

Marianne van Dijke-Droogers

Freudenthal Instituut

Universiteit Utrecht

m.j.s.vandijke-droogers@uu.nl

Onderwijs Een leertraject met een black box en statistisch modelleren via simulatiesoftware

Statistische inferentie in vwo 3

Ruim vier jaar lang werkte Marianne van Dijke-Droogers als docent-onderzoeker aan een promotieonderzoek naar het aanleren van statistische inferentie, waarop zij op 30 juni zal promoveren. Gedurende deze onderzoeksperiode bestonden haar werkzaamheden uit twee dagen lesgeven op de middelbare school en drie dagen onderzoek, gefinancierd door het DUDOC-programma van het ministerie van OCW. Het onderzoek vond plaats op het Freudenthal Instituut van de Universiteit Utrecht, onder begeleiding van Paul Drijvers en Arthur Bakker. In dit onderzoek werd met name gekeken of de inferentiële statistiek op een informele manier al in de onderbouw kan worden geïntroduceerd. Dit artikel presenteert het ontwerp en de leereffecten van het ontworpen leertraject.

Belang van statistische inferentie

De overweldigende hoeveelheid data, grafieken en voorspellingen met betrekking tot de COVID-pandemie in het afgelopen jaar illustreert het essentiële belang van statistiek. Op basis van data worden ingrijpende beslissingen genomen en uitspraken gedaan, zowel door burgers als professionals. Het is daarom van belang om de statistische geletterdheid van onze leerlingen te ontwikkelen. Dit houdt in dat leerlingen toegerust worden om statistische informa-

tie te interpreteren, kritisch te beoordelen en hier conclusies uit te trekken [9].

Het werken met inferenties (gevolgtrekkingen) vormt het hart van de statistiek. Als zodanig is het essentieel dat leerlingen zich bekwamen in het interpreteren van inferenties. Bij inferenties worden op basis van steekproefdata conclusies getrokken over een groter geheel of proces. Het interpreteren van de onzekerheid en het duiden van de waarschijnlijkheid van de conclusie is veelzijdig en complex.

Problemen aanleren inferentiële statistiek

In veel landen, waaronder Nederland, wordt statistische inferentie daarom pas behandeld in de bovenbouw van het voortgezet onderwijs of in het hoger onderwijs. Uit onderzoek blijkt dat inferentiële statistiek hier een struikelblok is voor veel leerlingen en studenten. De moeilijkheden van leerlingen worden met name veroorzaakt door een beperkt begrip van kernconcepten die nodig zijn voor inferenties [5,11] zoals steekproef, variatie en verdelingen. Deze conceptuele problemen worden verergerd door een sterke onderwijsfocus op het aanleren van complexe, formele procedures.

Inzet statistische modellen en modelleren

Om de moeilijkheden van leerlingen te overbruggen is in de afgelopen decennia gezocht naar *informele* onderwijsbenaderingen om conceptueel begrip te promo-

ten. Het aanbieden van informele inferentiële activiteiten op jongere leeftijd zou het leren van de complexere inferentiële statistiek op latere leeftijd kunnen vereenvoudigen [19].

De inzet van recente digitale leeromgevingen met opties voor statistisch modelleren, zoals de simulatiesoftware van VUstat en TinkerPlots, biedt een informele aanpak om het begrip van statistische concepten en modellen te verdiepen [3,8]. Inzicht in statistische modellen is van fundamenteel belang voor het interpreteren van statistische inferenties [13].

Het inbedden van informele statistische inferentie in eerdere leerjaren lijkt veelbelovend, met name in combinatie met het gebruik van digitale leermiddelen. Er is echter nog weinig bekend over hoe we dit op een effectieve manier kunnen inzetten in ons huidige onderbouwcurriculum.

Onderzoeksvraag en methode

Dit promotieonderzoek richtte zich op de vraag:

Hoe kan een theoretisch en empirisch gebaseerd leertraject leerlingen uit vwo 3 laten kennismaken met statistische inferentie?

Dit onderzoeksproject volgde een ontwerpgerichte aanpak [1]. Deze aanpak kenmerkt zich door een cyclisch proces waarin onderwijsmateriaal voor leeromgevingen wordt ontworpen, geïmplementeerd en geëvalueerd, voor vervolgcycli van (her) ontwerp en testen [14]. Het ontwerp van het leertraject werd geïnformeerd door theorieën over herhaalde steekproeven en statistisch modelleren [2,10,12,15,18] met behulp van een black box paradigmatische context [7]. In de beginfase richtten we ons vooral op de ontwikkeling van een theoretisch gefundeerd ontwerp, met daarin een specificatie van beoogde leerdoelen en de uitwerking hiervan in een (hypothetisch) leertraject. Naarmate het onderzoek vorderde, werden meerdere interventies met het leertraject uitgevoerd in de lespraktijk en geëvalueerd. Deze interventies werden in iedere cyclus opgeschaald in zowel de lengte van het leertraject als in het aantal deelnemers.

Ontwerp van het leertraject

In dit artikel zoomen we in op de resultaten van de derde, en tevens laatste, onderzoekscyclus. Op basis van de eerste

twee onderzoekscycli, een verdiepende casestudie en aanvullend literatuuronderzoek werd het (hypothetische) leertraject (her)ontworpen voor de derde cyclus. Deze cyclus omvatte het gehele traject van acht leerstappen, opgesplitst in twee vergelijkbare delen van vier: (1) experimenteren met een fysieke black box, (2) visualiseren van verdelingen, (3) onderzoeken van steekproevenverdelingen met behulp van simulatiesoftware, (4) interpreteren van steekproevenverdelingen voor inferenties in realistische contexten. De stappen 1 tot en met 4 zijn alleen gericht op categoriale data en in de stappen 5 tot en met 8 wordt gewerkt met numerieke data. Een overzicht van het gehele leertraject is weergegeven in Tabel 1. De tijdsomvang van het leertraject bestond uit zes lesuren per deel, met een totaal van 12 lesuren.

Implementatie van het leertraject

Het leertraject werd geïmplementeerd in een interventie onder 267 leerlingen in dertien klassen op verschillende scholen. Voor de analyse van de resultaten gebruikten we pre- en posttests en interventiematerialen — bestaande uit werkbladen van leerlingen, notities van docenten en onderzoeker. Om de significante leerwinst te interpreteren, hebben we de pre- en posttestresultaten van de leerlingen uit de interventiegroep vergeleken met nationale en internationale prestaties. Voor de nationale vergelijking gebruikten we de testresultaten van een vergelijkingsgroep ($n = 217$) die het reguliere vwo 3-curriculum volgde, en de internationale vergelijking werd gedaan aan de hand van een Australische studie [4] met vergelijkbare testopzet.

De pre- en posttest

De reguliere aanpak in de nationale vergelijkingsgroep ($n = 217$) bestond uit tien tot twaalf lessen gericht op beschrijvende statistiek. Om de resultaten van de interventie- en vergelijkingsgroep te kunnen vergelijken werden pre- en posttests ontwikkeld op basis van reeds bestaande tests om de statistische geletterdheid — en dan met name de statistische inferentie — van leerlingen te onderzoeken [6,16,17]. Statistische geletterdheid kan onderverdeeld worden in drie domeinen: (1) Statistische Inferentie, (2) Grafieken en Veranderingen, (3) Kans en Gemiddelde. Het ontworpen leertraject was vooral gericht op statisti-

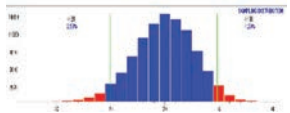
sche inferentie, het eerste domein. We vermoedden echter dat het leertraject ook een positieve invloed zou hebben op de andere twee domeinen. Het reguliere vwo 3-curriculum was voornamelijk gericht op domein twee en drie. De inzet van tests gericht op alle domeinen van statistische geletterdheid maakte het mogelijk om de resultaten van de interventiegroep te vergelijken met die van de vergelijkingsgroep.

Effecten van het leertraject

De nationale vergelijking van (post)testresultaten toonde aan dat de interventiegroep significant hoger scoorde op statistische geletterdheid, en in het bijzonder op het domein van statistische inferentie. Tabel 2 toont een overzicht van de resultaten voor de interventie- en vergelijkingsgroep. De laatste kolom in het vak met posttestresultaten toont het verschil in score tussen de interventie- en vergelijkingsgroep. Een one-way ANOVA toonde aan dat de interventiegroep significant hoger scoorde met $+0,33$ op statistische geletterdheid en $+0,67$ op statistische inferentie, beide met $p < 0,0005$. Tevens vonden we aanzienlijk positieve effecten voor de andere twee domeinen. Hoewel het leertraject niet gericht was op de andere domeinen, bleek het leertraject — bestaande uit een onderzoekgerichte aanpak met digitale middelen en meer complexe leeractiviteiten voor statistische inferentie — ook hier een positief effect te hebben.

De vergelijking van pre- en posttestresultaten maakte het mogelijk om de vooruitgang van leerlingen te onderzoeken. Deze vooruitgang is af te lezen in het onderste vak van Tabel 2 'Pre naar post'. Een gepaarde t -test toonde aan dat de interventiegroep bij de posttest significant hoger scoorde dan bij de pretest op alle domeinen van statistische geletterdheid, en met name op het domein statistische inferentie $+0,89$ met $p < 0,0005$. Tussen de pre- en posttest volgde de vergelijkingsgroep geen statistieklessen, wat zichtbaar is in de gelijkwaardige resultaten op beide tests voor deze groep. De leerlingen van de vergelijkingsgroep volgden de statistieklessen voorafgaand aan de pretest, waardoor hun resultaten op de pretest hoger zijn dan voor de interventiegroep.

De vergelijking met de internationale studie maakte het mogelijk om het geobserveerde verschil tussen beide groepen

Leerstep	Beschrijving	Voorbeeld van activiteiten	Leerdoel	Opbouw in leerstappen
<i>Categoriale data</i>				
1. Experimenteren met een fysieke black box	Fysiek 'black box met balletjes'-experiment (met klein en groot kijkvenster)	Schat het aantal gele balletjes in de black box, gevuld met 1000 balletjes, door het (herhaaldelijk) schudden en onderzoeken van het aantal zichtbare balletjes	Leerlingen formuleren inferenties en maken kennis met concepten als steekproef, steekproefomvang, steekproefvariatie, frequentie en centrum- en spreidingsmaten, in de context van herhaalde steekproeven met een fysieke black box	<p>Leerlingen ervaren dat steekproefresultaten variëren en dat een grotere steekproefomvang en meer herhalingen leiden tot een betere schatting van de populatie.</p> <p>Vervolg vraag: Wat gebeurt er als we de steekproefomvang en het aantal herhaalde steekproeven verder vergroten? Het uitvoeren van meer en grotere steekproeven is tijdrovend, daarom kan een gedachte-experiment helpen.</p> <p>De steekproevenverdeling van herhaalde steekproeven kan gebruikt worden om de waarschijnlijkheid van specifieke steekproefresultaten te bepalen.</p> <p>Vervolg vraag: Hoe kunnen we snel en eenvoudig de steekproevenverdeling van herhaalde steekproeven verkrijgen? Het gebruik van technologie kan helpen.</p> <p>Statistisch modelleren— inclusief het interpreteren van de steekproevenverdeling bij herhaalde steekproeven— kan gebruikt worden om de waarschijnlijkheid van specifieke steekproefresultaten te bepalen.</p> <p>Vervolg vraag: Kan statistisch modelleren meer algemeen gebruikt worden, in andere situaties en contexten?</p> <p>Vanuit stap 1 tot en met 4 ontstaat de vraag hoe statistisch modelleren gebruikt kan worden bij andere—niet categoriale—data</p>
				
2. Visualiseren van verdelingen	Grafiek als model (visualisatie) voor de frequentieverdeling bij herhaalde steekproeven met een black box	Schets de frequentieverdeling die je verwacht als het black box-experiment met groot venster 100000 keer wordt herhaald	Leerlingen kunnen een visualisatie tekenen van de verwachte steekproevenverdeling bij herhaalde steekproeven, leerlingen interpreteren de steekproevenverdeling om inferenties te formuleren over de waarschijnlijkheid van specifieke steekproefresultaten	
				
3. Modelleren van een black box (ICT)	Simulaties van herhaalde steekproeven met een gemodelleerde black box in een steekproevenverdeling, voor het bepalen en interpreteren van de waarschijnlijkheid van specifieke resultaten	Gebruik TinkerPlots om de meest voorkomende steekproefresultaten te bepalen voor een black box gevuld met 750 gele en 250 oranje balletje, bij een steekproefomvang van 40	Leerlingen gebruiken statistisch modelleren in de digitale omgeving van TinkerPlots om (on)waarschijnlijke steekproefresultaten te bepalen, in de context van een black box [statistisch modelleren omvat het invoeren van een model, simuleren van (herhaalde) steekproeven, visualiseren van de steekproevenverdeling en interpreteren van de resultaten]	
				
4. Modelleren van realistische contexten (ICT)	Modelleren van een specifieke situatie in TinkerPlots, voor het simuleren en interpreteren van de steekproevenverdeling en het beter begrijpen van de situatie (inclusief de bijbehorende onzekerheid)	Gebruik TinkerPlots om te bepalen welke steekproefresultaten je kunt verwachten als een steekproef met 30 leerlingen wordt uitgevoerd op een school met 300 leerlingen, bij onderzoek naar het aantal leerlingen dat dagelijks ontbijt (gegeven dat gemiddeld 70% van de vo-leerlingen dagelijks ontbijt)	Leerlingen gebruiken statistisch modelleren in de digitale omgeving van TinkerPlots voor het formuleren van inferenties, in de context van realistische probleemstellingen	
				

Tabel 1 Overzicht van stap 1–8 van het leertraject.

Leerstep	Beschrijving	Voorbeeld van activiteiten	Leerdoel	Opbouw in leerstappen
<i>Numerieke data</i>				
5. Experimenteren met een fysieke black box	Fysiek 'black box met briefjes'-experiment (de black box is gevuld met 4000 briefjes, elk briefje bevat gegevens over het geslacht en de lichaamslengte van een leerling uit klas 2, bijvoorbeeld: jongen – 155 cm)	Neem een steekproef van 40 briefjes en vat de steekproefdata samen (bereken centrum- en spreidingsmaten, gebruik visualisaties), maak een schatting van de lichaamslengte van de 4000 leerlingen uit klas 2	Leerlingen formuleren inferenties in de context van een fysieke black box met briefjes (geslacht en lichaamslengte van 4000 leerlingen uit klas 2) met inachtneming van steekproefomvang, steekproefvariatie en centrummaten	<p>Leerlingen bespreken hoe numerieke data van herhaalde steekproeven gebruikt kunnen worden om inferenties over de populatie te formuleren.</p> <p>Vervolg vraag: Hoe kan de onderliggende populatieverdeling – de lichaamslengte van de 4000 leerlingen op de briefjes in de black box – weergegeven worden op basis van de gevonden steekproefresultaten?</p> <p>Leerlingen formuleren inferenties over het populatiegemiddelde en de populatieverdeling op basis van de gevonden steekproefdata.</p> <p>Vervolg vraag: Wat zijn de effecten van grotere en meer steekproeven op de schatting van het populatiegemiddelde en de populatieverdeling? Het gebruik van technologie kan helpen bij het onderzoeken van deze effecten.</p> <p>Vanuit leerstep 7 ontstaat de vraag hoe het statistisch modelleren met numerieke data kan worden toegepast in andere contexten en situaties</p>
6. Visualiseren van verdelingen	Verwachtingen over de populatie (lengte van 4000 leerlingen) samenvatten en visualiseren op basis van de gevonden steekproefdata in leerstep 5	Schets de frequentieverdeling die je verwacht voor de totale populatie, op basis van de gevonden steekproefresultaten in leerstep 5	Leerlingen schetsen een visualisatie van de populatieverdeling die ze verwachten op basis van de gevonden steekproefresultaten, leerlingen maken inferenties over de populatieverdeling en het populatiegemiddelde	
7. Modelleren van een black box (ICT)	Experimenteren met simulaties van (herhaalde) steekproeven, bij variatie in omvang en aantal herhalingen, in de context van de black box met briefjes uit leerstep 5 en 6	Gebruik TinkerPlots om de meest voorkomende steekproefresultaten te bepalen voor de (gegeven) gemodelleerde black box met briefjes die hoort bij leerstep 5	Leerlingen gebruiken statistisch modelleren in de digitale omgeving van TinkerPlots om (on)waarschijnlijke steekproefresultaten te bepalen, in de context van de black box met briefjes [statistisch modelleren omvat hier het simuleren van (herhaalde) steekproeven vanuit een <i>gegeven</i> model, visualiseren van de steekproevenverdeling van het steekproefgemiddelde, en het interpreteren van de resultaten]	
8. Modelleren van realistische contexten (ICT)	Modelleren van een realistische situatie in TinkerPlots, voor het simuleren en interpreteren van de steekproevenverdeling, en het beter begrijpen van de situatie (inclusief de bijbehorende onzekerheid)	Gebruik TinkerPlots voor het simuleren van herhaalde steekproeven (omvang 200) vanuit een verborgen populatiemodel van 4000 personen om de sporttijd van deze personen te onderzoeken en bepalen	Leerlingen gebruiken statistisch modelleren in de digitale omgeving van TinkerPlots om inferenties te formuleren, in realistische contexten en probleemstellingen	

Tabel 1 Overzicht van stap 1–8 van het leertraject. (vervolg)

		Interventiegroep (n = 267) M (SD)	Vergelijkingsgroep (n = 217) M (SD)	Interventiegroep – vergelijkingsgroep M(I) – M(V)
Pretest	SG	2,60 (0,61)	2,97 (0,68)	– 0,37***
	SI	2,45 (0,65)	2,72 (0,71)	– 0,27***
	GV	2,07 (0,63)	2,29 (0,58)	– 0,22***
	KG	3,29 (1,38)	3,92 (1,31)	– 0,63***
Posttest	SG	3,28 (0,69)	2,95 (0,78)	+ 0,33***
	SI	3,34 (0,84)	2,67 (0,84)	+ 0,67***
	GV	2,59 (0,81)	2,38 (0,88)	+ 0,21*
	KG	3,92 (0,88)	3,80 (1,06)	+ 0,12
Pre naar post	SG	+ 0,68 (0,86)***	– 0,02 (0,73)	0,70***
	SI	+ 0,89 (0,92)***	– 0,04 (0,71)	0,93***
	GV	+ 0,52 (0,98)***	+ 0,09 (0,94)	0,43***
	KG	+ 0,63 (1,53)*	– 0,11 (1,45)	0,74***

* $p < 0,05$, ** $p < 0,005$ en *** $p < 0,0005$
 Noot: SG = statistische geletterdheid; SI, GV, KG zijn de drie domeinen binnen SG; SI = statistische inferentie; GV = grafieken en veranderingen; KG = kans en gemiddelde.

Tabel 2 Gemiddelde leerlingsscore op de domeinen van statistische geletterdheid bij de pre- en posttest voor de interventie- en vergelijkgroep, inclusief vooruitgang van pre naar post.

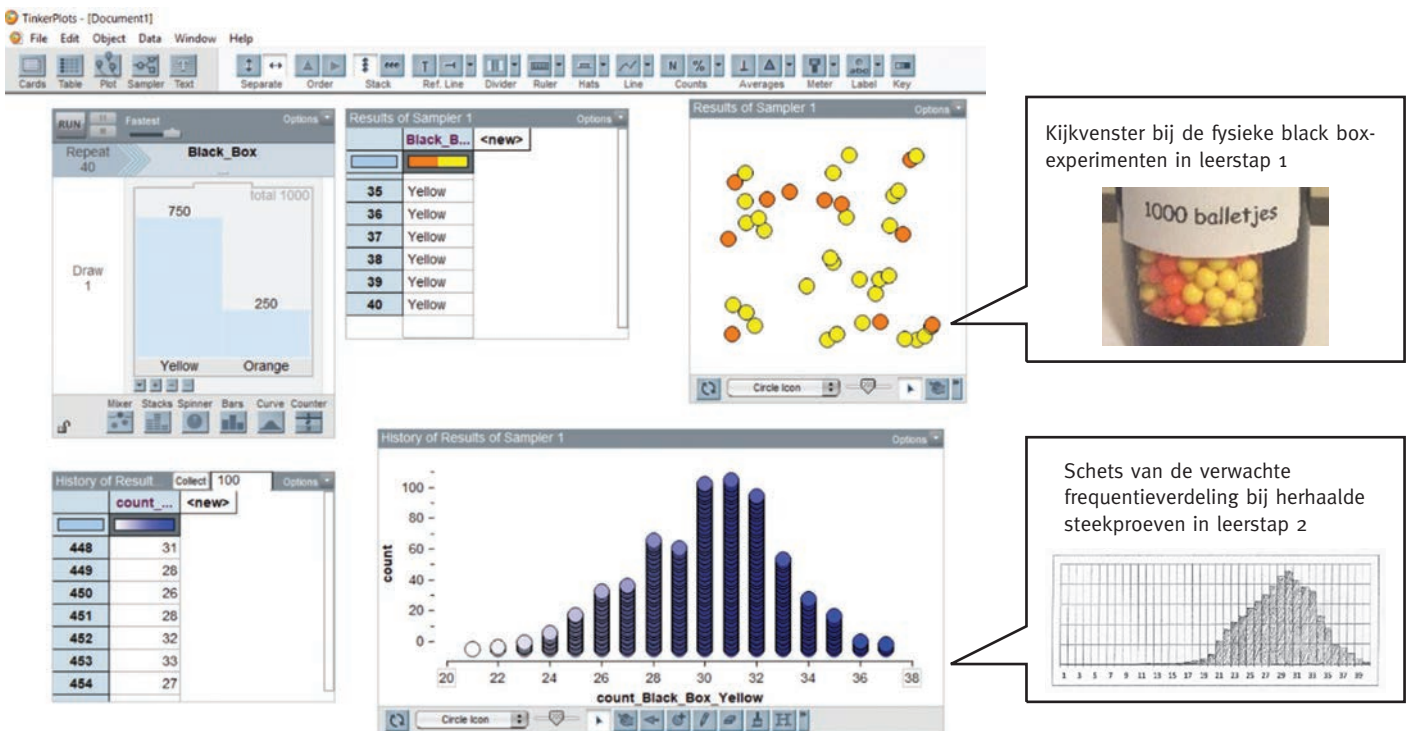
nader te interpreteren. Deze vergelijking toonde aan dat de posttestresultaten van de interventiegroep met 14–15-jarigen op statistische geletterdheid het meest overeenkwamen met die van Australische leerlingen in Grade 7–8 met een leeftijd van ongeveer 13 jaar. De resultaten van de vergelijkgroep met 14–15-jarigen waren het meest vergelijkbaar met die van

Australische leerlingen in Grade 6–7 met een leeftijd van ongeveer 12 jaar. Uit deze internationale vergelijking kunnen we concluderen dat de interventiegroep aanzienlijk hoger scoorde dan de vergelijkgroep met ongeveer één leerjaar verschil. Opvallend is dat de resultaten van beide groepen met leerlingen in de leeftijd van 14–15 jaar overeenkwamen met die van

aanzienlijk jongere Australische leerlingen. Vermoedelijk komt dit doordat het statistiekaanbod in Australië uitgebreider is dan in Nederland. De bevindingen vanuit zowel de nationale als internationale vergelijking toonden aan dat het leertraject een significant positief effect had op de statistische geletterdheid van de leerlingen, en in het bijzonder op het domein van statistische inferenties. Tevens signaleerden we positieve effecten voor de andere domeinen.

De leerstappen nader onderzocht

Om te onderzoeken *hoe* het leertraject werkt, werden zowel leerstapgerelateerde posttestitems als interventiematerialen – leerling werkbladen en docentnotities – geanalyseerd. De posttestresultaten toonden aan dat leerlingen die les kregen vanuit het leertraject significant hoger scoorden op *alle* specifiek leerstapgerelateerde doelen uit het leertraject dan leerlingen van de vergelijkgroep die het reguliere curriculum volgden. Deze leerstapgerelateerde doelen omvatten in leerstap 1 en 5 het gebruik van steekproeven, in leerstap 2 en 6 het visualiseren van verdelingen, in leerstap 3 en 7 het effect van herhaalde steekproeven en steekproefomvang, en in leerstap 4 en 8 het interpreteren van inferenties in realistische contexten. Dit



Figuur 1 De samenhang tussen het werken in de digitale omgeving van TinkerPlots in leerstap 3 (screenshot), en de fysieke black box-activiteiten in leerstappen 1 en 2 (tekstwolfsjes).

betekent dat elk onderdeel uit de opbouw in leerstappen, zoals gepresenteerd in de laatste kolom van Tabel 1, van essentieel belang is voor het totale leertraject. De analyse van werkbladen en notities van docenten en onderzoeker bevestigden het belang van de sterke samenhang in opbouw tussen elke leerstap uit het traject. Met name de koppeling tussen het fysieke experiment met de black box en de digitale leeromgeving (zoals weergegeven in Figuur 1) bevorderde het inzicht van leerlingen in statistische modellen en modelle- ren. Dit inzicht maakte het voor leerlingen

mogelijk om vervolgens in leerstap 4 en 8 inferenties te interpreteren in realistische contexten.

Conclusie en aanbeveling

De bevindingen in deze studie toonden aan dat en hoe het ontworpen leertraject werkt. Een aanpak gebaseerd op herhaalde steekproeven met een black box gecombineerd met statistisch modelleren in de digitale omgeving van TinkerPlots, bleek vruchtbaar voor het introduceren van statistische inferentie. Beide ideeën hebben tevens potentieel voor inbedding

in meer complexe vervolgvaciviteiten, zoals het toetsen van hypothesen en het vergelijken van groepen. Op basis hiervan kunnen we constateren dat huidige statistiecurricula met een sterk beschrijvende focus verrijkt kunnen worden met een inferentiële focus — in ieder geval voor de onderbouw van het vwo. Het voordeel hiervan is dat leerlingen meer leren over statistische inferenties en niet minder over de andere domeinen van statistische geletterdheid, om zo beter te anticiperen op vervolgstappen van de leerling binnen statistiekonderwijs. ❖

Referenties

- 1 A. Bakker, *Design Research in Education. A Practical Guide for Early Career Researchers*, Routledge, 2018.
- 2 R. Biehler, D. Ben-Zvi, A. Bakker en K. Maker, Technology for enhancing statistical reasoning at the school level, in M.A. Clements, A. Bishop, C. Keitel, J. Kilpatrick en F. Leung, eds., *Third International Handbook of Mathematics Education*, Springer, 2013, pp. 643–690.
- 3 R. Biehler, D. Frischemeier en S. Podworny, Editorial: Reasoning about models and modeling in the context of informal statistical inference, *Statistics Education Research Journal* 16(2) (2017), 8–12.
- 4 R. Callingham en J.M. Watson, The development of statistical literacy at school, *Statistics Education Research Journal* 17(1) (2017), 181–201.
- 5 A.E. Castro Sotos, S. Vanhoof, W. Van Den Noortgate en P. Onghena, Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education, *Educational Research Review* 1(2) (2007), 90–112.
- 6 R.C. delMas, J. Garfield, A. Ooms en B. Chance, Assessing students' conceptual understanding after a first course in statistics, *Statistics Education Research Journal* 6 (2007), 28–58.
- 7 M.J.S. van Dijke-Droogers, P.H.M. Drijvers en A. Bakker, Repeated sampling with a black box to make informal statistical inference accessible, *Mathematical Thinking and Learning* 22(2) (2020), 116–138.
- 8 M.J.S. van Dijke-Droogers, P.H.M. Drijvers en A. Bakker, Statistical modeling processes through the lens of instrumental genesis, *Educational Studies in Mathematics* (2021), <https://doi.org/10.1007/s10649-020-10023-y>.
- 9 I. Gal, Adults' statistical literacy: Meaning, components, responsibilities, *International Statistical Review* 70(1) (2002), 1–25.
- 10 J. Garfield, D. Ben-Zvi, L. Le en A. Zieffler, Developing students' reasoning about samples and sampling variability as a path to expert statistical thinking, *Educational Studies in Mathematics* 88(3) (2015), 327–342.
- 11 C. Konold en A. Pollatsek, Data analysis as the search for signals in noisy processes, *Journal for Research in Mathematics Education* 33(4) (2002), 259–289.
- 12 H. Manor en D. Ben-Zvi, Students' articulations of uncertainty in informally exploring sampling distributions, in A. Zieffler en E. Fry, eds., *Reasoning About Uncertainty: Learning and Teaching Informal Inferential Reasoning*, Catalyst Press, 2015, pp. 57–94.
- 13 H. Manor en D. Ben-Zvi, Students' emergent articulations of statistical models and modeling in making informal statistical inferences, *Statistics Education Research Journal* 16(2) (2017), 116–143.
- 14 S. McKenney en T.C. Reeves, *Conducting Educational Design Research*, Routledge, 2012.
- 15 L.A. Saldanha en P.W. Thompson, Conceptions of sample and their relationship to statistical inference, *Educational Studies in Mathematics* 51(3) (2002), 257–270.
- 16 J.M. Watson en R. Callingham, Statistical literacy: A complex hierarchical construct, *Statistics Education Research Journal* 2 (2003), 3–46.
- 17 J. Watson en R. Callingham, Statistical literacy: From idiosyncratic to critical thinking, in G. Burrill en M. Camden, eds., *Curricular Development in Statistics Education: International Association for Statistical Education Roundtable*, International Association for Statistical Education, 2004, pp. 116–137.
- 18 J. Watson en B. Chance, Building intuitions about statistical inference based on resampling, *Australian Senior Mathematics Journal* 26(1) (2012), 6–18.
- 19 A. Zieffler, J. Garfield, R. delMas en C. Reading, A framework to support research on informal inferential reasoning, *Statistics Education Research Journal* 7(2) (2008), 40–58.