamen. Het aantal redeneerfouten hing echter niet samen met de vaardigheid in begrijpend lezen en rekenen en de verschillen tussen leerlingen wat betreft het aantal redeneerfouten konden niet worden verklaard door de gezamenlijke vaardigheden in begrijpend lezen en rekenen.

Samenhang leerlingkenmerken en redeneerfouten

Bij het tweede deel van de onderzoeksvraag werd verwacht dat een lagere score op begrijpend lezen zou samenhangen met meer redeneerfouten op alle onderzoeksvaardigheden afzonderlijk (Van de Sande et al., 2019) en gezamenlijk (Lazonder & Janssen, 2021; Mayer et al., 2014; Schiefer et al., 2021). Voor rekenvaardigheid werd verwacht dat een lagere score

zou samenhangen met meer redeneerfouten op alle onderzoeksvaardigheden tezamen, en met de afzonderlijke deelvaardigheid ‘bewijs evalueren’ (Lazonder & Janssen, 2021; Mayer et al., 2014; Schlatter et al., 2021). Gezien het ontoereikende aantal deelnemers kon echter geen uitspraak worden gedaan over de afzonderlijke onderzoeksvaardigheden en is enkel gekeken naar alle onderzoeksvaardigheden gezamenlijk. Uit de analyses bleek een negatieve samenhang tussen zowel IEP-score Lezen als IEP-score Rekenen en het totale aantal redeneerfouten. Hoe hoger de IEP-score, hoe minder redeneerfouten en dus hoe meer goed beredeneerde antwoorden. Deze correlaties waren echter niet significant, zoals wel in eerdere onderzoeken naar voren kwam voor zowel begrijpend lezen (Koerber et al., 2015; Lazonder & Janssen, 2021; Mayer et al., 2014; Schiefer et al., 2019; Van de Sande et al., 2019) als rekenen (Lazonder & Janssen, 2021; Mayer et al., 2014). Daarnaast bleken de scores op IEP-Rekenen en IEP-Lezen samen een niet-significante invloed te hebben op het gemiddelde aantal redeneerfouten. Wat betreft rekenvaardigheid komt dit resultaat overeen met het onderzoek van Schlatter et al. (2021), waarin dit geen significante voorspeller bleek van de variantie in gemiddelde scores op een onderzoeksvaardigheidentest.

Het feit dat deze leerlingkenmerken niet significant voorspellend zijn en geen significante samenhang vertonen, kan ten eerste komen doordat de steekproef te klein was. Er hadden volgens het gebruikte regressiemodel met 2 voorspellers eigenlijk 33 leerlingen moeten deelnemen, en dit waren er nu 22. Ook is het leerlingkenmerk rekenvaardigheid vanwege praktische redenen gemeten met de gehele IEP-toets en dus alle bijbehorende domeinen, open en gesloten vragen en kale sommen en contextsommen. Wellicht dat bepaalde onderdelen van rekenvaardigheid wel een significante samenhang vertonen met onderzoeksvaardigheden. In eerdere onderzoeken is dit bijvoorbeeld gemeten met ruimtelijk inzicht (Mayer et al., 2014). Daarnaast zaten in de gebruikte test relatief weinig vragen die enige mate van rekenvaardigheid vereisten. Verder waren de meegenomen IEP-scores afkomstig uit het midden van het schooljaar (november 2022 - januari 2023) en dus wellicht niet compleet representatief voor het niveau op het moment van huidig onderzoek.

Soorten redeneerfouten

Voor het eerste deel van de onderzoeksvraag was geen hypothese opgesteld. Uit de exploratieve resultaten bleek dat in totaal 11 soorten redeneerfouten werden gemaakt. Fouten 1, 4, 5, 6, 7, 8 en 9 kwamen al eerder naar voren in de review van Kranz et al. (2022), waarbij fout 9 (‘afgaan op voorkennis’) daar niet als losstaand werd gezien, maar deel uitmaakte van andere fouten. Opvallend was dat fouten 2 (niet-sluitend experiment kiezen) en 3 (niet de juiste ontwerpfout ontdekken) naar verhouding veel voorkwamen. Uit eerder onderzoek blijkt

dat leerlingen rond groep 3/4 al een sluitende test kunnen kiezen om een hypothese te testen en onderscheid kunnen maken tussen het testen van hypothesen en het produceren van effecten (fout 2) (Piekny et al., 2014; Sodian et al., 1991). Er waren in deze eerdere onderzoeken echter twee ‘experimentopties’ om tussen te kiezen en in huidig onderzoek was er ook een ‘tussenoptie’; de antwoordoptie die vaak is gekozen, met bijbehorende ‘logische’ maar foutieve redenering. Ook werd in eerder onderzoek meer inhoudelijk verduidelijkende informatie gegeven over het experiment. Hoewel het in verhouding grote aantal redeneerfouten bij deze vraag lijkt in te gaan tegen eerdere onderzoeken, is dit verschil in verwacht niveau dus wellicht te verklaren door het verschil in gebruikte meetinstrumenten en procedure. Daarnaast lijkt het erop dat leerlingen in groep 6-8 wel onderscheid kunnen maken tussen hypothese testen en effect produceren, maar dat ze dit alleen kunnen in eenvoudige situaties. Bij meer dan twee opties gaat het vaak fout en blijkt uit de redenering dat de leerlingen kiezen voor de ‘tussenoptie’, waarbij ze een experiment kiezen dat enkel één hypothese test en de andere hypothese dus niet uitsluit.

Fout 3 kwam enkel voor bij vraag 10. In eerder onderzoek zijn sterk gelijksoortige vragen voorgelegd aan leerlingen van 8-10 jaar (Osterhaus et al., 2015). Daar wordt bij de drie mogelijke antwoordopties de ‘tussenoptie’ (tussen goed en fout) omschreven als ‘intermediate’; als die wordt gekozen, toont dat een eerste inzicht in de noodzaak van hypothese testen en het bestaan van relevante ontwerpkenmerken, zoals het belang van steekproefomvang. Dit belang werd in het huidige scriptieonderzoek vaak als onderbouwing gebruikt voor de gegeven antwoordoptie. Daarmee scoren de leerlingen in dit onderzoek beter dan in het onderzoek van Osterhaus et al. (2015), waar voornamelijk het foute antwoord werd gekozen. Wellicht dat dit komt doordat de leerlingen in huidig onderzoek ouder zijn (9-13 jaar).

Fout 10 is nog niet eerder in de literatuur over onderzoeksvaardigheden naar voren gekomen. Dit zoeken naar overeenkomstige woorden in de vraag en de antwoordopties is een strategie van testwijsheid. Dit is het vermogen om teststrategieën te gebruiken bij het beantwoorden van meerkeuzevragen, zonder de inhoudelijke kennis of vaardigheden te bezitten die worden getest (Allan, 1992). Bij huidig onderzoek werkte deze strategie echter niet en leidde automatisch tot een onjuiste redenering.

Tot slot fout 11, waarbij leerlingen geen redenering gaven. Ondanks dat vooraf werd aangegeven dat de test nergens voor meetelde, zou het uitblijven van een inhoudelijke reden kunnen komen doordat de leerlingen het idee hadden geen fouten te mogen maken en zich nog niet goed genoeg op hun gemak voelden bij de onderzoeker. Het onderwerp van de test

was veelal onbekend terrein voor de leerlingen en het kan zijn dat zij zich verbaal niet zeker genoeg voelden om een redenering te geven. Ook kan het zijn dat bepaalde leerlingen minder motivatie hadden om mee te doen en daarom snel door de vragen heen gingen.

Opvallend was dat meer dan de helft van de gemaakte redeneerfouten gebaseerd was op verkeerde voorkennis/opvattingen (fout 9). Hoewel het bij de vaardigheid ‘hypothese opstellen’ niet per definitie verkeerd is om voorkennis te gebruiken als je deze hypothese zelf moet opstellen, en voorkennis zelfs leidend kan zijn bij het selecteren van een hypothese (Croker & Buchanan, 2011; Klahr & Dunbar, 1988), was dit in het huidige onderzoek anders. Bij de betreffende vragen stond namelijk informatie in de vorm van tekst en plaatjes die gebruikt moesten worden om de juiste hypothese te kiezen. Ook bij de andere onderzoeksvaardigheden gingen leerlingen vaak af op voorkennis. Dit komt overeen met eerdere onderzoeken, waarin leerlingen bijvoorbeeld bij het evalueren van bewijs nauwelijks gebruikmaken van oordelen gebaseerd op bewijs, maar afgaan op hun voorkennis en opvattingen (Amsel & Brock, 1996; Masnick & Morris, 2008). Verder viel op dat sommige leerlingen bij de vragen met een grafiek of tabel (deelvaardigheid bewijs evalueren) deze oversloegen, wat een verklaring kan zijn voor het in plaats daarvan afgaan op voorkennis. Tot slot zou dit vele afgaan op voorkennis kunnen komen doordat de leerlingen op school niet of nauwelijks kennis over of ervaring hadden met onderzoeksvaardigheden; pas vanaf schooljaar 2023-2024 wordt schoolbreed W&T aangeboden.

Concluderend kan gesteld worden dat er veel redeneerfouten zijn gemaakt, maar de vraag is wat dit aantal en de aard van de fouten zegt over de beheersing van de onderzoeksvaardigheden. Ten eerste bleek het afgaan op voorkennis een veelvoorkomende strategie bij alle onderzoeksvaardigheden. Het woord ‘hypothese’ bleek veelal onbekend, en zelfs na een korte toelichting hierop gebruikten leerlingen niet de beschikbare informatie om de gevraagde hypothese op te stellen. Hypotheses opstellen zoals in dit onderzoek bedoeld, dus met behulp van de gegeven informatie en daarin bijvoorbeeld de juiste variabelen identificeren, is lastig.

Bij het opzetten en uitvoeren van een onderzoek werd de CVS soms niet of verkeerd gebruikt. Daarnaast kwamen redeneerfouten 2 (niet-sluitend experiment kiezen) en 3 (niet de juiste ontwerpfout ontdekken) in verhouding vaak voor, maar deze verdienen een genuanceerder beeld. Beide fouten behoorden toe aan de ‘tussenoptie’ van de beschikbare antwoordopties en waren dus niet de minst goede keuze. De leerlingen kunnen kort gezegd al wel enig onderscheid maken tussen hypothese testen en effecten genereren wanneer er drie antwoordopties zijn; met twee opties wordt waarschijnlijk de juiste gekozen. Daarnaast blijkt

uit de redeneringen bij fout 3 dat de leerlingen al een eerste inzicht hebben in het belang van hypothese testen en relevante ontwerpkenmerken. Een goed experiment bedenken en opzetten blijkt voor de leerlingen dus zelfs met meerkeuzevragen lastig, al geven de redeneringen wel enige blijk van deze onderzoeksvaardigheid.

Wat betreft de laatste onderzoeksvaardigheid, bewijs evalueren, werden diverse fouten een paar keer gemaakt. Leerlingen negeerden bijvoorbeeld de verstoorde aard van de data, maar verwezen hierbij wel naar de data. Ook werden grafieken wel gebruikt als informatiebron, maar werd bijvoorbeeld het aflezen ervan verkeerd gedaan. Tot slot kan het niet geven van een redenering of het zoeken van associaties verder aangeven dat de leerlingen onbekend zijn met de manier van vragen en de onderzoeksvaardigheden hierin.

Hoewel er dus geen W&T-onderwijs wordt gegeven op de school en er veel redeneerfouten zijn gemaakt, laten de redeneringen ook al enkele belangrijke aspecten van de onderzoeksvaardigheden zien. De leerlingen zijn zich hier echter waarschijnlijk niet van bewust en het verdient dan ook de aandacht om deze al aanwezige (beginnende) vaardigheden meer onder de aandacht te brengen, dit verder uit te breiden en de redeneerfouten aan te pakken.

Limitaties

Het huidige onderzoek heeft een aantal beperkingen. Ten eerste heeft maar één school deelgenomen aan het onderzoek. Ook hadden er 33 deelnemers moeten zijn voor een betrouwbaar regressiemodel en dit waren er slechts 22. De resultaten zijn daarom niet te generaliseren en het is daardoor moeilijk om stellige uitspraken te doen over de resultaten; deze moeten meer als richtinggevend worden beschouwd. Verder is de onderzoeksvaardigheid ‘onderzoeksvraag opstellen’ in huidig onderzoek buiten beschouwing gelaten, waardoor geen compleet beeld van het construct onderzoeksvaardigheden is verkregen. Deze vaardigheid zou in toekomstig onderzoek wel meegenomen kunnen worden om ook te zien hoe deze samenhangt met redeneerfouten. Daarnaast is het leerlingkenmerk rekenvaardigheid enkel gemeten als geheel, en konden er dus geen uitspraken worden gedaan over bepaalde onderdelen van rekenvaardigheid. Toekomstig onderzoek zou wel onderscheid kunnen maken tussen de verschillende domeinen en de meest relevante hiervan voor wat betreft onderzoeksvaardigheden meenemen.

Tevens is er voor de keuze van de vragen voornamelijk gekeken naar de fouten uit Kranz et al. (2022) die daarin voor zouden kunnen komen (zie Tabel 2 en 3). Aangezien er maar beperkte tijd was per leerling zijn slechts 11 vragen meegenomen. Daardoor zijn

wellicht ook andere redeneerfouten gemist die wel gemaakt hadden kunnen worden als er meer of andere vragen waren meegenomen, zoals het verkeerd aflezen van instrumenten.

Bovendien bevatten de antwoordopties bij de vragen uit de test van Osterhaus et al. (2020) al redeneringen. Dit was voornamelijk het geval bij vragen 7 (Simanezen) en 10 (directeur). Bij bijvoorbeeld een ja/nee-vraag had dus wellicht verder geen uitleg moeten staan, zodat de leerlingen deze zelf konden geven in de vorm van een redenering. Nu is de antwoordoptie in feite al de reden en werden leerlingen wellicht dus al de kant opgeduwd van een redenering en daarmee ook een eventuele fout.

Tot slot moet bij de evaluatie van de resultaten rekening worden gehouden met de mogelijke communicatie-effecten tussen de leerlingen en de onderzoeker. Ondanks dat er een instructietekst was opgesteld om de afnames zo gelijk mogelijk te houden, zijn af en toe vragen gesteld over de antwoorden en gedachten van de leerlingen. Het kan zijn dat zij hierdoor zijn aangezet tot nadenken en hun antwoordoptie en/of redenering hebben herzien.

Wat betreft de sterke punten kan worden gezegd dat de gebruikte vragen valide en betrouwbaar zijn, aangezien deze afkomstig zijn uit de valide en betrouwbare testen van Kruit et al. (2018) en Osterhaus et al. (2020). Ook de gebruikte IEP-toetsen zijn voldoende valide en betrouwbaar (Expertgroep Toetsen PO, z.d.) en er bleek na het coderen van de redeneerfouten sprake van een hoge interbeoordelaarsbetrouwbaarheid. Tot slot geeft het hoge aantal redeneerfouten aan dat er voldoende ruimte was voor foute redeneringen; dit impliceert dat de vragen toereikend waren voor het doel van het onderzoek.

Implicaties en vervolgonderzoek

De bevindingen van het huidige onderzoek hebben implicaties voor zowel de theorie als de praktijk op het gebied van onderzoeksvaardigheden. Ten eerste blijkt uit de resultaten dat het aangeven van de juiste antwoordoptie niet betekent dat leerlingen een juiste redenering geven en dus de getoetste vaardigheid voldoende begrijpen of beheersen. Het lijkt dus van belang dat leerkrachten niet alleen bij foute, maar ook bij goede antwoordopties nagaan wat de redenering hierachter is, om eventuele misconcepties tegen te gaan. Nog beter zou zijn om onderzoeksvaardigheden met open vragen of praktische opdrachten te toetsen. Dit is meteen een implicatie voor toetsontwikkelaars; toetsen ontwikkelen met open vragen, of in het geval van meerkeuzevragen ruimte om het gekozen antwoord toe te lichten. Daarbij moet in acht worden genomen dat er geen redenering in de antwoordoptie moet staan; in dat geval is er grote kans op het herhalen van de antwoordoptie als redenering. Verder ligt het voor de hand om bij deze leeftijdsgroep twee antwoordopties te geven bij vragen over het onderscheid tussen hypothese testen en effect genereren; dit sluit aan bij het gemiddelde

vaardigheidsniveau. Een vraag met drie antwoordopties kan worden gebruikt om te toetsen hoeveel leerlingen deze vaardigheid volledig beheersen. Daarnaast zou het van belang kunnen zijn om bij (schriftelijke) vragen duidelijk aan te geven dat naar een plaatje/grafiek/tabel gekeken moet worden, in plaats van dit meer beschrijvend of terloops te noemen in de vraag. Op die manier is er wellicht minder kans dat leerlingen deze relevante informatie negeren en daardoor bijvoorbeeld afgaan op (foutieve) voorkennis om de vraag te beantwoorden.

Verder wordt aangeraden om meer in te zetten op W&T-onderwijs in onderzoeksvaardigheden. De resultaten uit huidig onderzoek suggereren het belang hiervan; deze zijn gemiddeld, maar er is nog voldoende ruimte voor groei en de deelnemende school biedt nog geen W&T-onderwijs aan. Hierbij is het echter wel belangrijk dat leerkrachten zelf voldoende onderzoeksvaardig zijn, zodat zij redeneerfouten herkennen en hierop kunnen inspelen. Goede begeleiding is namelijk van belang om het verwerven van onderzoeksvaardigheden effectief te laten zijn (zie o.a. Aditomo & Klieme, 2020; D'Angelo et al., 2014; Lazonder & Harmsen, 2016).

Met betrekking tot vervolgonderzoek is het interessant om het onderzoek uit te breiden naar een grotere leerlingpopulatie met meer onderzoeksvaardigheden, om zo de generaliseerbaarheid van de resultaten te vergroten. Daarnaast is het zinvol om andere variabelen te onderzoeken die mogelijk van invloed zijn op het aantal redeneerfouten bij onderzoeksvaardigheden. In eerder onderzoek zijn leerlingkenmerken als SES (Koerber et al., 2015), intelligentie (Koerber et al., 2015; Mayer et al., 2014; Schiefer et al., 2021) en probleemoplossingsvaardigheden (Mayer et al., 2014; Schlatter et al., 2021) meegenomen en die bleken (in enkele gevallen) een significant effect te hebben op de gemiddelde scores op een onderzoeksvaardigheidentest. Ook motivatie voor W&T-onderwijs is in eerder onderzoek naar voren gekomen als een variabele die een significante invloed heeft op de prestaties van leerlingen op dit vlak (Cavas, 2011; Chan & Norlizah, 2017). Het onderzoeken van deze variabelen kan een bredere kijk bieden op de factoren die van invloed zijn op het aantal redeneerfouten in onderzoeksvaardigheden. Tot slot is het interessant om het huidige onderzoek met deze verbeteringen te herhalen bij jongere leerlingen (tot 10 jaar). Op die manier komen wellicht andere of juist vergelijkbare redeneerfouten aan het licht. Als deze eenmaal bekend zijn, kunnen leerkrachten hier in een vroeger stadium rekening mee houden en formatieve beoordeling gericht toepassen.

Conclusie

De huidige studie vond 11 soorten redeneerfouten in een schriftelijke toets naar onderzoeksvaardigheden, die veelvuldig en in verschillende mate voorkwamen. Verder bleek

het aantal redeneerfouten onafhankelijk van rekenvaardigheid en begrijpend lezen. Het grote aantal redeneerfouten suggereert het belang van goed W&T-onderwijs, waarbij leerkrachten goed op de hoogte dienen te zijn van de mogelijke fouten en deze gebruiken als formatieve informatie voor de inrichting van toekomstige lessen. Voor toetsontwikkelaars geldt dat zij enkele toetsinhoudelijke aanpassingen kunnen doen om dit (formatief) beoordelen van onderzoeksvaardigheden eenvoudiger te maken en nog beter de redenering achter het uitvoeren van onderzoeksvaardigheden te kunnen achterhalen.

Referenties

- Aditomo, A., & Klieme, E. (2020). Forms of inquiry-based science instruction and their relations with learning outcomes: Evidence from high and low-performing education systems. *International Journal of Science Education*, 42(4), 504-525.
<https://doi.org/10.1080/09500693.2020.1716093>
- Alfieri, L., Brooks, P. J., Aldrich, N. J., & Tenenbaum, H. R. (2011). Does discovery-based instruction enhance learning?. *Journal of Educational Psychology*, 103(1), 1-18.
<https://doi.org/10.1037/a0021017>
- Allan, A. (1992). Development and validation of a scale to measure test-wiseness in EFL/ESL reading test takers. *Language Testing*, 9(2), 101-119.
<https://doi.org/10.1177/02655322920090020>
- Allen, P., Bennett, K., & Heritage, B. (2014). *SPSS statistics version 22: A practical guide* (3e editie). Cengage Learning Australia Pty Limited.
- Amsel, E., & Brock, S. (1996). The development of evidence evaluation skills. *Cognitive Development*, 11(4), 523-550. [https://doi.org/10.1016/S0885-2014\(96\)90016-7](https://doi.org/10.1016/S0885-2014(96)90016-7)
- Baur, A. (2021). Errors made by 5th-, 6th-, and 9th-graders when planning and performing experiments: Results of video-based comparisons. *Zeitschrift für Didaktik Der Biologie (ZDB)-Biologie Lehren Und Lernen*, 25, 45-63.
<https://doi.org/10.11576/zdb-3576>
- Baur, A. (2023). Which student problems in experimentation are related to one another? *International Journal of Science Education*, 1-25.
<https://doi.org/10.1080/09500693.2023.2175334>
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31.
<https://doi.org/10.1007/s11092-008-9068-5>
- Boaventura, D., Faria, C., Chagas, I., & Galvão, C. (2013). Promoting science outdoor activities for elementary school children: Contributions from a research laboratory. *International Journal of Science Education*, 35(5), 796-814.
<https://doi.org/10.1080/09500693.2011.583292>
- Boston, C. (2002). The concept of formative assessment. *Practical Assessment, Research, and Evaluation*, 8. <https://doi.org/10.7275/kmcq-dj31>
- Bullock, M., & Ziegler, A. (1999). Scientific reasoning: Developmental and individual differences. In F. E. Weinert & W. Schneider (Red.), *Individual development from 3-*

- 12: *Findings from the Munich longitudinal study* (pp. 38-54). Cambridge University Press.
- Cavas, P. (2011). Factors affecting the motivation of Turkish primary students for science learning. *Science Education International*, 22(1), 31-42.
- Chan Y. L., & Norlizah C. H. (2017). Students' motivation towards science learning and students' science achievement. *International Journal of Academic Research in Progressive Education and Development*, 6(4), 174-189.
<https://doi.org/10.6007/IJARPED/v6-i4/3716>
- Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development*, 70(5), 1098–1120.
<https://doi.org/10.1111/1467-8624.00081>
- Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research*, 63(1), 1-49. <https://doi.org/10.3102/00346543063001001>
- Croker, S., & Buchanan, H. (2011). Scientific reasoning in a real-world context: The effect of prior belief and outcome on children's hypothesis-testing strategies. *British Journal of Developmental Psychology*, 29(3), 409-424.
<https://doi.org/10.1348/026151010X496906>
- D'Angelo, C., Rutstein, D., Harris, C., Bernard, R., Borokhovski, E., & Haertel, G. (2014). Simulations for STEM learning: Systematic review and meta-analysis. *Menlo Park: SRI International*, 5(23), 1-5.
- De Jong, T., & Van Joolingen, W. R. (1998). Scientific discovery learning with computer simulations of conceptual domains. *Review of Educational Research*, 68(2), 179-201.
<https://doi.org/10.3102/0034654306800217>
- Dean, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education*, 91(3), 384-397. <https://doi.org/10.1002/sce.20194>
- Djoyoadhiningrat-Hol, K., & Klein Tank, M. (2022). *Trendanalyse Wetenschap & Technologie: Ontwikkelingen en uitdagingen bij Wetenschap & Technologie in het basisonderwijs*. SLO. Geraadpleegd op 13 maart 2023, van <https://www.slo.nl/publicaties/@21403/trendanalyse-wetenschap-technologie/>
- Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science*, 17(3), 397-434. https://doi.org/10.1207/s15516709cog1703_3

- Dunbar, K., & Klahr, D. (1989). Developmental differences in scientific discovery processes. In D. Klahr & K. Kotovsky (Red.), *Complex information processing* (pp. 129-164). Psychology Press.
- Dunbar, K., & Klahr, D. (2012). Scientific thinking and reasoning. In K. J. Holyoak & R. G. Morrison (Red.), *The Oxford handbook of thinking and reasoning* (pp. 701–718). Cambridge University Press.
<https://doi.org/10.1093/oxfordhb/9780199734689.013.0035>
- Expertgroep Toetsen PO. (z.d.). *Kwaliteitsoordelen van LVS-instrumenten*. Geraadpleegd op 14 juli 2023, van <https://www.expertgroepoetsenpo.nl/c/lvs-instrumenten#1b65274/kwaliteitsoordelen-van-lvs-instrumenten>
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., Reinhard, P., Neuhaus, B., Dorner, B., Pankofer, S., Fischer, M., Strijbos, J., Heene, M., & Eberle, J. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, 2(3), 28-45.
<https://doi.org/10.14786/flr.v2i2.96>
- Furtak, E. M., Seidel, T., Iverson, H., & Briggs, D. C. (2012). Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis. *Review of Educational Research*, 82(3), 300-329. <https://doi.org/10.3102/0034654312457206>
- Gott, R., & Duggan, S. (2003). *Understanding and using scientific evidence: How to critically evaluate data*. Sage.
- Hamman, M., Phan, T. T. H., Ehmer, M., & Grimm, T. (2008). Assessing pupils' skills in experimentation. *Journal of Biological Education*, 42(2), 66-72.
<https://doi.org/10.1080/00219266.2008.9656113>
- Harlen, W., & Jelly, S. (1990). *Developing science in the primary classroom*. Vintage.
- Hmelo-Silver, C. E., Duncan, R. G., & Chinn, C. A. (2007). Scaffolding and achievement in problem-based and inquiry learning: A response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42(2), 99-107.
<https://doi.org/10.1080/00461520701263368>
- Kanari, Z., & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching*, 41(7), 748–769.
<https://doi.org/10.1002/tea.20020>
- Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based,

- experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75-86.
https://doi.org/10.1207/s15326985ep4102_1
- Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. MIT press.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12(1), 1-48. https://doi.org/10.1207/s15516709cog1201_1
- Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science*, 15(10), 661-667. <https://doi.org/10.1111/j.0956-7976.2004.00737.x>
- Klahr, D., Fay, A. L., & Dunbar, K. (1993). Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology*, 25(1), 111-146.
<https://doi.org/10.1006/cogp.1993.1003>
- Koerber, S., & Osterhaus, C. (2019). Individual differences in early scientific thinking: Assessment, cognitive influences, and their relevance for science learning. *Journal of Cognition and Development*, 20(4), 510-533.
<https://doi.org/10.1080/15248372.2019.1620232>
- Koerber, S., Mayer, D., Osterhaus, C., Schwippert, K., & Sodian, B. (2015). The development of scientific thinking in elementary school: A comprehensive inventory. *Child Development*, 86(1), 327-336. <https://doi.org/10.1111/cdev.12298>
- Koerber, S., Sodian, B., Thoermer, C., & Nett, U. (2005). Scientific reasoning in young children: Preschoolers' ability to evaluate covariation evidence. *Swiss Journal of Psychology*, 64(3), 141-152. <https://doi.org/10.1024/1421-0185.64.3.141>
- Köksal-Tuncer, Ö., & Sodian, B. (2018). The development of scientific reasoning: Hypothesis testing and argumentation from evidence in young children. *Cognitive Development*, 48, 135-145. <https://doi.org/10.1016/j.cogdev.2018.06.011>
- Kranz, J., Baur, A., & Möller, A. (2022). Learners' challenges in understanding and performing experiments: a systematic review of the literature. *Studies in Science Education*, (59)2, 321-367. <https://doi.org/10.1080/03057267.2022.2138151>
- Krummenauer, J., & Kuntze, S. (2019, 6-10 februari). *Primary students' reasoning and argumentation based on statistical data* [paperpresentatie]. In the Eleventh Congress of the European Society for Research in Mathematics Education, Utrecht, Nederland. Geraadpleegd op 29 maart 2023, van <https://hal.science/hal-02398118/document>
- Kruit, P. M., Oostdam, R. J., van den Berg, E., & Schuitema, J. A. (2018). Assessing students' ability in performing scientific inquiry: Instruments for measuring science skills in

- primary education. *Research in Science & Technological Education*, 36(4), 413-439.
<https://doi.org/10.1080/02635143.2017.1421530>
- Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science*, 16(11), 866-870. <https://doi.org/10.1111/j.1467-9280.2005.01628.x>
- Kuhn, D., Amsel, E., O'Loughlin, M., Schauble, L., Leadbeater, B., & Yotive, W. (1988). *The development of scientific thinking skills*. Academic Press.
- Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction*, 9(4), 285-327.
https://doi.org/10.1207/s1532690xci0904_1
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174. <https://doi.org/10.2307/2529310>
- Lazonder, A. W., & Harmsen, R. (2016). Meta-analysis of inquiry-based learning: Effects of guidance. *Review of Educational Research*, 86(3), 681-718.
<https://doi.org/10.3102/0034654315627366>
- Lazonder, A. W., & Janssen, N. (2021). Development and initial validation of a performance-based scientific reasoning test for children. *Studies in Educational Evaluation*, 68, 100951. <https://doi.org/10.1016/j.stueduc.2020.100951>
- Mayer, D., Sodian, B., Koerber, S., & Schwippert, K. (2014). Scientific reasoning in elementary school children: Assessment and relations with cognitive abilities. *Learning and Instruction*, 29, 43-55.
<https://doi.org/10.1016/j.learninstruc.2013.07.005>
- Masnack, A. M., & Klahr, D. (2003). Error matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development*, 4(1), 67-98. <http://dx.doi.org/10.1080/15248372.2003.9669683>
- Masnack, A. M., & Morris, B. J. (2008). Investigating the development of data evaluation: The role of data characteristics. *Child Development*, 79(4), 1032-1048.
<https://doi.org/10.1111/j.1467-8624.2008.01174.x>
- Masnack, A. M., Klahr, D., & Knowles, E. R. (2017). Data-driven belief revision in children and adults. *Journal of Cognition and Development*, 18(1), 87-109.
<https://doi.org/10.1080/15248372.2016.1168824>
- Mommers, A., Van Langen, A., Mulder, L., Netten, A., & Lamers, I. (2016). *Peilingsonderzoek natuur & techniek: Rapportage school- en leerlingvragenlijst*. ITS, Radboud Universiteit Nijmegen. Geraadpleegd op 27 maart, van

- <https://www.onderwijsinspectie.nl/documenten/rapporten/2017/05/31/peilingsonderzoek-natuur-techniek-its>
- Osterhaus, C., Koerber, S., & Sodian, B. (2015). Children's understanding of experimental contrast and experimental control: An inventory for primary school. *Frontline Learning Research, 3*(4), 56-94. <https://doi.org/10.14786/flr.v3i4.220>
- Osterhaus, C., Koerber, S., & Sodian, B. (2020). The Science-P Reasoning Inventory (SPR-I): Measuring emerging scientific-reasoning skills in primary school. *International Journal of Science Education, 42*(7), 1087-1107. <https://doi.org/10.1080/09500693.2020.1748251>
- Pedaste, M., Baucal, A., & Reisenbuk, E. (2021). Towards a science inquiry test in primary education: development of items and scales. *International Journal of STEM Education, 8*(1), 1-19. <https://doi.org/10.1186/s40594-021-00278-z>
- Pedaste, M., Mäeots, M., Leijen, Ä., & Sarapuu, T. (2012). Improving students' inquiry skills through reflection and self-regulation scaffolds. *Technology, Instruction, Cognition and Learning, 9*(1-2), 81-95.
- Pedaste, M., Mäeots, M., Siiman, L. A., De Jong, T., Van Riesen, S. A., Kamp, E. T., Manoli, C. C., Zacharia, Z. C., & Tsourlidaki, E. (2015). Phases of inquiry-based learning: Definitions and the inquiry cycle. *Educational Research Review, 14*, 47-61. <https://doi.org/10.1016/j.edurev.2015.02.003>
- Piekny, J., & Maehler, C. (2013). Scientific reasoning in early and middle childhood: The development of domain-general evidence evaluation, experimentation, and hypothesis generation skills. *British Journal of Developmental Psychology, 31*(2), 153-179. <https://doi.org/10.1111/j.2044-835X.2012.02082.x>
- Piekny, J., Grube, D., & Maehler, C. (2014). The development of experimentation and evidence evaluation skills at preschool age. *International Journal of Science Education, 36*(2), 334-354. <https://doi.org/10.1080/09500693.2013.776192>
- Rijksoverheid. (2013). *Nationaal Techniepact 2020*. Geraadpleegd op 14 maart 2023, van <https://www.rijksoverheid.nl/documenten/convenanten/2013/05/13/nationaal-techniepact-2020>
- Sandoval, W. A., Sodian, B., Koerber, S., & Wong, J. (2014). Developing children's early competencies to engage with science. *Educational Psychologist, 49*(2), 139-152. <https://doi.org/10.1080/00461520.2014.917589>
- Schauble, L., Glaser, R., Duschl, R. A., Schulze, S., & John, J. (1995). Students' understanding of the objectives and procedures of experimentation in the science

- classroom. *The Journal of the Learning Sciences*, 4(2), 131-166.
https://doi.org/10.1207/s15327809jls0402_1
- Schiefer, J., Golle, J., Tibus, M., & Oschatz, K. (2019). Scientific reasoning in elementary school children: Assessment of the inquiry cycle. *Journal of Advanced Academics*, 30(2), 144-177. <https://doi.org/10.1177/1932202X1882515>
- Schlatter, E., Lazonder, A. W., Molenaar, I., & Janssen, N. (2021). Individual differences in children's scientific reasoning. *Education Sciences*, 11(9), 471.
<https://doi.org/10.3390/educsci11090471>
- Siegler, R. S., & Chen, Z. (1998). Developmental differences in rule learning: A microgenetic analysis. *Cognitive Psychology*, 36(3), 273-310.
<https://doi.org/10.1006/cogp.1998.0686>
- Siler, S., Klahr, D., Magaro, C., Willows, K., & Mowery, D. (2010, 14-18 juni). *Predictors of transfer of experimental design skills in elementary and middle school children* [Paperpresentatie]. Intelligent Tutoring Systems: 10th International Conference, Pittsburgh, Pennsylvania, Verenigde Staten. Geraadpleegd op 2 april 2023, van https://link.springer.com/chapter/10.1007/978-3-642-13437-1_20
- Sodian, B., Zaitchik, D., & Carey, S. (1991). Young children's differentiation of hypothetical beliefs from evidence. *Child Development*, 62(4), 753-766.
<https://doi.org/10.1111/j.1467-8624.1991.tb01567.x>
- Tschirgi, J. E. (1980). Sensible reasoning: A hypothesis about hypotheses. *Child Development*, 51(1), 1-10. <https://doi.org/10.2307/1129583>
- Van de Sande, E., Kleemans, T., Verhoeven, L., & Segers, E. (2019). The linguistic nature of children's scientific reasoning. *Learning and Instruction*, 62, 20-26.
<https://doi.org/10.1016/j.learninstruc.2019.02.002>
- Van der Graaf, J., Segers, E., & Verhoeven, L. (2015). Scientific reasoning abilities in kindergarten: dynamic assessment of the control of variables strategy. *Instructional Science*, 43(3), 381-400. <https://doi.org/10.1007/s11251-015-9344-y>
- Van der Graaf, J., Segers, E., & Verhoeven, L. (2016). Scientific reasoning in kindergarten: Cognitive factors in experimentation and evidence evaluation. *Learning and Individual Differences*, 49, 190-200. <https://doi.org/10.1016/j.lindif.2016.06.006>
- Van der Graaf, J., Segers, E., & Verhoeven, L. (2018). Individual differences in the development of scientific thinking in kindergarten. *Learning and Instruction*, 56, 1-9. <https://doi.org/10.1016/j.learninstruc.2018.03.005>

- Van Graft, M., Tank, M. K., Beker, T., & Van der Laan, A. (2014). *Wetenschap & technologie in het basis en speciaal onderwijs: richtinggevend leerplankader bij het leergebied Oriëntatie op jezelf en de wereld*. SLO, nationaal expertisecentrum leerplanontwikkeling.
- Van Joolingen, W. R., & De Jong, T. (1991). Supporting hypothesis generation by learners exploring an interactive computer simulation. *Instructional Science*, 20(5-6), 389-404. <https://doi.org/10.1007/BF00116355>
- Van Someren, M., Barnard, Y. F., & Sandberg, J. (1994). *The think aloud method: A practical approach to modelling cognitive processes*. Academic Press.
- Wagensveld, B., Segers, E., Kleemans, T., & Verhoeven, L. (2015). Child predictors of learning to control variables via instruction or self-discovery. *Instructional Science*, 43(3), 365-379. <https://doi.org/10.1007/s11251-014-9334-5>
- Wenham, M. (1993). The nature and role of hypotheses in school science investigations. *International Journal of Science Education*, 15(3), 231-240. <https://doi.org/10.1080/0950069930150301>
- Wininger, S. R. (2005). Using your tests to teach: Formative summative assessment. *Teaching of Psychology*, 32(3), 164-166. https://doi.org/10.1207/s15328023top3203_7
- Wu, H. K., & Hsieh, C. E. (2006). Developing sixth graders' inquiry skills to construct explanations in inquiry-based learning environments. *International Journal of Science Education*, 28(11), 1289-1313. <https://doi.org/10.1080/09500690600621035>
- Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review*, 20(1), 99-149. <https://doi.org/10.1006/drev.1999.0497>
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27(2), 172-223. <https://doi.org/10.1016/j.dr.2006.12.001>
- Zion, M., Cohen, S., & Amir, R. (2007). The spectrum of dynamic inquiry teaching practices. *Research in Science Education*, 37(4), 423-447. <https://doi.org/10.1007/s11165-006-9034-5>

Bijlage A

Test onderzoeksvaardigheden

Naam leerling:

School:

Juf/meester:

Groep:

Leeftijd:

Meisje: 0

Jongen: 0

Nodig:

✓ Potlood



IDnummer leerling:

Voorbeeldopgave

Mick en Diana werkten samen aan een borduurwerkje. Mick had sneller de draad in de naald dan Diana. Ze besloten om te onderzoeken of jongens sneller waren in het draad in de naald doen dan meisjes. Ze vroegen 10 jongens en 10 meisjes om hen te helpen. Ze verzamelden draad, naalden en een stopwatch.

Welke onderzoeksvraag hadden zij bedacht?

- A. Zijn jongens beter in borduren dan meisjes?
- B. Kan Mick de draad sneller in de naald doen dan Diana?
- C. Kunnen jongens de draad sneller in de naald doen dan meisjes?
- D. Waarom kunnen jongens de draad sneller in de naald doen dan meisjes?

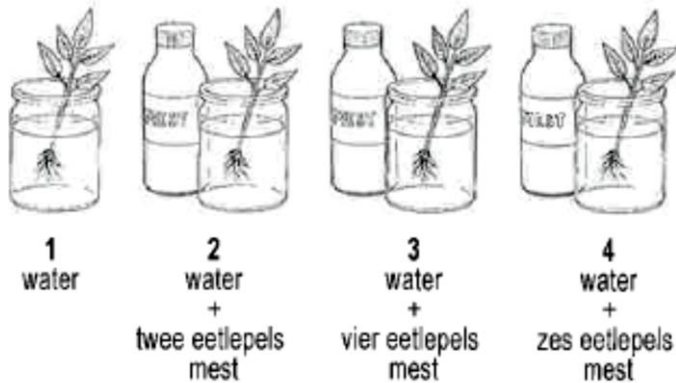
**Probeeropgave**

Dinah vroeg zich af of het vaste land en de oceanen even snel door de zon opwarmen. Zij besloot om dat te onderzoeken. Zij vulde een emmer met zand en een emmer met water en zette de emmers in de zon. De temperatuur in beide emmers werd elk uur opgemeten.

Welke onderzoeksvraag hoort bij dit onderzoek?

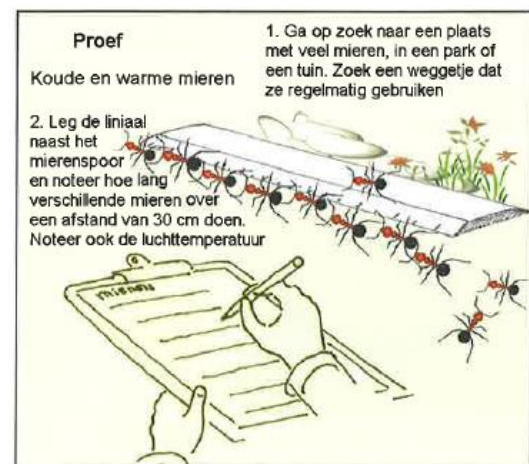
- A. Worden het zand en het water warmer als er een grotere hoeveelheid zonlicht is?
- B. Worden het zand en het water warmer als ze langer in de zon staan?
- C. Is het zand eerder opgewarmd door de zon dan het water?
- D. Komen verschillende hoeveelheden zonlicht binnen op verschillende momenten van de dag?

1. Op school doet Kim een proef. Ze zet vier dezelfde kamerplanten in vier potten. Onder de tekening kun je lezen wat ze verder nog in de potten doet. Daarna zet ze de potten op de vensterbank. Na enkele weken wil ze de planten met elkaar vergelijken. Vooraf bedenkt Kim een voorspelling over wat er gaat gebeuren. Welke voorspelling is het best?



- A: De hoeveelheid water is van invloed op hoe snel de planten groeien.
B: De hoeveelheid mest is van invloed op hoe snel de planten groeien.
C: De plant met 6 eetlepels mest groeit het snelst.
D: De plant zonder mest groeit het snelst.
2. Om de snelheid van mieren te onderzoeken doet Sam een experiment zoals op de uitlegkaart beschreven staat. Hij denkt dat mieren sneller lopen als de temperatuur buiten hoger is. Hij noteert de volgende resultaten:

Temperatuur	Tijd gelopen voor afstand van 30 cm
10 ⁰ C	10 minuten
15 ⁰ C	6 minuten
20 ⁰ C	2 minuten
25 ⁰ C	3 minuten
30 ⁰ C	4 minuten



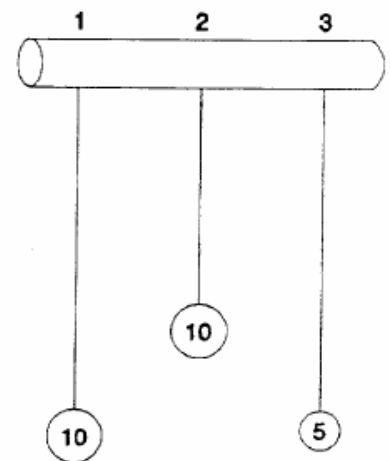
Welke conclusie kan Sam trekken uit de resultaten?

- A. Hoe warmer het wordt, hoe sneller de mieren gaan lopen.
B. De warmte heeft invloed op de snelheid van het lopen.
C. Bij steeds hogere temperaturen neemt de snelheid van lopen eerst toe en daarna weer af.
D. Bij 20⁰ C doen de mieren er 2 minuten over.




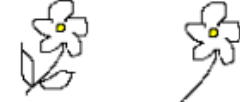

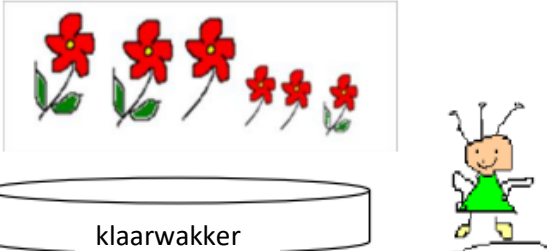
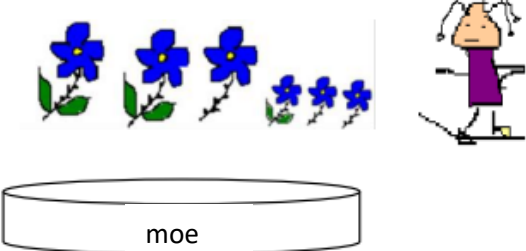
3. Jet wil weten hoe hoog de bloeddruk is van leerlingen van 12 jaar. Zij meet de bloeddruk van een jongen en van een meisje van 12 jaar. Ze berekent het gemiddelde. Hoe kan Jet haar onderzoek verbeteren?
- A. Door alleen bij jongens of alleen bij meisjes te meten.
 - B. Door alleen bij leerlingen te meten die veel sporten.
 - C. Door bij veel meer leerlingen van 12 jaar te meten.
 - D. Door een leerkracht de meting te laten doen.

Vraag 4 gaat over het plaatje hieronder. Op het plaatje zie je drie touwtjes met daaraan een gewichtje hangen. Touwen 1 en 3 zijn even lang en touw 2 heeft een andere lengte. Aan twee touwen hangt hetzelfde gewichtje en aan het andere touw een kleiner gewichtje. De touwtjes kunnen heen en weer slingeren en de tijd om 5x heen en weer te slingeren kan gemeten worden.

4. Stel dat je wil weten of het gewicht aan de touwtjes invloed heeft op de tijd die het kost om 5x heen en weer te slingeren. Welk touw of welke touwtjes moet je dan gebruiken om hier achter te komen?
- A. alleen één touwtje
 - B. alle touwtjes
 - C. touwtje 2 en touwtje 3
 - D. touwtje 1 en touwtje 3
 - E. touwtje 1 en touwtje 2

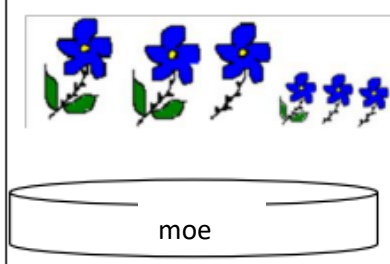
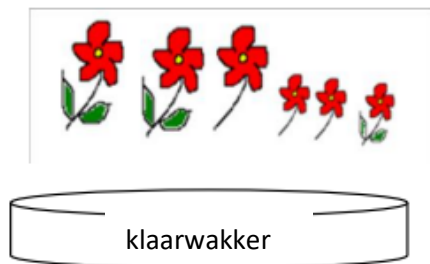


Vraag 7

<p>Op de planeet 'Sima' staan bloemen. Als de Simanezen deze bloemen plukken, worden ze moe óf klaarwakker. Onderzoekers willen uitzoeken hoe dit kan. Ze denken na of de reden zou kunnen zijn...</p>	
<p>...de grootte?</p>	 <p style="text-align: center;">groot klein</p>
<p>...de kleur?</p>	 <p style="text-align: center;">rood blauw</p>
<p>...de blaadjes?</p>	 <p style="text-align: center;">met blaadjes zonder blaadjes</p>
<p>....de stekels?</p>	 <p style="text-align: center;">met stekels zonder stekels</p>
<p>De onderzoekers deden een onderzoek. De Simanezen plukten verschillende bloemen. Daarna schreven de onderzoekers op of de Simanezen moe óf klaarwakker werden.</p> <p>Kijk naar de twee plaatjes hieronder. Daar zie je wat de onderzoekers ontdekten.</p>	
 <p style="text-align: center;">klaarwakker</p>	 <p style="text-align: center;">moe</p>

De onderzoekers ontdekten dat het niet uitmaakt of de bloemen groot of klein zijn. Soms waren de Simanezen moe en soms waren ze klaarwakker.

Kijk nog een keer naar de plaatjes!



Welke reden zouden de onderzoekers kunnen geven voor hun mening?

De onderzoekers zouden kunnen zeggen: “Het maakt niet uit of de bloemen groot of klein zijn...”

Ze zouden dit zeggen

Ze zouden dit niet zeggen

1. ...omdat de Simanezen die klaarwakker zijn grote én kleine bloemen plukten, net als de Simanezen die moe zijn.”





2. ...omdat de grote bloemen vroeger klein waren.”

3. ...omdat grote bloemen de Simanezen moe maakten, net als de kleine bloemen.”



Wat is het beste antwoord?

No. _____





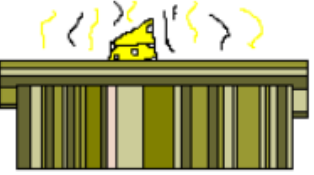
Vraag 8

<p>In het dorpje Kruidenweide woont een oude vrouw die veel weet over kruiden. Ze wil een kruidendrankje maken dat buikpijn geneest. Ze heeft veel kruiden, maar weet niet zeker welke ze moet gebruiken:</p>	
<p>Gedroogde of verse kruiden?</p>	 <p>gedroogd vers</p>
<p>Veldkruiden of boskruiden?</p>	 <p>veldkruiden boskruiden</p>
<p>Met of zonder wortel?</p>	 <p>met wortel zonder wortel</p>
<p>De oude vrouw maakt verschillende drankjes en probeert ze uit bij de inwoners met buikpijn. Zo kan ze zien door welke kruiden de inwoners genezen en welke niet. Kijk naar de plaatjes en zie wat de oude vrouw ontdekte.</p>	
 <p>gezond</p>	 <p>ziek</p>

Kijk nog een keer goed naar de plaatjes!

 <p style="text-align: center;">gezond</p>	 <p style="text-align: center;">ziek</p>	
Welke kruiden genezen de inwoners? Wat zou de oude vrouw zeggen?		
De oude vrouw zou zeggen:	Ze zou dit zeggen	Ze zou dit <u>niet</u> zeggen
1. "De wortels hebben de inwoners genezen. Omdat de kracht in de wortels zit."	<input type="checkbox"/>	<input type="checkbox"/>
2. "Het zouden de verse kruiden en de wortels kunnen zijn die de inwoners hebben genezen."	<input type="checkbox"/>	<input type="checkbox"/>
3. "De verse kruiden hebben de inwoners genezen, omdat iedereen die het drankje met verse kruiden heeft gedronken geen buikpijn meer heeft."	<input type="checkbox"/>	<input type="checkbox"/>
Wat is het <u>beste</u> antwoord?	No. _____	

Vraag 9

<p>Paula en Suzie hebben een nieuw huisdier. Het is een heel zeldzaam dier, een aardvarken. Paula en Suzie hebben verschillende ideeën over of hun nieuwe huisdier goed kan ruiken of niet.</p>		
<p>Ze pakken een doos, kaas die je heel goed ruikt en brood dat je niet goed ruikt. Nu denken ze na over waar ze de kaas of het brood moeten neerleggen om te onderzoeken of hun huisdier een goede neus heeft of niet.</p>		
<p>Waar zouden Paula en Suzie de kaas of het brood moeten neerleggen om uit te vinden of hun huisdier wel of niet goed kan ruiken?</p>		
	<p>goed</p>	<p>fout</p>
<p>1.  brood in de doos</p>	<input type="checkbox"/>	<input type="checkbox"/>
<p>2.  kaas in de doos</p>	<input type="checkbox"/>	<input type="checkbox"/>
<p>3.  kaas boven op de doos</p>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Wat is het <u>beste</u> antwoord?</p>	<p>No. _____</p>	

Vraag 10

<p>De leerlingen in School X halen heel slechte cijfers. Volgens de directeur kunnen daar drie redenen voor zijn.</p>		
<p>1. De leerkrachten zijn te streng. 2. De leerlingen zijn te lui. 3. De klassen zijn te klein en de leerlingen kunnen zich niet goed concentreren.</p>		
<p>Op een school in de buurt halen de leerlingen heel goede cijfers. De directeur vraagt zijn collega over de grootte van de klaslokalen in die school. En inderdaad: de klaslokalen van de andere school zijn veel groter.</p>		
<p>De directeur denkt dat de leerlingen van zijn school slechte cijfers halen omdat de klaslokalen te klein zijn en de leerlingen zich niet goed kunnen concentreren.</p>		
<p>Heeft de directeur gelijk?</p>		
	<p>ja</p>	<p>nee</p>
<p>1. Nee. De directeur had meer dan één andere school moeten vragen.</p>	<input type="checkbox"/>	<input type="checkbox"/>
<p>2. Ja, hij heeft gelijk. Leerlingen kunnen zich niet concentreren in kleine klaslokalen.</p>	<input type="checkbox"/>	<input type="checkbox"/>
<p>3. Nee. Het zou ook kunnen dat de leerkrachten op deze school te streng zijn of dat de leerlingen te lui zijn.</p>	<input type="checkbox"/>	<input type="checkbox"/>
<p>Wat is het <u>beste</u> antwoord?</p>	<p>Nr. _____</p>	

Vraag 11

Sandra wil onderzoeken waardoor het waterpeil in een waterbak stijgt als je er een blokje in stopt. (Het blokje zinkt.)

Sandra heeft verschillende ideeën. Eerst wil ze het volgende idee testen:
Door grote blokjes stijgt het waterpeil meer dan door kleine blokjes!

Sandra wil haar idee testen. Daarvoor pakt ze twee waterbakken met dezelfde hoeveelheid water erin. Nu kan ze in elke waterbak een blokje stoppen en deze vergelijken om te zien in welke waterbak het waterpeil hoger stijgt.



Sandra denkt na over welke blokjes ze moet gebruiken om te laten zien dat grote blokjes zorgen voor een hogere stijging van het waterpeil dan kleine blokjes.

Hoe kan Sandra laten zien dat grote blokjes zorgen voor een grotere stijging van het waterpeil dan de kleine blokjes?

		Een goed idee	<u>Geen</u> goed idee
1. Ze kan een groot en een klein blokje gebruiken die een verschillend gewicht hebben.		<input type="checkbox"/>	<input type="checkbox"/>
2. Ze kan een groot en een klein blokje gebruiken die hetzelfde gewicht hebben.		<input type="checkbox"/>	<input type="checkbox"/>
3. Ze kan blokjes gebruiken die verschillen in grootte en in gewicht.		<input type="checkbox"/>	<input type="checkbox"/>
Wat is het <u>beste</u> antwoord?		No. _____	

Bijlage B

Antwoordenblad test onderzoeksvaardigheden: juist antwoord met juiste redenering

Vraag 1

- juist antwoord: B
- juiste redenering in de trant van: omdat de hoeveelheid water in alle potten gelijk is, en je dus kijkt of de hoeveelheid mest (het verschil tussen de potten) dan uitmaakt voor hoe snel de planten groeien.

Vraag 2

- juist antwoord: C
- juiste redenering in de trant van: in de tabel kan je zien dat hoe hoger de temperatuur, tot 20°C, hoe korter de mieren lopen over 30 cm. Daarna lopen de minuten weer op als de temperatuur omhoog gaat.

Vraag 3

- juist antwoord: C
- juiste redenering in de trant van: twee leerlingen meten is te weinig, al zijn het een jongen en een meisje. Door bij veel meer leerlingen van 12 jaar te meten, krijg je meer informatie en een beter, meer betrouwbaar gemiddelde.

Vraag 4

- juist antwoord: D
- juiste redenering in de trant van: je wil weten of het gewicht dat aan de touwtjes hangt invloed heeft op de tijd dat het kost om 5x heen en weer te slingeren. Touwtje 1 en 3 zijn even lang, maar hebben wel een verschillend gewicht. Zo is er maar 1 ding verschillend tussen de touwtjes en kan je de verschillende gewichten en bijbehorende slingertijd vergelijken.

Vraag 5

- juist antwoord: B
- juiste redenering in de trant van: in de tabel staat helemaal rechts de informatie voor de grootste fles, en je ziet dat die ook het grootst aantal minuten brandde.

Vraag 6

- juist antwoord: A
- juiste redenering in de trant van: omdat de soort aarde en de hoeveelheid zon in de bakjes gelijk is, en je dus kijkt of de hoeveelheid water (het verschil tussen de bakjes) dan uitmaakt voor hoe snel de bonenzaadjes groeien.

Vraag 7

- juist antwoord: 1
- juiste redenering in de trant van: omdat je op de plaatjes kan zien dat zowel de Simanezen die klaarwakker zijn als de Simanezen die moe zijn, grote en kleine bloemen plukten. Het verschil tussen die twee groepen kan dus niet komen door de grootte van de bloemen.

Vraag 8

- juist antwoord: 2
- juiste redenering in de trant van: de gezonde mensen hebben verse kruiden en de zieke mensen gedroogde kruiden + de gezonde mensen hebben kruiden met wortels en de zieke mensen kruiden zonder wortels. De andere soorten kruiden zijn hetzelfde tussen de twee groepen, dus het moeten de verse kruiden of de kruiden met wortels zijn die het verschil maken, maar je weet niet welke van de twee.

Vraag 9

- juist antwoord: 1
- juiste redenering in de trant van: als je het brood in de doos doet, wat geen sterke geur heeft, en het aardvarken ruikt het toch, dan weet je dat hij goed kan ruiken. Ruikt hij het brood niet, dan weet je dat hij niet goed kan ruiken.

Vraag 10

- juist antwoord: 3
- juiste redenering in de trant van: de directeur heeft nu alleen maar naar de derde reden gekeken (de klassen zijn te klein en de leerlingen kunnen zich niet goed concentreren) en daarop zijn conclusie gebaseerd. Hij moet ook nog kijken naar andere mogelijke redenen, namelijk dat de leerkrachten te streng zijn of de leerlingen te lui. Misschien

ligt het wel aan (één van) die redenen, maar dat weet je nu niet. Als er maar één variabele was (grootte van de lokalen), dan was antwoord 1 goed geweest.

Vraag 11

- juist antwoord: 2
- juiste redenering in de trant van: het gewicht van beide blokjes is gelijk (50 g), maar de grootte is verschillend. Nu is er maar één verschil tussen de blokjes, de grootte, en kan ze kijken of grotere blokjes inderdaad zorgen voor een grotere stijging van het waterpeil dan kleinere blokjes.

Bijlage C

Instructietekst afname test onderzoeksvaardigheden

- Voorafgaand aan de test

Welkom (naam), wat fijn dat je mij wil helpen met dit onderzoek! Zometeen ga je een toets maken over onderzoeksvaardigheden. Het is belangrijk om te weten dat er voor jou niks afhangt van deze toets; alleen ik weet hoe je de toets hebt gemaakt en de resultaten worden alleen gebruikt in mijn onderzoek.

Deze toets bestaat uit 11 vragen. Bij elke vraag staat eerst een verhaaltje, met daarna een vraag en 3, 4 of 5 antwoordopties. Tijdens het maken van de toets wil ik je vragen om hardop te denken. Dit betekent dat je tijdens het maken van elke vraag probeert uit te leggen wat je denkt en waarom je wel of niet voor een bepaald antwoord kiest. Waarom denk jij dat een antwoord goed of fout is? Soms kan ik je vragen om even te stoppen, zodat ik kort iets kan opschrijven. Het maken van de toets duurt ongeveer 30-45 minuten. Mocht je tijdens de toets een vraag hebben of iets niet begrijpen, dan kan je die aan me stellen. Maar: ik kan je niet helpen met het beantwoorden van de vragen, dat moet je zelf doen. Tijdens het maken van de toets mag je op elk moment besluiten om te stoppen; hier hoeft je geen reden voor te geven en dit heeft geen negatieve gevolgen. Heb je hier tot nu toe vragen over? **Zo ja, vra(a)g(en) beantwoorden.** Geef je toestemming om mee te doen aan dit onderzoek door het maken van de toets? **Zo nee, dan kan de leerling nog vertrekken.** Dan gaan we nu beginnen. Je mag eerst de vragen op de voorkant van de toets maken. **Leerling vult vragen in.** Eerst heb ik een voorbeeldopgave, zodat je weet wat de bedoeling is tijdens het maken van de toets. Deze zal ik nu voorlezen. Je kan meelesen met wat ik doe.

Voorbeeldopgave

Mick en Diana werkten samen aan een borduurwerkje. Mick had sneller de draad in de naald dan Diana. Ze besloten om te onderzoeken of jongens sneller waren in het draad in de naald doen dan meisjes. Ze vroegen 10 jongens en 10 meisjes om hen te helpen. Ze verzamelden draad, naalden en een stopwatch.

Welke onderzoeksvraag hadden zij bedacht?

- A. *Zijn jongens beter in borduren dan meisjes?*
- B. *Kan Mick de draad sneller in de naald doen dan Diana?*
- C. *Kunnen jongens de draad sneller in de naald doen dan meisjes?*
- D. *Waarom kunnen jongens de draad sneller in de naald doen dan meisjes?*

Als ik nadenk over welk antwoord ik het best vind, dan gaat dat als volgt. Ik lees A. Ik denk dat het A niet is, want dat antwoord gaat over borduren. Het verhaaltje gaat eerst wel over borduren, maar daarna staat er: *‘Ze besloten om te onderzoeken of jongens sneller waren in het draad in de naald doen dan meisjes.* Het gaat dus meer over het draad in de naald doen. Ik lees B. B is ook niet goed, want het antwoord op die vraag weten ze al, dat stond in het verhaaltje. Ik lees C. C zou goed kunnen zijn, want dat gaat over wie er sneller is, jongens of meisjes, en ze vragen 10 jongens en meisjes om mee te doen. D is niet goed, want dat is een waarom-vraag. Met het onderzoekje dat Mick en Diana willen doen, kunnen ze geen antwoord geven op deze vraag bij D. Ze willen ook niet weten waaróm, maar juist wie er sneller is. Je weet nog niet of jongens sneller zijn dan meisjes, daarom doen ze juist dit onderzoek. Daarom is antwoord C het juiste antwoord denk ik. Je hoeft zometeen niet zo uitgebreid antwoord te geven als ik nu heb gedaan, als ik maar snap waarom je kiest voor het antwoord dat jij denkt dat het beste is.

Nu mag jij ditzelfde proberen bij de oefenvraag. Probeer net als ik hardop te denken en uit te leggen waarom je voor een bepaald antwoord kiest. **Leerling maakt oefenvraag. Eventueel helpen met het hardop denken als dat nog niet goed lukt.** Dan mag je nu beginnen met de echte toets, succes!

- Tijdens de test

Leerling kent woord niet → Woord nader uitleggen.

Leerling valt stil → Ik merk dat je niet meer hardop denkt. Vergeet niet om uit te leggen wat je denkt en waarom je voor een bepaald antwoord kiest.

Leerling noemt alleen het antwoord, zonder verdere uitleg → Ik zie dat je voor antwoord X hebt gekozen. Kan je ook nog aan mij uitleggen waarom je voor antwoord X hebt gekozen?

Leerling zegt dat hij het antwoord niet weet → Probeer dan het antwoord te kiezen waarvan jij denkt dat het het best is. Of de antwoorden weg te strepen waarvan je denkt dat ze niet goed zijn.

- Na afloop van de test

Bedankt (naam) voor het maken van de toets! Hiermee heb je mij erg geholpen. Het enige wat je nu nog moet weten, is dat het belangrijk is dat je niks over deze toets doorvertelt aan andere leerlingen. Dit zodat zij niet weten wat ze te wachten staat, als zij nog aan de beurt komen.

Nogmaals bedankt voor het maken van de test, je mag nu terug naar de klas.