

Visualisaties van onzekerheid in Cito testcores

Naam: Merle van Tuinen
Studentnummer: S3722864
Datum: Juni 2022
Cursus: Masterthese Orthopedagogiek PAMA5166
Faculteit: Gedrags- en Maatschappijwetenschappen,
Rijksuniversiteit Groningen
Eerste beoordelaar: dr. N. Frans
Tweede beoordelaar: L. Van Haaften, MA
Aantal woorden: 6460

Samenvatting

In het huidige onderwijsklimaat wordt veel gebruikgemaakt van data-gestuurde besluitvorming. De toetsscores die hier de basis van vormen gaan gepaard met onzekerheid. Leerkrachten blijken moeite te hebben met het interpreteren van toetsscores en de bijbehorende onzekerheid op basis van de huidige rapportages. Ook lijken leerkrachten niet-toereikende statistische kennis te hebben om deze onzekerheid te begrijpen. Een manier om de interpretaties te verbeteren, is het gebruik van andersoortige visualisaties. In dit onderzoek zijn de puntschatting met en zonder betrouwbaarheidsinterval (PS en PS+BHI) vergeleken met de *Probability Density Function* (PDF) en de *Quantile Dot Plot* (QDP) in hun bijdrage in de bekwaamheid van leerkrachten om onzekerheid te interpreteren. Het onderzoek volgde een experimenteel between-subject design, met een online vragenlijst onder 194 leerkrachten en studenten van de (academische) pabo. De interpretatie van onzekerheid is gemeten met de KIWS-vragenlijst, waarin participanten kansinschattingen maken op basis van de geobserveerde score en de geassocieerde nauwkeurigheid. Daarnaast is de statistische kennis van de participanten getoetst aan de hand van de LOCUS vragenlijst, om te onderzoeken of statistische kennis leidt tot accuratere interpretaties. Regressieanalyses laten geen significante verschillen zien in de accuratesse van kansinschattingen tussen de verschillende visualisaties. Wel lijkt de QDP tot betere kansinschattingen te leiden. Tussen de LOCUS en de KIWS is een trend te zien, waarbij meer statistische kennis gepaard gaat met accuratere kansinschattingen. Het is aan te raden dat in de docentenopleiding aandacht wordt besteed aan statistische vakken en dat de onzekerheid in scorerapportages op een andere manier worden weergegeven.

Abstract

In the current educational climate, many decisions are data-driven. The test scores that lay the foundation for these decisions are accompanied by uncertainty. However, teachers seem to struggle with interpreting these test scores and the accompanying uncertainty based on the current reports. Furthermore, the teachers appear to have non-sufficient statistical knowledge to fully understand the uncertainty of such scores. A way to improve the interpretations, is to visualize the uncertainty of test scores. This thesis compares the point-estimate and the point-estimate with written confidence interval with de Probability Density Function and the Quantile Dot Plot. The research followed an experimental between-subject design, in which 194 teachers and students filled in an online questionnaire. The interpretations of uncertainty were measured by use of the KIWS-questionnaire, in which participants make probability estimates of student's actual scores. Moreover, the statistical knowledge of the participants was tested with the LOCUS questionnaire, to examine whether an increase in statistical knowledge leads to more accurate interpretations. Although the QDP seem to lead to more accurate probability estimates, a regression analysis showed no significant differences. The regression analysis between the LOCUS and KIWS only shows a significant difference between the lowest and highest score. However, there seems to be a trend, where a higher score on the LOCUS is accompanied by more accurate probability estimates. Based on these findings, it is recommended that statistical courses be integrated in teacher's education and that uncertainty is visualized differently in score reports.

Data-gestuurde besluitvorming is tegenwoordig niet meer weg te denken uit het onderwijs. Ter ondersteuning van besluitvorming wordt er systematisch data verzameld, geanalyseerd en geïnterpreteerd (Datnow & Hubbard, 2015). Met name kwantitatieve data krijgen een steeds grotere rol in het besluitvormingsproces van leerkrachten (Marsh et al., 2006). Op deze data wordt bijvoorbeeld gebaseerd welke instructie een leerling nodig heeft, of welk niveau voor vervolgonderwijs passend is. In Nederland zijn leerkrachten zelfs verplicht om te werken met een leerlingvolgsysteem, waarin de resultaten en vorderingen binnen de schoolvakken van de leerlingen op de voet worden gevolgd (Ministerie van Algemene Zaken, 2021). Het meest gebruikte leerlingvolgsysteem is dat van het *Centraal Instituut voor ToetsOntwikkeling* (Cito; Van der Kleij & Eggen, 2013). Ieder half jaar worden gestandaardiseerde toetsen afgenomen bij de leerlingen om hun groei en ontwikkeling in de gaten te houden.

Testscores gaan altijd gepaard met onzekerheid (Padilla et al., 2020), zo ook die van Cito. Deze onzekerheid kunnen we uitdrukken in meetfouten en worden veroorzaakt door factoren die los staan van de vaardigheid van de leerling, maar wel de testscore beïnvloeden (Lichtenstein, 2020). Zo kan een leerling slecht hebben geslapen of is er veel onrust in de klas tijdens de toets. De mate waarin externe factoren van invloed zijn op testcores kunnen we uitdrukken als de betrouwbaarheid van een test (Gardner, 2013). De invloed van deze externe factoren vallen samen met een veelgenoemde kritiek op Cito-toetsen, namelijk dat het een momentopname is. Een leerling heeft soms een betere of mindere dag of heeft bijvoorbeeld veel stress van toetsen wat voor een afwijkend resultaat kan zorgen (Schuwirth, 2007). Om de mogelijkheid op deze meetfouten inzichtelijk te maken gebruikt Cito betrouwbaarheidsintervallen, naast de puntscore wordt het 67%-betrouwbaarheidsinterval (BHI) gegeven (Hop et al., 2016). Hiermee wordt getoond dat het werkelijke niveau van de leerling kan afwijken van de geobserveerde score. Het BHI geeft een scorebereik rondom de geobserveerde score waarin de werkelijke score van de leerling zich met 67% zekerheid bevindt (Ensie, 2016). Op deze manier kun je rekening houden met de onzekerheid van de score die voortkomt uit de aanwezigheid van meetfouten (Curran-Everett, 2009). Dit geeft een betere weergave van het werkelijke niveau (Charter & Feldt, 2002), je kunt namelijk met meer zekerheid zeggen dat het werkelijke niveau van de leerling in het interval ligt.

Uit onderzoek blijkt echter dat leerkrachten moeite hebben met het interpreteren van scorerapportages en ook minimaal gebruik maken van de informatie die de rapportages bevatten (Meijer et al., 2011; Ledoux et al., 2009). Ook lijkt het toevoegen van het betrouwbaarheidsinterval in Citorapportages niet het beoogde effect te hebben (Van der Kleij

& Eggen, 2013). Leerkrachten geven aan het interval niet te gebruiken en het concept van het betrouwbaarheidsinterval is hen niet bekend (Gardner, 2013; Zenisky & Hambleton, 2012). De keuze van Cito om de betrouwbaarheidsintervallen te gebruiken in de scorerapportages draagt dus niet bij aan de bekwaamheid van leerkrachten om de onzekerheid van scores te interpreteren. Uit verschillende studies blijkt ook dat leerkrachten niet de benodigde kennis en vaardigheden hebben om data te interpreteren en beslissingen te maken op basis van die data (Earl & Fullan, 2003; Verhaege, 2011).

In de huidige Citorapportages ligt de focus op de groei van de leerlingen (Hop et al., 2016). Het interpreteren van onzekerheid rondom groei is voor leerkrachten vaak een nog grotere uitdaging (Van der Kleij & Eggen, 2013). In de visualisaties van de groei wordt helemaal geen rekening gehouden met de onzekerheid van scores. Dit is problematisch, aangezien het vergelijken van twee scores met meetfouten meer onzekerheid met zich mee brengt dan bij een individuele score (Grimm et al., 2021). Door de onzekerheid is het namelijk niet mogelijk om de groei volledig toe te schrijven aan de vooruitgang in het niveau van de leerling. De waargenomen groei zou (deels) toegeschreven kunnen worden aan meetfouten in de testcores (Lichtenstein, 2021). Leerkrachten kunnen worden ondersteund in hun begrip van de onzekerheid door de onzekerheid te visualiseren, maar dat is in de Citorapportages nu dus niet het geval.

Bij het interpreteren van de toetsscores is het juist belangrijk om de onzekerheid goed te begrijpen en te interpreteren. Een onjuiste interpretatie van de toetsscores kan leiden tot inadequate besluiten en inadequaat handelen (Hopster-den Otter et al., 2019; Newton, 2005). In de onderwijssetting zou dit kunnen betekenen dat leerproblemen niet tijdig worden gesignaleerd, waardoor de leerling niet de ondersteuning krijgt die hij nodig heeft (Van der Kleij & Eggen, 2013).

Onzekerheid is een abstract concept, het visualiseren van deze onzekerheid kan het begrip vergroten (Padilla et al., 2020). Zoals eerder benoemd gebruikt Cito een puntschatting met een geschreven betrouwbaarheidsinterval in de scorerapportages. Deze visualisatie draagt niet veel bij aan het begrip en de interpretatie van onzekerheid (Belia et al., 2005; Correll & Gleicher, 2014). De interpretatie van een betrouwbaarheidsinterval is ook afhankelijk van de waarden van het interval, waardoor de accuratesse van de interpretaties niet constant is (Fernandes et al., 2018). Naast deze visualisaties is er echter nog legio aan alternatieven (Padilla et al., 2020).

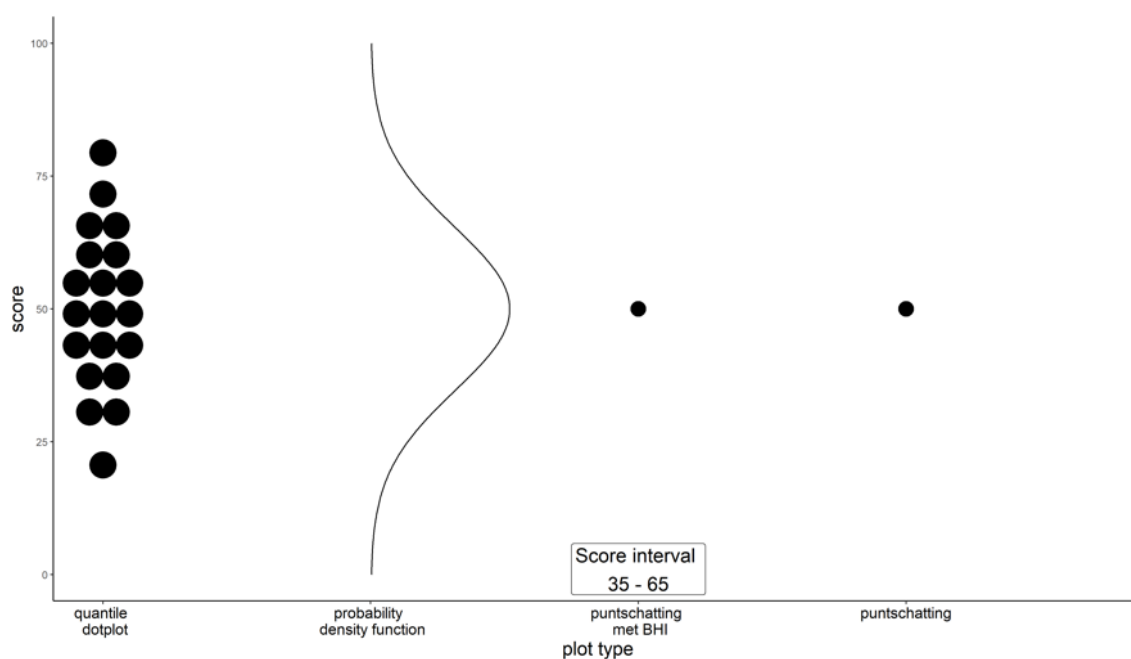
Een eerste alternatieve visualisatie is de *Quantile Dot Plot* (QDP). Deze visualisaties geeft met bolletjes de waarschijnlijkheid van een score weer. Hoe meer bolletjes, des te waarschijnlijker is die score (zie Figuur 1). Uit onderzoek blijkt dat visualisaties op basis van

frequentie kunnen zorgen voor betere kans inschattingen, dit soort visualisaties wordt intuïtiever begrepen (Padilla et al., 2020). De QDP is hier een voorbeeld van (Fernandes et al., 2018; Kay et al., 2016).

Een tweede alternatief voor de visualisatie van onzekerheid is de *Probability Density Function* (PDF) (Padilla, et al. 2020). De PDF is een grafiek waarin de hoogte afhangt van de waarschijnlijkheid van de score. Hoe hoger de grafiek is, des te groter is de kans dat die score het niveau van de leerling weergeeft (zie Figuur 1). Er is nog geen onderzoek gedaan naar het gebruik van deze visualisatie in een onderwijssetting. Wel is dit een veelvoorkomende visualisatie die intuïtief wordt begrepen in interpretaties (Padilla et al., 2020). In andere contexten laat de PDF vergelijkbare resultaten zien met de QDP (Fernandes et al, 2018).

Figuur 1

Voorbeelden van de vier visualisaties van meetnauwkeurigheid.



Aangezien het onvermijdelijk is dat testcores in de onderwijssetting gepaard gaan met onzekerheid en meetfouten, is het van belang dat de leerkrachten zo goed mogelijk worden ondersteund in hun begrip en interpretatie van die onzekerheden (Van der Kleij & Eggen, 2013). Om volledig te begrijpen wat de onzekerheid inhoudt, is een bepaalde statistische kennis van de leerkrachten vereist (Belia et al., 2009). Het aanbieden van statistiek op de lerarenopleiding zou een manier zijn om de bekwaamheid van de leerkrachten te vergroten. Statistiek is echter nog geen onderdeel van het curriculum op de pabo (Van Casteren et al.,

2018). De academische pabo, die in 2008 is opgericht, biedt een combinatie van de reguliere pabo en de universitaire opleiding pedagogische wetenschappen. In deze opleiding is de nadruk op vakken als statistiek en testtheorie veel sterker. Of deze vakken ook daadwerkelijk bijdragen aan een beter begrip van de onzekerheid van de testcores is nog niet onderzocht.

In het huidige onderzoek wordt gekeken naar de verschillen in het begrip van onzekerheid van leerkrachten en de juistheid van hun interpretaties op basis van verschillende visualisaties. Er zullen vier verschillende visualisaties vergeleken worden. De eerste visualisatie is de puntschatting. Ten tweede de visualisatie zoals die op dit moment is in cito-rapportages: een puntscore met een geschreven betrouwbaarheidsinterval. Ten derde de *Quantile Dot Plot* (QDP) en tot slot de *Probability Density Function* (PDF). De verwachting is dat de PDF en de QDP gepaard gaan met accuratere interpretaties door leerkrachten dan de puntschatting met en zonder betrouwbaarheidsinterval. De onderzoeksvraag die hierbij hoort is:

In hoeverre bestaat er een relatie tussen verschillende visualisaties van onzekerheid in testcores en de bekwaamheid van leerkrachten om de werkelijke scores te kunnen inschatten?

Vervolgens wordt de statistische kennis van leerkrachten in verband gebracht met hun bekwaamheid in het interpreteren van de onzekerheid van testcores. Hiermee wordt onderzocht of het hebben van een statistische basiskennis bevorderlijk is voor de interpretatie van onzekerheid. De verwachting is dat naarmate de statistische kennis van de participanten toeneemt, de inschattingen van de werkelijke scores accurater zullen zijn. De bijbehorende onderzoeksvraag luidt:

In hoeverre bestaat er een relatie tussen de statistische kennis van leerkrachten en hun bekwaamheid om de werkelijke scores in te kunnen schatten?

Ten derde wordt onderzocht of het type opleiding dat de participanten hebben gevolgd invloed heeft op de voorgenoemde relaties. Er wordt verwacht dat participanten die de academische pabo hebben gevolgd een betere statistische kennis hebben en hierdoor ook accuratere kansinschattingen maken dan participanten van de reguliere pabo.

Methode

Design

Om antwoord te geven op de gestelde onderzoeksvragen is een cross-sectioneel kwantitatief onderzoek uitgevoerd. Het design van het onderzoek is een experimenteel between-subject design. Er zijn vier visualisaties van onzekerheid onderzocht. Naast een situatie waarin geen onzekerheid wordt weergegeven zijn de Quantile Dot Plot, de Probability Density Function en

de puntscore met geschreven betrouwbaarheidsinterval onderzocht. De verschillende visualisaties zijn gerandomiseerd over de participanten, iedere participant heeft slechts één van de vier visualisatie te zien gekregen. De participanten hebben op basis van deze visualisaties kansinschattingen gemaakt over het niveau van de leerlingen.

Populatie en steekproef

De doelpopulatie van het onderzoek bestond uit huidige en toekomstige basisschoolleerkrachten in Nederland die bekend zijn met het leerlingvolgsysteem van Cito. Dit onderzoek heeft gebruikgemaakt van een gelegenheidssteekproef van basisschoolleerkrachten en (academische) pabo-studenten vanaf het derde leerjaar. Hier is voor gekozen, omdat de studenten in het derde leerjaar voor het eerst leren werken met Cito en ze vanaf dan worden verwacht de uitslagen juist te kunnen interpreteren. Participanten die aangaven niet bekend te zijn met Cito zijn uit de vragenlijst geleid.

Volgens Cohen (1992) zijn bij een medium effect minimaal 45 participanten per subgroep nodig om een power van ongeveer .80 te bereiken, bij een groot effect zijn dit 18 participanten. In dit onderzoek zijn er twaalf subgroepen: vier visualisaties en drie groepen participanten (basisschoolleerkrachten, pabo-studenten en studenten van de academische pabo). Het streven van dit onderzoek was daarom om een steekproef van 216 tot 540 participanten te verzamelen.

Instrumenten

Om de bekwaamheid van de participanten in het inschatten van de onzekerheid te meten is de Kans Inschatting Werkelijke Score (KIWS) schaal ingezet (Ettema, 2021). De participanten kregen vijf plots te zien, waarin een score van een leerling al dan niet gepaard ging met een visualisatie van onzekerheid. Per plot werden twee vragen gesteld, in totaal bestaat deze schaal dus uit tien vragen (zie Bijlage A). De vragen waren “*Wat is de kans dat het werkelijke niveau van deze leerling **lager/hoger** is dan niveau X?*” En “*Wat is de kans dat het **werkelijke niveau** van deze leerling niveau X is?*”. De participanten gaven antwoord op een schaal van 1 tot 100. De plots zijn gebaseerd op de scorerapporten van Cito in kleur en vormgeving, op de plaats van de X is daarom een score van I t/m V ingevuld. Deze schaal laat in eerder onderzoek een betrouwbaarheid van .90 op Guttman’s Lamda² zien (Ettema, 2021). Voor het huidige onderzoek is de betrouwbaarheid opnieuw vastgesteld.

Daarnaast is de statistische kennis van de participanten getest aan de hand van *de Levels of Conceptual Understanding in Statistics* (LOCUS) vragenlijst (Whitaker, et al., 2015). Deze vragenlijst is ontwikkeld om de statistische kennis van studenten in opleiding tot (wiskundig) docent te meten. De vragenlijst bestaat uit vier subschalen: formuleren van vragen, dataverzameling, data-analyse en interpretatie van resultaten. Voor dit onderzoek is de

subschaal van data-analyse gebruikt, aangezien dit onderdeel zich bezig houdt met het begrijpen van grafieken en het duiden van scores. De vragen hadden elk vier antwoordopties, waarvan één antwoord het juiste was. Een hogere score op deze vragenlijst betekent dat de participant meer statistische kennis heeft. Voor de data-analyse schaal is in het pilot-onderzoek een betrouwbaarheid van .87 gevonden (Whitaker, et al., 2015). Voor het huidige onderzoek zijn de vragen vertaald naar het Nederlands, waarna de betrouwbaarheid opnieuw is bepaald.

Naast deze twee instrumenten zijn de sociaal-demografische kenmerken van de participanten bevraagd aan de hand van drie vragen. Ten eerste gaven de participanten aan welke opleiding zij hebben genoten, de reguliere pabo, de academische pabo of een vrij in te vullen alternatief. Ten tweede werd gevraagd of de participant nog bezig is met de opleiding, en in welk studiejaar de participant zich dan bevindt, of dat de opleiding al is afgerond. Op basis van de eerste twee vragen zijn de participanten ingedeeld in de drie subgroepen: basisschoolleerkrachten, pabostudenten en studenten van de academische pabo. Ten derde werd gevraagd in welke provincie de participant werkt of studeert. Hiermee is de geografische spreiding van de steekproef gecontroleerd.

Procedure

Voor dit onderzoek is een online Qualtrics vragenlijst gebruikt. De verspreiding van de vragenlijst is gebeurd aan de hand van e-mail en sociale media. In eerste instantie werden de persoonlijke contacten van de onderzoekers benaderd. Ook werden scholen door Nederland heen benaderd met de vraag of zij mee wilden werken. De contactgegevens van deze scholen zijn verkregen uit het bestand van basisscholen van de Dienst Uitvoering Onderwijs (DUO, 2021). Vervolgens werd gebruikgemaakt van snowball sampling, om het bereik van het onderzoek te vergroten.

Voorafgaand aan deelname aan het onderzoek werden de participanten geïnformeerd over het doel van het onderzoek: het onderzoeken van visualisaties van onzekerheid in Cito-testscores. Er werd aangegeven dat de vragenlijst ongeveer 15 minuten in beslag zou nemen, uiteindelijk waren de participanten gemiddeld 13 minuten kwijt. Vervolgens werd de anonimiteit benadrukt. Gegevens die terug geleid zouden kunnen worden naar een participanten, zoals een IP-adres, zijn niet opgeslagen. Ook werd aangegeven dat deelname vrijwillig is en dat het de participant vrij stond om op ieder moment te stoppen met de vragenlijst. Tot slot werd verteld dat de data van het onderzoek vijf jaar lang worden opgeslagen op een beveiligde server van de Rijksuniversiteit Groningen. Na deze informatie werd de participant gevraagd om schriftelijke toestemming voor het gebruik van de data binnen dit

onderzoek. Voorafgaand aan de dataverzameling heeft de ethische commissie van Pedagogische en Onderwijswetenschappen toestemming verleend.

Het eerste deel van de vragenlijst betrof de algemene vragen. Vervolgens kwamen de KIWS vragen aan bod. En tot slot de LOCUS vragen. De respondenten kregen gerandomiseerd één van de visualisaties toebedeeld. De plots zijn gemaakt naar voorbeeld van de huidige Cito rapportages in kleur en vormgeving. De getoonde scores in de plots variëren van 60 tot 75. De scores liggen ieder in een grensgebied tussen twee scorecategorieën, waarbij de visualisaties van onzekerheid door verschillende categorieën lopen. In het onderzoek van Ettema (2021) zijn de verschillende type vragen van de KIWS (hoger/lager en werkelijk niveau) als twee subschalen beschouwt. Er is geen verschil gevonden tussen de schalen. In het huidige onderzoek zijn echter enkele veranderingen gemaakt aan de vragenlijst, bijvoorbeeld de kleurstelling en de geobserveerde scores. De subschalen worden daarom beide meegenomen in dit onderzoek en opnieuw met elkaar vergeleken.

Data-analyse

De verzamelde data zijn verwerkt in het computerprogramma SPSS (versie 27). Voorafgaand aan de analyses zijn problemen in dataset verholpen, zo werden de participanten die vroegtijdig zijn gestopt en daarom niet de volledige KIWS schaal hadden ingevuld verwijderd. Participanten die na de KIWS schaal zijn gestopt zijn wel meegenomen. Aansluitend is de steekproef aan de hand van frequentietabellen beschreven op basis van de bevraagde sociaal-demografische kenmerken.

Vervolgens zijn de kansinschattingen op de KIWS-schaal geanalyseerd. De scores op de KIWS zijn op twee manieren berekend. Ten eerste is het gemiddelde van de absolute verschillen tussen de ingeschatte kans en de daadwerkelijke kans op de tien vragen genomen ($KIWS_{abs}$). Deze score geeft aan hoe veel procentpunten de participant gemiddeld afwijkt van het juiste antwoord. Deze score is gebruikt voor de beschrijvende resultaten, aangezien de scores op deze manier praktisch te interpreteren zijn. Daaropvolgend zijn de scores een logistische transformatie ondergaan. Hiervoor is gekozen, omdat de maximaal mogelijke verschillen tussen de vragen sterk verschillen. Een vraag waarop het juiste antwoord 8% is, kan bijvoorbeeld een verschillen van maximaal 92 opleveren, terwijl dit bij een antwoord van rond de 50% veel kleiner is. De transformatie die is uitgevoerd is: $KIWS_{log} = \ln(KIWS_{abs} + .005) - \ln(.005)$. Deze score is gebruikt in de regressieanalyses en *t*-toetsen.

Voor de LOCUS vragenlijst is per goed antwoord een punt toegekend, het aantal juiste antwoorden is bij elkaar opgeteld om tot een totaalscore te komen. De participanten konden een score van 0 tot 5 halen op deze schaal. Verder is voor zowel de KIWS-schaal als de LOCUS

de betrouwbaarheid geschat. Hiervoor is Guttman's Lamda² gebruikt en is een item-analyse uitgevoerd.

Om de relatie tussen de visualisaties en de kansinschattingen te onderzoeken is in eerste instantie gebruikgemaakt van boxplots. Hiervoor zijn de gemiddelde absolute verschillcores op de KIWS per visualisatie uitgesplitst. Vervolgens is een regressieanalyse uitgevoerd. Voor de KIWS is hiervoor de getransformeerde variabele gebruikt ($KIWS_{log}$). Voor de visualisaties zijn dummy-variabelen aangemaakt, waarbij de puntscore de referentiecategorie is.

Ook voor de LOCUS score zijn dummy-variabelen opgesteld, met als referentiecategorie 0 goed. Deze variabele is in het tweede regressiemodel toegevoegd. Het derde regressiemodel neemt naast de visualisatie en de LOCUS de opleiding van de participant en de afrondingsstatus van die opleiding mee in de regressie.

Voorafgaand aan de analyses zijn de assumpties van normaliteit en homoscedasticiteit gecheckt. Daarnaast zijn de data ook gecontroleerd voor uitbijters, hieronder vallen alle scores waarvan het gestandaardiseerde residu <-2 of >2 is.

Resultaten

Problemen in de dataset

In totaal zijn er 373 participanten begonnen aan de vragenlijst. Van deze participanten vallen 12 niet in de doelgroep. Vier participanten gingen niet akkoord met de verwerking van de gegevens, vier participanten waren niet bekend met Cito en vier participanten zaten nog in het eerste jaar van de opleiding tot leerkracht. Deze participanten zijn uit de vragenlijst geleid.

Verder hebben 15 participanten anders, namelijk ingevuld bij de vraag welke opleiding zij hebben gevolgd. Hiervan zijn 5 participanten die de KLOS hebben gevolgd en 1 participant die onderwijswetenschappen heeft gedaan. Deze mensen zijn wel meegenomen in het onderzoek, maar bij opleiding worden deze waarden als missing weergegeven. De overige 9 participanten zijn op basis van hun antwoord ingedeeld in één van de subgroepen. Participanten die een voorganger of vergelijkbare opleiding van de pabo hebben gevolgd, zijn in de groep pabo geplaatst. Dit betreft 7 participanten. Tot slot waren er 2 participanten die een academische vervolgopleiding hebben gevolgd na hun pabo opleiding. Deze participanten zijn in de groep academische pabo geplaatst.

Er zijn in totaal 159 mensen vroegtijdig gestopt met de vragenlijst, deze participanten zijn verwijderd uit de dataset. Binnen deze groep zijn 23 participanten gestopt na de algemene vragen, 98 participanten zijn na de uitleg van de KIWS opgehouden en 38 participanten zijn gestopt tijdens de vragen van de KIWS. Van de groepen participanten die in ieder geval de start

van de KIWS heeft gezien ($N = 136$), hebben 33 de visualisatie puntschatting gekregen, 29 de puntschatting met het BHI, 42 de PDF en 32 de QDP. De verdeling van deze plottypes is gecontroleerd met een Chi-kwadraattoets. Uit de toets blijkt geen significant verschil in de proporties: $X^2(3) = 2.765$, $p = .429$. Er is dus geen reden om aan te nemen dat een bepaalde visualisatie leidt tot meer uitval.

Steekproef

Er zijn 194 participanten die in ieder geval de KIWS vragen volledig hebben ingevuld. Hiervan hebben 153 (78,9%) participanten de volledige vragenlijst afgemaakt en zijn 41 (21,1%) gestopt voor of tijdens de LOCUS vragen.

De vragenlijst is ingevuld door 174 leerkrachten (89,7%) en 20 studenten (10,3%). Van de groep studenten volgen 12 participanten de academische pabo (60%) en 8 de reguliere pabo (40%). De groep leerkrachten bestaat uit 158 participanten (90,8%) die de pabo hebben afgerond en 16 participanten (9,2%) die de academische pabo hebben gevolgd. Het grootste deel van de participanten heeft de pabo als opleiding gedaan ($n = 188$, 86,2%), tegenover een groep van 26 participanten (13,8%) die de academische pabo hebben gevolgd.

De meeste participanten zijn werkzaam in Zuid-Holland ($n = 44$, 22,7%) en Drenthe ($n = 31$, 16,0%). De provincies met de minste representatie zijn Zeeland (2, 1,0%) en Flevoland (2, 0,5%). De rest van de provincies laten een spreiding van 10 tot 22 participanten zien.

In Tabel 1 is een overzicht te zien van de verdeling van de plottypes die de participanten te zien hebben gekregen. De *Probability Density Function* (PDF) komt duidelijk het minst vaak voor. Voor de verdeling van de plottypes is een Chi-kwadraattoets uitgevoerd. De verschillen in proporties zijn niet significant: $\chi^2(3) = 2.907$, $p = .406$.

De tabel toont ook een overzicht van de plottypes per subgroep. Hierin valt op dat binnen de subgroep studenten de *puntschatting met BHI* (PS+BHI) slecht 3 keer voorkomt. Dit valt deels te verklaren door de participanten die zijn gestopt met de vragenlijst.

Tabel 1

Verdeling plottypes

		Totaal	Leerkracht	Student
Plotype	PS	54 (27,8%)	47 (27,0%)	7 (35%)
	PS + BHI	53 (27,3%)	50 (28,7%)	3 (15%)
	PDF	39 (20,1%)	35 (20,1%)	4 (20%)
	QDP	48 (24,7%)	42 (24,1%)	6 (30%)
Totaal		194 (100%)	174 (100%)	20 (100%)

Analyse meetinstrumenten

De KIWS vragenlijst geeft een Guttman's Lambda2 waarde van .75. In het onderzoek van Ettema (2021) zijn geen verschillen gevonden tussen de subschalen. Aangezien de vragenlijst enkele veranderingen heeft ondergaan voor het huidige onderzoek, is dit opnieuw gecontroleerd. Wederom was er geen verschil te zien tussen de twee schalen.

In Tabel 3 zijn de juiste scores, gemiddelde scores, gemiddelde absolute afwijking en item-restcorrelaties van de items weergegeven. Bij de gemiddelde absolute afwijking van de participanten valt op dat de eerste vraag per plot wordt overschat, terwijl de tweede vraag juist meestal wordt onderschat. De eerste vraag betreft de hoger/lager vraag, waar de tweede vraag het werkelijke niveau vraagt. Bij de gemiddelde absolute afwijking is echter geen groot verschil te zien tussen het type vragen. Vraag 1B lijkt het minst goed gemaakt te zijn, met de grootste gemiddelde absolute afwijking ($M = 22.46$, $SD = 13.92$). Vraag 4B is juist het best gemaakt ($M = 14.83$, $SD = 11.42$). Er zijn geen grote verschillen in moeilijkheid van de vragen. De item-rest correlaties van vragen 2B en 5B zijn erg laag. Dit lijkt het gevolg te zijn van enkele uitschieters die op deze vraag veel beter of slechter scoren dan op de rest van de vragen.

Tabel 2

Item-analyse KIWS

Item	Juiste antwoord	Gemiddelde afwijking	<i>SD</i>	Gem. Absolute afwijking	<i>SD</i>	Item-rest correlatie
1 hoger/lager	8	16.62	18.65	17.79	17.53	.26
1 between	39	17.05	20.21	22.46	13.92	.29
2 hoger/lager	40	3.09	20.74	16.24	13.22	.34
2 between	55	-0.61	19.38	15.46	11.65	.02
3 hoger/lager	39	5.89	18.45	15.20	11.96	.24
3 between	58	-5.47	18.64	14.95	12.37	.26
4 hoger/lager	39	0.89	21.70	17.17	13.25	.30
4 between	59	-0.62	18.74	14.83	11.42	.28
5 hoger/lager	38	4.92	20.74	17.19	12.56	.24
5 between	62	-7.52	18.71	16.05	12.17	.15
	Totaal			16.73	13.01	-

In Tabel 4 zijn de gemiddelden, standaarddeviaties en item-restcorrelaties van de LOCUS vragen weergegeven. De LOCUS vragen laten een goede spreiding in moeilijkheid zien. De best gemaakte vraag is door 66% van de participanten goed beantwoord. Voor de minst goed

gemaakte vraag is dit 41%. De scores zijn redelijk normaal verdeeld, waarbij de grootste groep participanten 3 vragen goed heeft beantwoord (22.7%).

De vragenlijst geeft een betrouwbaarheid van .27 op Guttman's Lamda2. Deze score is erg laag. De eerste vraag van de vragenlijst lijkt niet goed samen te hangen met de rest van de schaal, het item correleert negatief met de andere vragen (item-restcorrelatie = -.10). Deze vraag is daarom niet meegenomen in de schaal. Na de verwijdering van dit item geeft Guttman's Lamda2 een waarde van .38. De gemiddelde totaalscore van de vier overgebleven vragen is 2.14 ($SD = 1.15$).

Tabel 3

Analyse LOCUS

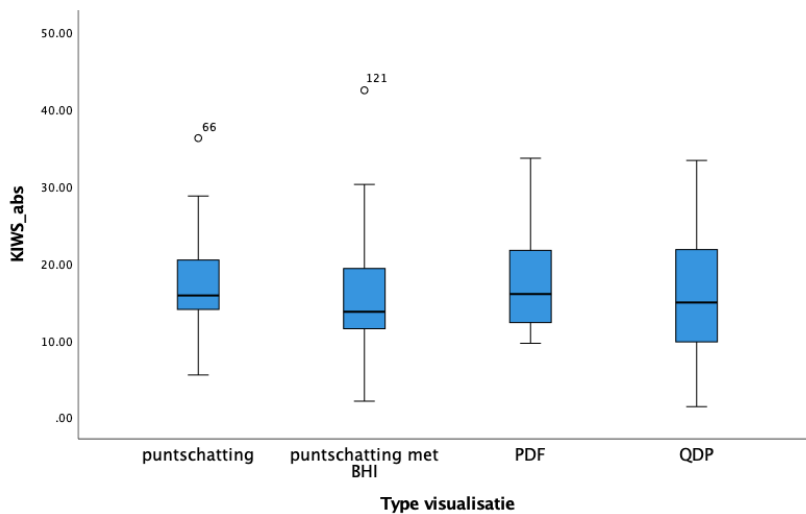
Item	Gemiddelde score	SD	Item-rest correlaties
Vraag 1	.46	0.50	-.10
Vraag 2	.66	0.48	.21
Vraag 3	.41	0.49	.21
Vraag 4	.56	0.50	.15
Vraag 5	.51	0.50	.01
Totaal	2.60	1.20	-

Kans inschattingen

In Figuur 1 wordt de nauwkeurigheid van de kans inschattingen in gemiddelde absolute verschillen weergegeven voor de verschillende plottypes. Er lijken op het eerste oog geen grote verschillen te zijn tussen de visualisaties. De mediaan van alle visualisaties ligt rond de 16. De QDP laat de grootste spreiding zien. De minimumscore van deze visualisatie ligt erg dicht bij de 0, wat inhoudt dat een participant erg accurate kansinschattingen heeft gemaakt. Het minimum van de PDF ligt beduidend hoger dan dat van de andere visualisaties. De meest accurate score binnen deze visualisatie is 10. De puntschatting en de puntschatting met BHI laten allebei een uitbijter aan de bovenkant zien. De inschattingen van deze participanten wijken sterk af van de juiste antwoorden.

Figuur 2

Boxplots gemiddelde absolute verschilcores kans inschattingen per plotype.



Note. 0 is perfect accuraat op alle vragen.

Bivariate relaties

Vervolgens zijn de bivariate relaties onderzocht. Bij deze analyses is rekening gehouden met de assumpties van normaliteit en homoscedasticiteit. Hierin zijn geen afwijkende waarden gevonden, er is voldaan aan deze assumpties. Voor de toetsen met betrekking tot de kansinschattingen is de $KIWS_{\log}$ gebruikt.

Participanten van de academische pabo waren gemiddeld iets accurater in hun kansinschattingen ($M = 2.96$, $SD = 0.72$) dan participanten van de reguliere pabo ($M = 3.51$, $SD = 0.38$). Dit verschil was significant $t(186) = 5.85$, $p < .001$. Het verschil in KIWS scores tussen studenten en werkende leerkrachten is ook significant bevonden: $t(186) = 2.32$, $p < .02$. De kansinschattingen van de studenten ($M = 3.21$, $SD = 0.76$) zijn gemiddeld accurater dan die van werkenden ($M = 3.47$, $SD = 0.44$).

De LOCUS laat in verband met de opleiding van de participanten een significant verband zien; $t(144) = -6.09$, $p < .001$. Participanten van de academische pabo ($M = 3.33$, $SD = 0.64$) halen gemiddeld een hogere score op de vragenlijst dan participanten van de reguliere pabo ($M = 1.95$, $SD = 1.07$). Verder is het verschil tussen studenten en werkenden een significant verschil gevonden: $t(148) = -2.33$, $p = .02$. Studenten scoorden gemiddeld 2.72 ($SD = 0.28$) punten, terwijl leerkrachten een gemiddelde van 2.06 ($SD = 0.10$) laten zien.

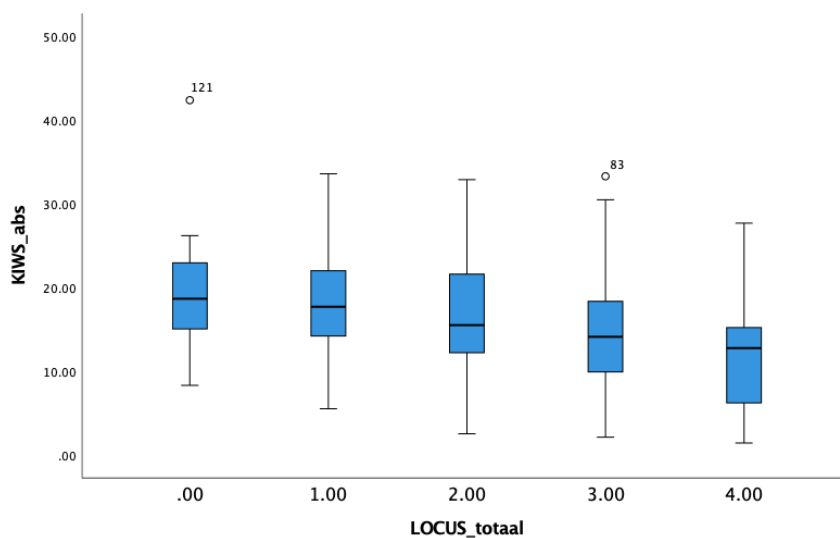
De variabelen opleiding en opleiding afgerond laten een sterke associatie met elkaar zien $X^2(1) = 34.16$, $p < .001$. Er is echter niet aan de assumpties voor een Chi-kwadraattoets voldaan, aangezien de verwachte cel-waarde niet overal boven de 5 was. Daarom is ook een Fisher's

exact test uitgevoerd, deze test is significant ($p < .001$). De variabelen zijn dus niet onafhankelijk van elkaar.

Figuur 3 laat de relatie tussen de gemiddelde verschillcores op de KIWS vragenlijst en de score op de LOCUS zien. Er lijkt een trend te zien te zijn, waarin een hogere score op de LOCUS gepaard gaat met accuratere kansinschattingen. De spreiding varieert tussen de verschillende scores. Bij 0 goed is de spreiding veel kleiner dan bijvoorbeeld bij 2 goed. Verder zijn er twee uitbijters te zien, één bij 0 goed en één bij 3 goed. De kansinschattingen van deze participanten wijken ver af van de goede antwoorden.

Figuur 3

Boxplots gemiddelde verschillcores KIWS per score op de LOCUS.



Regressieanalyses

In Tabel 3 zijn de coëfficiënten, standaard fouten, significanties en verklaarde varianties van regressie-Model 1, -Model 2 en Model 3 te zien. Het eerste model brengt de visualisaties in verband met de kans inschattingen. In het tweede model is de statistische kennis toegevoegd. Het derde model voegt de covariaten opleiding en opleiding afgerond toe. Voor alle regressieanalyses zijn de assumpties van normaliteit en homoscedasticiteit. Ook is gecontroleerd op invloedrijke observaties. Aan deze assumpties is voldaan.

Tabel 4

Regressiemodellen met als afhankelijke variabele $KIWS_{log}$ en in model 1 visualisatie, in model 2 visualisatie en statistische kennis en model 3 visualisatie, statistische kennis, opleiding en opleiding afgerond als onafhankelijke variabelen.

	Model 1				Model 2				Model 3			
	R ²	<i>b</i>	<i>SE</i>	<i>p</i>	R ²	<i>b</i>	<i>SE</i>	<i>p</i>	R ²	<i>b</i>	<i>SE</i>	<i>p</i>
	.03				.12				.22			
Intercept		3.49	0.08	<.001		3.60	0.17	<.001		4.12	0.23	<.001
Visualisatie [puntschatting]												
BHI		-0.09	0.12	.43		0.05	0.11	.68		-0.09	0.11	.41
PDF		-0.02	0.13	.85		0.01	0.13	.96		-0.04	0.12	.77
QDP		-0.22	0.12	.07		-0.14	0.12	.23		-0.20	0.11	.07
Statistische kennis [0 goed]												
1 goed		-	-	-		0.01	0.18	.97		-0.00	0.17	.99
2 goed		-	-	-		-0.05	0.17	.76		-0.03	0.16	.86
3 goed		-	-	-		-0.23	0.17	.18		-0.09	0.17	.57
4 goed		-	-	-		-0.48	0.19	.01		-0.19	0.20	.34
Opleiding [academisch]		-	-	-		-	-			-0.52	0.14	<.001
Opleiding afgerond [student]		-	-	-		-	-			-0.03	0.14	.81

Het eerste regressiemodel, met alleen visualisatietype, geeft een proportie verklaarde variantie van slechts 3%. De waarden van de visualisatie (BHI, PDF en QDP) worden vergeleken met de waarde van de puntschatting (het intercept). In het eerste model verschillen de kansinschattingen van de visualisaties niet significant met die van de puntschatting (intercept). De coëfficiënten van de visualisaties laten negatieve waarden zien. Dit zou betekenen dat de visualisaties gepaard gaan met meer accurate kansinschattingen. De QDP laat het grootste verschil zien ten opzichte van de puntschatting. Dit verschil is echter dus niet significant ($p = .07$).

In het tweede model wordt naast de visualisatie ook de statistische kennis meegenomen. Dit model geeft een proportie verklaarde variantie van 12%, dit is een sterke toename ten opzichte van het eerste model. De statistische kennis scores 1 tot en met 3 laten geen significant verschil zien ten opzichte van het intercept (0 goed). Het verschil tussen een score van 4 goed is wel significant bevonden ten opzichte van 0 goed ($p = .01$). Ook lijkt er over alle scores een trend te zijn; een hogere score op de LOCUS gaat gepaard met een lagere score op de KIWS en een lagere p -waarde. De enige score die niet binnen deze trend valt is 1 goed. Participanten met deze score laten iets minder accurate kansinschattingen zien dan participanten met 0 goed. De visualisaties laten hogere p -waarden zien dan in het eerste model, de verschillen met de puntschatting zijn ook minder groot. Er is wederom geen sprake van significante verschillen met de puntschatting.

In het derde model worden naast visualisatie en statistische kennis de covariaten opleiding en opleiding afgerond meegenomen. Dit model geeft een proportie verklaarde variantie van 22%, een flinke toename van de voorgaande modellen. De variabele opleiding laat een significant verband zien ($p < .001$). De bijbehorende coëfficiënt is negatief. Dit houdt in dat participanten van de academische pabo accuratere kansinschattingen hebben gemaakt. De variabele opleiding afgerond geeft geen significant verschil. De p -waarden van de visualisaties zijn in dit model lager dan in model 1 en 2, maar zijn wederom niet significant. De scores op de LOCUS laten in dit model weer een aflopende trend zien, de coëfficiënten zijn in model 3 wel kleiner. De p -waarden liggen hoger dan in model 2.

Het derde model lijkt het best passend. Aangezien de covariaat opleiding een significante relatie laat zien. Daarnaast is de proportie verklaarde variantie sterk toegenomen na de toevoeging van de covariaten.

Discussie

In dit onderzoek zijn vier visualisatievormen (puntschatting, puntschatting met betrouwbaarheidsinterval, *Probability Density Function* en *Quantile Dot Plot*) vergeleken in hun bijdrage aan de bekwaamheid van leerkrachten om toetscores te interpreteren. Vervolgens is de statistische kennis van de participanten in verband gebracht met de bekwaamheid van de leerkrachten. Tot slot was er aandacht voor de verschillen tussen participanten die de reguliere, dan wel academische pabo hebben gevolgd.

De resultaten tonen geen significant verschil tussen de kansinschattingen van participanten met de verschillende visualisatie van onzekerheid en alleen de puntschatting. Participanten lijken gemiddeld de meest accurate inschattingen te maken bij de QDP. Dit komt overeen met de verwachting dat visualisaties op basis van frequentie leiden tot betere kansinschattingen (Fernandes et al., 2018). De PDF laat nauwelijks verschil zien met de puntschatting. In het tweede model zijn de kansinschattingen van de participanten met de PDF zelf minder accuraat dan die van participanten met de puntschatting. Dit is in strijd met de verwachtingen (Fernandes et al., 2018). Dit zou het gevolg kunnen zijn van de context. Deze visualisatie is niet eerder onderzocht in de onderwijssetting. Deze doelgroep is wellicht minder bekend met deze visualisatie.

De puntschatting met het BHI laat ook geen betere resultaten zien dan de puntschatting zonder onzekerheid. De verwachtingen over deze visualisatie waren gemixt. Enerzijds zou het begrip van de onzekerheid vergroten (Curran-Everett, 2009), anderzijds zouden leerkrachten moeite hebben met de interpretatie (Van der Kleij & Eggen, 2013). Voor de doelgroep en de setting lijkt deze visualisatie dus niet geschikt te zijn.

In de relatie tussen de kansinschattingen en de statistische kennis van de participanten is in het tweede model één significant verschil gevonden. Participanten die alle vier de vragen van de LOCUS goed hebben beantwoord laten significant accuratere kansinschattingen zien dan participanten die geen vragen goed hadden. Over alle scores is een duidelijke trend zichtbaar, waarbij meer statistische kennis gepaard gaat met accuratere kansinschattingen. Dit is in lijn met de verwachting dat enige statistische kennis vereist is om onzekerheid juist te kunnen interpreteren (Belia et al., 2009).

De opleiding die de participanten hebben gevolgd is significant bevonden. Participanten die de academische pabo hebben gedaan scoren hoger op de statistische kennis en maken accuratere kansinschattingen. Dit komt overeen met de verwachtingen. Deze opleiding biedt namelijk statistische vakken aan in het curriculum. Dit lijkt dus ook daadwerkelijk te leiden tot meer statistische kennis dan bij participanten die de reguliere pabo hebben gevolgd. Ook is het

in lijn met de verwachting dat meer statistische kennis leidt tot betere kansinschattingen (Belia et al., 2009).

Van de 373 participanten die aan de vragenlijst zijn gestart waren slechts 194 bruikbaar voor het onderzoek. Dit heeft negatieve gevolgen voor de power van het onderzoek. Het streven was een steekproef van 216 tot 540 participanten. Daarnaast zou het afhaken van de participanten kunnen leiden tot bias. Er lijken echter geen verschillen in uitval te zijn tussen de verschillende subgroepen en type visualisaties. Door de randomisatie van de visualisaties over de participanten is er in de groep overgebleven participanten een goede spreiding te zien tussen de visualisaties. Bias is lijkt hierdoor geen groot probleem.

Wel is het opmerkelijk dat bijna de helft van de participanten vroegtijdig is gestopt. Een mogelijke verklaring is de moeilijkheid van de vragenlijst. Participanten moesten eerst een redelijk lange uitleg lezen, om vervolgens in een onbekende situatie daar vragen over te beantwoorden. Ook kunnen de statistische vragen als frustrerend worden ervaren, bijna alle participanten hebben hier fouten in gemaakt. Er zijn dan ook veel participanten afgehaakt voor of tijdens de statistische kennis vragen. Deze respondenten konden echter nog wel worden meegenomen in de analyses voor de kansinschattingen. Mochten de statistische kennis vragen eerder in de vragenlijst hebben gezeten, waren deze respondenten wellicht ook niet bruikbaar geweest voor de analyses van de kansinschattingen. Daarnaast kan ook de lengte van de vragenlijst een rol hebben gespeeld. De motivatie voor het invullen van een vragenlijst neemt af, naarmate deze langer duurt. Dit effect is nog sterker bij online vragenlijsten (De Leeuw, 2012), wat in dit onderzoek ook het geval is.

Uit eerder onderzoek naar de LOCUS vragenlijst blijkt een hoge betrouwbaarheid en validiteit (Whitaker et al., 2015), dit is echter in het huidige onderzoek niet teruggevonden. Met name de eerste vraag past niet goed in de schaal. De lage item-restcorrelatie heeft ertoe geleid dat deze vraag is verwijderd uit de schaal. Eén van de foute antwoordopties is bijna net zo vaak gekozen als het juiste antwoord. Deze antwoordoptie gaf aan dat er te weinig informatie was om de plot goed af te kunnen lezen. Participanten met een goede kennisbasis geven eerder deze afleider als antwoord, wanneer ze niet helemaal zeker zijn. Terwijl participanten met mindere kennis deze optie juist als minder veilig beschouwen (Frary, 1991). Hierdoor kan deze vraag een vertekend beeld hebben gegeven van het niveau van de participanten. De betrouwbaarheid van de schaal is wellicht ook lager door de lengte van de vragenlijst. De vragen werden aan het einde van de vragenlijst gesteld, waardoor de focus en motivatie van de participanten wellicht was afgenomen. Daarnaast had de volgorde van de vragen gerandomiseerd kunnen worden. De lage betrouwbaarheid geeft aan dat de vragen niet geschikt zijn voor het meten van de

statistische kennis. De resultaten op basis van deze vragenlijst zijn dus wellicht niet juist. Voor vervolgonderzoek is het daarom aangeraden om een andere vragenlijst voor de statistische kennis te gebruiken. De KIWS vragenlijst is door zijn goede betrouwbaarheid wel aan te raden voor vervolgonderzoek naar kansinschattingen.

Voorafgaand aan de analyses van de KIWS zijn de scores een logistische regressie ondergaan. Hierdoor is de invloed van de verschillende vragen op de totaalscore gelijkwaardiger geworden. Wel is hierdoor de praktische interpretatie van de regressies bemoeilijkt. In de regressieanalyses zijn ook de covariaten opleiding en opleiding afgerond meegenomen. De subgroepen academische pabo en studenten zijn echter aan de kleine kant. De gevonden effecten van deze covariaten moeten daarom met voorzichtigheid worden geïnterpreteerd.

Zoals benoemd creëert de lengte van de vragenlijst op meerdere punten problemen voor het onderzoek. Voor vervolgonderzoek is het aan te raden om de vragenlijst in te korten. Een manier waarop dit mogelijk is, is door de KIWS aan te passen. Op het moment worden er per visualisatie twee vragen gesteld. Uit eerder onderzoek blijkt geen verschil tussen de twee subschalen, dat is in het huidige onderzoek bevestigd. In het vervolg kan er dus per visualisatie één van de subschalen worden gebruikt, waardoor de vragenlijst korter wordt.

Op basis van huidig onderzoek kunnen er meerdere aanbevelingen worden gedaan voor de praktijk. Ten eerste kunnen visualisaties uitmaken in de interpretatie van toetsscores. De huidige visualisatie van Cito is de puntschatting met het BHI. Het toevoegen van het betrouwbaarheidsinterval zou bijdragen aan het begrip van de onzekerheid rondom de toetsscores. Dit lijkt echter niet het geval. Er zijn aanwijzingen gevonden dat de visualisatie van de QDP hier geschikter voor zou zijn. De effectgrootte van deze visualisatie is niet erg groot (Pearson's r rond de .15). Er zijn wellicht visualisaties die een grotere praktische relevantie hebben.

Ten tweede is er een verband gevonden tussen de kansinschattingen door de participanten en hun statistische kennis. De participanten van de academische pabo laten meer statistische kennis zien en scoren ook beter op de KIWS. Het aangeboden krijgen van statistische vakken lijkt hier een logische verklaring voor. Door ook op de reguliere pabo statistische vakken op te nemen in het curriculum, kan de statistische kennis van alle (toekomstige) leerkrachten worden verbeterd. Wat op zijn beurt hopelijk leidt tot een betere interpretatie van de onzekerheid rondom toetsscores.

Concluderend zijn er in dit onderzoek aanwijzingen gevonden voor het belang van visualisaties en statistische kennis bij het interpreteren van de onzekerheid rondom toetsscores.

Visualisaties zoals de QDP ondersteunen leerkrachten beter in deze interpretaties dan de huidige visualisaties van Cito. Daarnaast is het van belang om (toekomstige) leerkrachten voldoende statistische kennis mee te geven om deze onzekerheid te kunnen begrijpen. Zodat leerkrachten in hun besluitvorming en handelen beter kunnen aansluiten bij de behoefte van de leerlingen en hen beter kunnen ondersteunen in hun leerproces.

Literatuur

- Belia, S., Fidler, F., Williams, J., & Cumming, G. (2005). Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods, 10*(4), 389–396. <https://doi.org/10.1037/1082-989x.10.4.389>
- Bich, W. (2012). From errors to probability density functions. Evolution of the concept of measurement uncertainty. *IEEE Transactions on Instrumentation and Measurement, 61*(8), 2153–2159. <https://doi.org/10.1109/tim.2012.2193696>
- Charter, R. A., & Feldt, L. S. (2002). The Importance of Reliability as It Relates to True Score Confidence Intervals. *Measurement and Evaluation in Counseling and Development, 35*(2), 104–112.
- Correll, M., & Gleicher, M. (2014). Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics, 20*(12), 2142–2151. <https://doi.org/10.1109/tvcg.2014.2346298>
- Curran-Everett, D. (2009). Explorations in Statistics: Confidence Intervals. *Advances Physiology Education, 33*(2), 87-90.
- Datnow, A., & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision making: A literature review of international research. *Journal of Educational Change, 17*(1), 7-28. <https://doi.org/10.1007/s10833-015-9264-2>
- De Leeuw, E. D. (2012). Counting and Measuring Online. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 114*(1), 68–78. <https://doi.org/10.1177/0759106312437290>
- DUO. (2021, 1 juni). *Hoofdvestigingen bo - Po - DUO Open onderwijsdata*. duo.nl. Geraadpleegd op 12 januari 2022, van https://www.duo.nl/open_onderwijsdata/primair-onderwijs/scholen-en-adressen/hoofdvestigingen-basisonderwijs.jsp
- Earl, L., & Fullan, M. (2003). Using Data in Leadership for Learning. *Cambridge Journal of Education, 33*(3), 383–394. <https://doi.org/10.1080/0305764032000122023>
- Ensie. (2016, 20 december). *Betrouwbaarheidsinterval*. Geraadpleegd op 17 juni 2022, van <https://www.ensie.nl/cito/betrouwbaarheidsinterval>
- Ettema, B. (2021). *Visuele weergave van onzekerheid in Cito testcores* (ongepubliceerde
- Fernandes, M., Walls, L., Munson, S., Hullman, J., & Kay, M. (2018). Uncertainty displays using quantile dot plots or CDFs improve transit decision-making. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, 144*, 1-12. <https://doi.org/10.1145/3173574.3173718>

- Frery, R. B. (1991). The None-of-the-Above Option: An Empirical Study. *Applied Measurement in Education*, 4(2), 115–124.
https://doi.org/10.1207/s15324818ame0402_2
- Gardner, J. (2013). The public understanding of error in educational assessment. *Oxford Review of Education*, 39(1), 72–92. <https://doi.org/10.1080/03054985.2012.760290>
- Grimm, K. J., Fine, K., & Stegmann, G. (2021). Accounting for Standard Errors of Measurement When Modeling Change. *International Journal of Behavioral Development*, 45(1), 11–18.
- Hop, M., Janssen, J., & Engelen, R. (2016). *Wetenschappelijke verantwoording Rekenen-Wiskunde 3.0 voor groep 5*. Cito B.V., Arnhem.
- Kay, M., Kola, T., Hullman, J. R., & Munson, S. A. (2016). When (ish) is My Bus?: User-centered visualizations of uncertainty in everyday, mobile predictive systems. Proceedings of the 2016 CHI Conference on Human factors in computing systems, 5092-5103. doi: 10.1145/2858036.2858558
- Ledoux, G., Blok, H., Boogaard, M., & Krüger, M. (2009). Opbrengstgericht werken: over de waarde van meetgestuurd onderwijs. (SCO-rapport; Nr. 812). SCO-Kohnstamm Instituut. <http://www.sco-kohnstammstituut.uva.nl/pdf/sco812.pdf>
- Lichtenstein, R. (2020). What You Don't Know about Measurement Error--And Why You Should Care. *Communique*, 48(8), 30–34.
- Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). Making sense of data-driven decision making in education: Evidence from recent RAND research. Santa Monica, CA: RAND Corporation masterscriptie). Rijksuniversiteit Groningen.
- Meijer, J., Ledoux, G., & Elshof, D. P. (2011). *Gebruikersvriendelijke leerlingvolgsystemen in het primair onderwijs*. Kohnstamm Instituut.
- Ministerie van Algemene Zaken. (2021, 30 september). *Hoe legt de basisschool de prestaties van mijn kind vast?* Rijksoverheid.nl. Geraadpleegd op 6 november 2021, van <https://www.rijksoverheid.nl/onderwerpen/basisonderwijs/vraag-en-antwoord/hoe-legt-de-basisschool-de-prestaties-van-mijn-kind-vast>
- Newton, P. E. (2005). The public understanding of measurement inaccuracy. *British Educational Research Journal*, 31, 419–442. doi:10.1080/01411920500148648
- Padilla, L., Kay, M., & Hullman, J. (2020). Uncertainty Visualization. PsyArXiv. <https://doi.org/10.31234/osf.io/ebd6r>
- Schuwirth, L. (2007). *Meten de maat genomen*. Maastricht: Maastricht University. <https://doi.org/10.26481/spe.20071012ls>

- Van Casteren, W., Brukx, D., & De Korte, K. (2018). Specialisatiemogelijkheden in lerarenopleidingen. *ResearchNed*. Published.
<https://www.tweedekamer.nl/downloads/document?id=9b7a0413-6dc1-450e-b3e6-e96456197592&title=Specialisatiemogelijkheden%20in%20lerarenopleidingen.pdf>
- Van der Kleij, F. M., & Eggen, T. J. (2013). Interpretation of the score reports from the Computer Program LOVS by teachers, internal support teachers and principals. *Studies in Educational Evaluation*, 39(3), 144–152.
<https://doi.org/10.1016/j.stueduc.2013.04.002>
- Verhaege, G., Vanhoof, J., Valcke, M., & Van Petegem, P. (2011). Effecten van ondersteuning bij schoolfeedbackgebruik. *Pedagogische Studien*, 88(2), 90–106.
- Whitaker, D., Foti, S., & Jacobbe, T. (2015). The Levels of Conceptual Understanding in Statistics (LOCUS) project: Results of the pilot study. *Numeracy*, 8(2).
<https://doi.org/10.5038/1936-4660.8.2.3>
- Zenisky, A. L., & Hambleton, R. K. (2012). Developing test score reports that work: The process and best practices for effective communication. *Educational Measurement: Issues and Practice*, 31, 21–26 <http://dx.doi.org/10.1111/j.1745-3992.2012.00231.x>.

Bijlage A

Vragenlijst

Visuele weergave van onzekerheid in Cito testcores

Onderaan de tekst kunt u verder door op het rode pijltje te klikken

Beste participant,

Bedankt voor uw interesse in dit onderzoek. Deze vragenlijst is onderdeel van een onderzoek naar de visuele weergave van onzekerheid in Cito testcores onder begeleiding van dr. Niek Frans.

Toetsscores, zoals de toetsen van het leerlingvolgsysteem van Cito, geven een schatting van het niveau van een leerling, welke gepaard gaat met een mate van onzekerheid. Deze onzekerheid wordt op dit moment niet weergegeven in de scoregrafieken van Cito. Dit onderzoek kijkt naar verschillende visuele weergaves van deze onzekerheid.

We zijn voor dit onderzoek op zoek naar leerkrachten in het basisonderwijs en toekomstige leerkrachten die bekend zijn met het leerlingvolgsysteem van Cito. Pabostudenten kunnen participeren als zij zich in het 3e of 4e leerjaar van de studie bevinden.

Het invullen van de vragenlijst neemt ongeveer 15 minuten van uw tijd in beslag. De vragenlijst bestaat uit drie onderdelen. In het eerste deel vragen we u om een aantal achtergrondkenmerken. In het tweede deel vragen we u om het werkelijke niveau van een leerling te schatten op basis van een visuele weergave van de score en de onzekerheid van deze score. Daarnaast vragen we u om een keuze te maken met betrekking tot een passende instructie op basis van deze gegevens. Het derde deel bestaat uit enkele vragen over algemene vaardigheden in het interpreteren van grafieken.

De ingevulde gegevens zullen te allen tijde anoniem verwerkt worden. Daarnaast wordt er geen informatie uitgevraagd die herleidbaar is en kunt u op elk gewenst moment stoppen met uw deelname aan het onderzoek. Als u op akkoord klikt begint u met de vragenlijst en geeft u schriftelijke toestemming voor het gebruik van de ingevulde gegevens voor dit onderzoek.

Als u vragen heeft of geïnteresseerd bent in de resultaten van dit onderzoek, mag u mailen naar één van de onderstaande e-mailadressen.

Vriendelijke groeten,

Femke Drost
Master Orthopedagogiek
Rijksuniversiteit Groningen
f.m.drost@student.rug.nl

Merle van Tuinen

Master Orthopedagogiek
Rijksuniversiteit Groningen
m.e.van.tuinen.1@student.rug.nl

Akkoord?

- Ja, ik ga akkoord
- Nee, ik ga niet akkoord

Deel 1. Algemene vragen

Bent u bekend met het leerlingvolgsysteem van Cito?

- Ja
- Nee

Welke opleiding heeft u gevolgd/volgt u?

- Pabo
- Academische Pabo
- Anders, namelijk

Heeft u deze opleiding afgerond?

- Ja
- Nee, ik ben nog bezig met de opleiding

In welk jaar van de opleiding zit u?

- Jaar 1
- Jaar 2
- Jaar 3
- Jaar 4
- Anders, namelijk

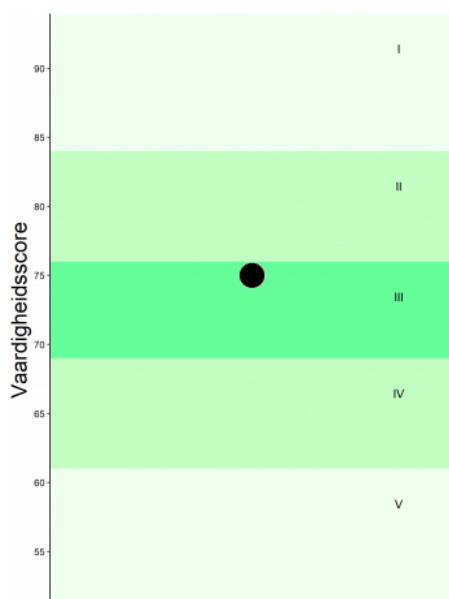
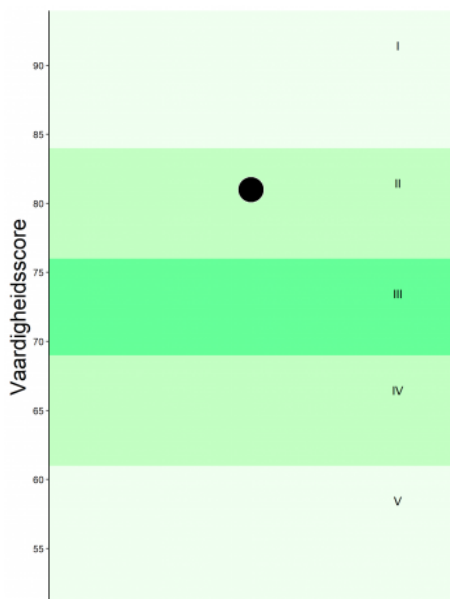
In welke provincie bent u (voornamelijk) werkzaam?

- Friesland
- Drenthe
- Groningen
- Flevoland
- Overijssel
- Noord-Holland
- Zuid-Holland
- Utrecht
- Gelderland
- Noord-Brabant
- Zeeland
- Limburg

Deel 2. Interpretatie onzekerheid en keuze instructie (Visualisatie puntschatting)

In dit deel van de vragenlijst krijgt u 5 keer een geobserveerde score van een leerling te zien. We stellen u per visualisatie twee vragen over uw inschatting van het niveau van de leerling. Deze kansinschatting loopt van 0 tot 100%. Vervolgens wordt er steeds gevraagd wat voor keuze u zou maken met betrekking tot een passende instructie voor de betreffende leerling. Aan het einde van dit onderdeel wordt u gevraagd om een toelichting te geven over deze keuzes.

Onderstaand figuur toont een voorbeeld van een leerlingrapport. De zwarte punt geeft de geobserveerde score van een leerling weer. De niveaus I t/m V zijn aangegeven door middel van gekleurde gebieden, rechts in de grafiek ziet u telkens de niveaus weergegeven.



Hoe schat u de kans in dat het werkelijke niveau van deze leerling lager is dan niveau III?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau II is?

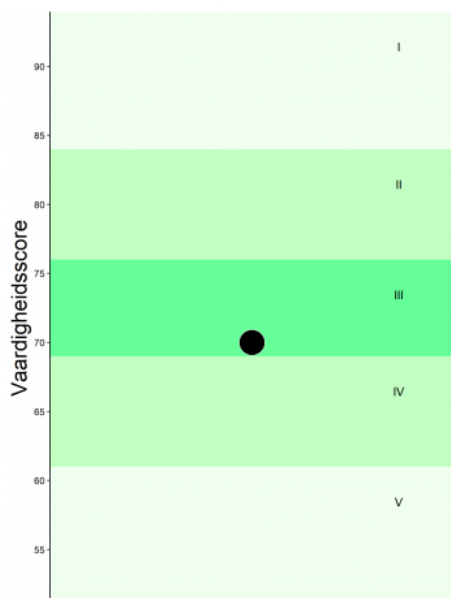
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling lager is dan niveau III?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau III is?

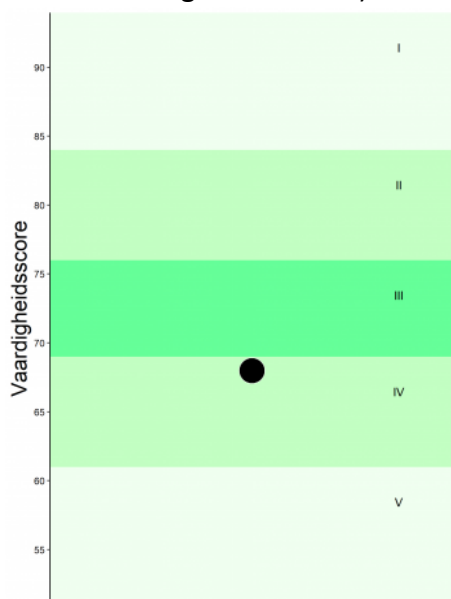
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling hoger is dan niveau IV?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau IV is?

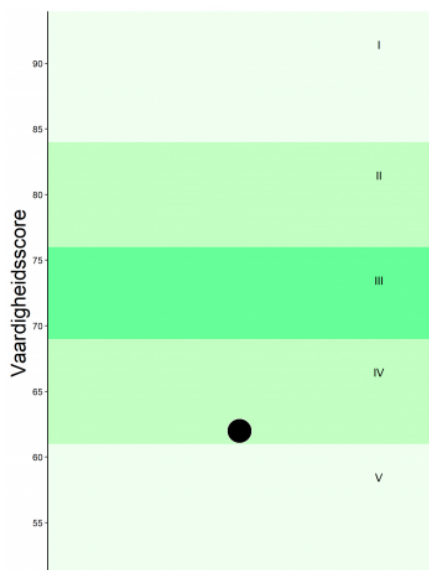
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling lager is dan niveau IV?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau IV is?

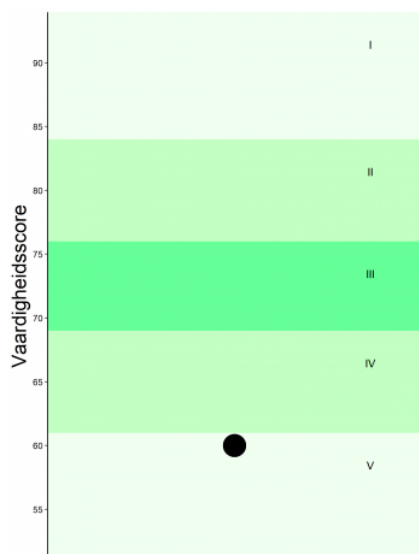
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling hoger is dan niveau V?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau V is?

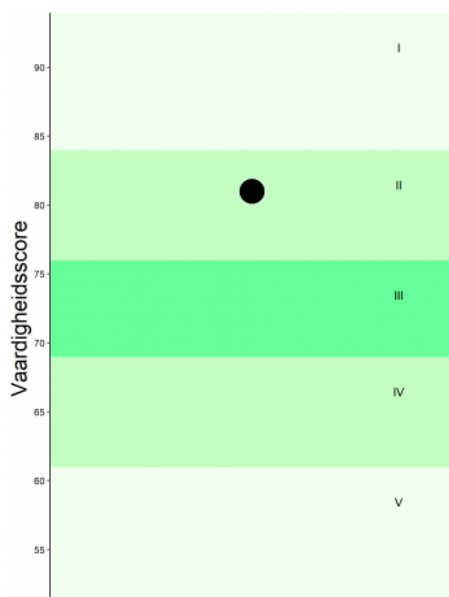
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)

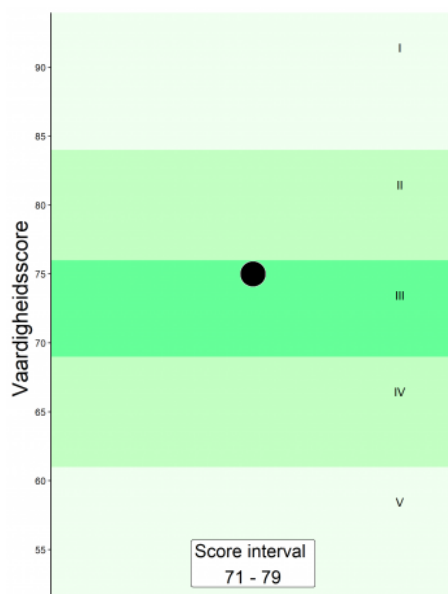
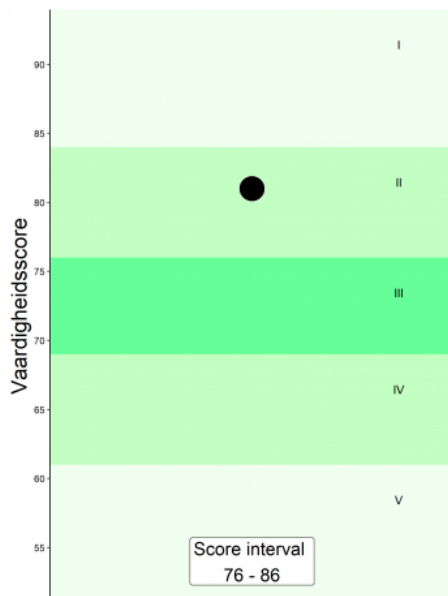


Hierboven ziet u nog een voorbeeld van de visualisatie. Kunt u toelichten waar u naar gekeken heeft in de vorige vijf visualisaties om een passende instructievorm te kiezen?

Deel 2. Interpretatie onzekerheid en keuze instructie (Visualisatie puntschatting + BHI)

In dit deel van de vragenlijst krijgt u 5 keer een geobserveerde score van een leerling met afgebeelde foutenmarge te zien. We stellen u per visualisatie twee vragen over uw inschatting van het niveau van de leerling. Deze kansinschatting loopt van 0 tot 100%. Vervolgens wordt er steeds gevraagd wat voor keuze u zou maken met betrekking tot een passende instructie voor de betreffende leerling. Aan het einde van dit onderdeel wordt u gevraagd om een toelichting te geven over deze keuzes.

Onderstaand figuur toont een voorbeeld van een leerlingrapport met foutmarge. De zwarte punt geeft de geobserveerde score van een leerling weer. Het 67% betrouwbaarheidsinterval staat onderaan de grafiek weergegeven. De niveaus I t/m V zijn aangegeven door middel van gekleurde gebieden, rechts in de grafiek ziet u telkens de niveaus weergegeven.



Hoe schat u de kans in dat het werkelijke niveau van deze leerling lager is dan niveau III?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau II is?

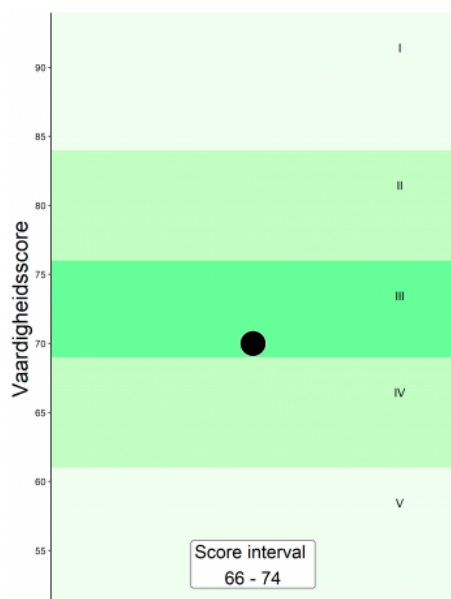
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling lager is dan niveau III?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau III is?

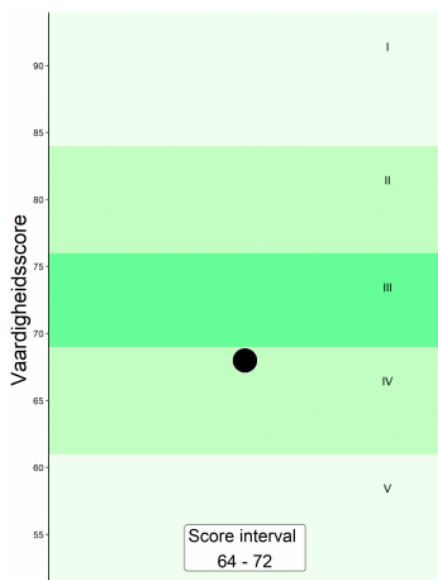
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling hoger is dan niveau IV?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau IV is?

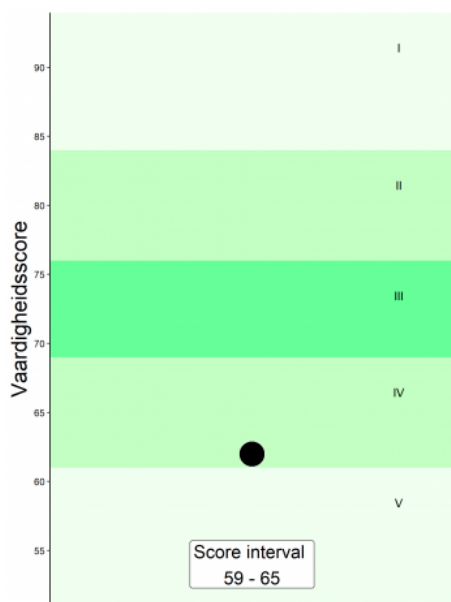
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling lager is dan niveau IV?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau IV is?

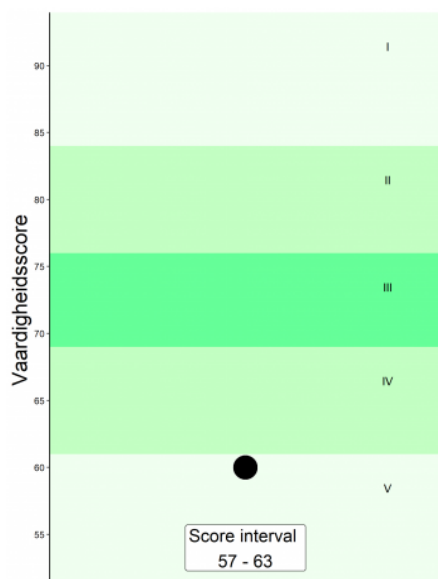
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling hoger is dan niveau V?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau V is?

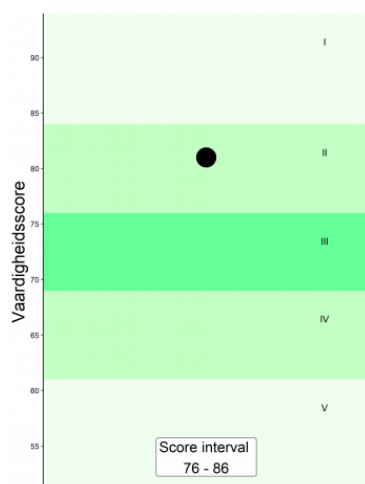
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



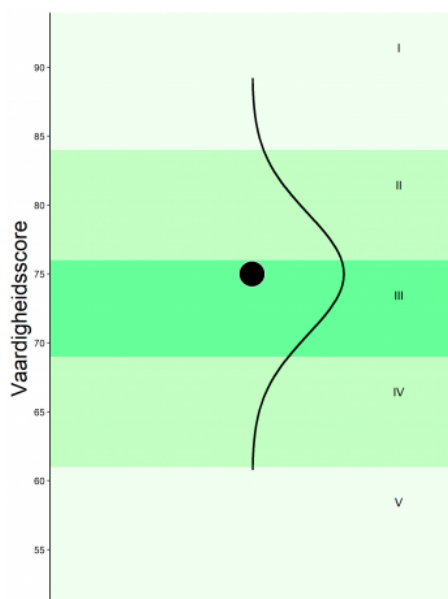
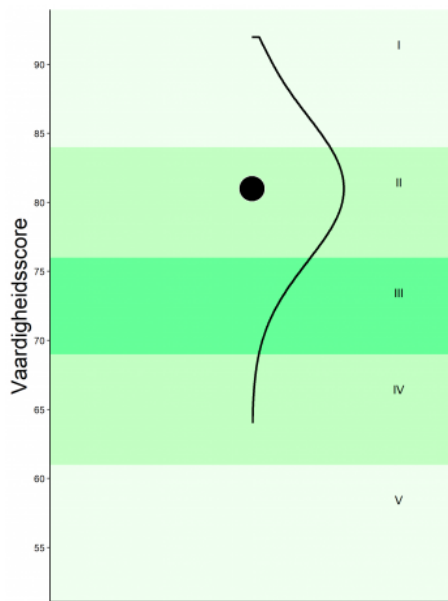
Hierboven ziet u nog een voorbeeld van de visualisatie. Kunt u toelichten waar u naar gekeken heeft in de vorige vijf visualisaties om een passende instructievorm te kiezen?

Deel 2. Interpretatie onzekerheid en keuze instructie (Visualisatie Probability Density Funtion)

In dit deel van de vragenlijst krijgt u 5 keer een geobserveerde score van een leerling met afgebeelde foutenmarge te zien. We stellen u per visualisatie twee vragen over uw inschatting van het niveau van de leerling. Deze kansinschatting loopt van 0 tot 100%. Vervolgens wordt er steeds gevraagd wat voor keuze u zou maken met betrekking tot een

passende instructie voor de betreffende leerling. Aan het einde van dit onderdeel wordt u gevraagd om een toelichting te geven over deze keuzes.

Onderstaand figuur toont een voorbeeld van een leerling rapport met foutmarge. De zwarte punt geeft de geobserveerde score van een leerling weer. De zwarte lijn geeft de waarschijnlijkheid van de werkelijke score weer. Hoe verder de lijn naar rechts ligt in een bepaald niveau, hoe waarschijnlijker dat het de werkelijke score van de leerling is. Het meest rechtse punt is altijd gelijk aan de geobserveerde score. De niveaus I t/m V zijn aangegeven door middel van gekleurde gebieden, rechts in de grafiek ziet u telkens de niveaus weergegeven.



Hoe schat u de kans in dat het werkelijke niveau van deze leerling lager is dan niveau III?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau II is?

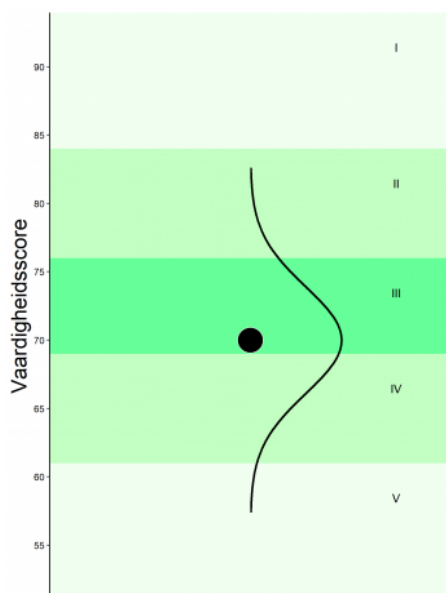
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling lager is dan niveau III?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau III is?

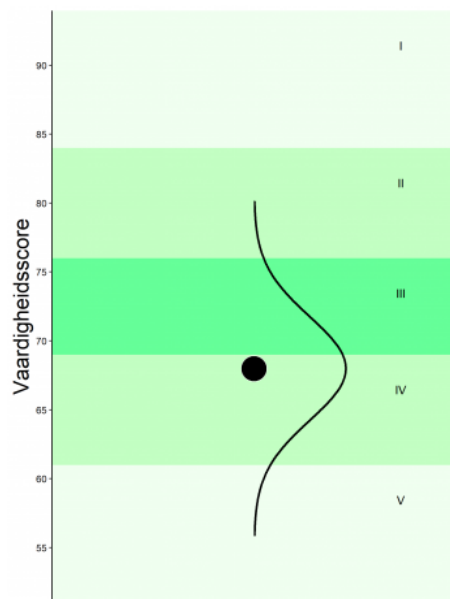
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling hoger is dan niveau IV?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau IV is?

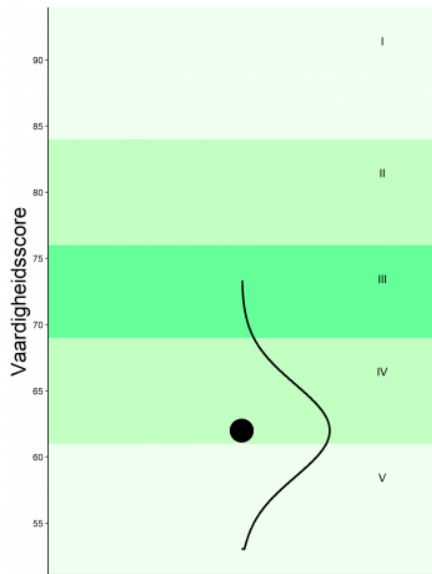
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

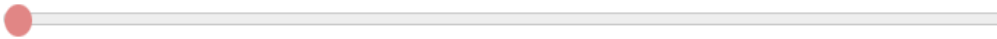
- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling lager is dan niveau IV?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau IV is?

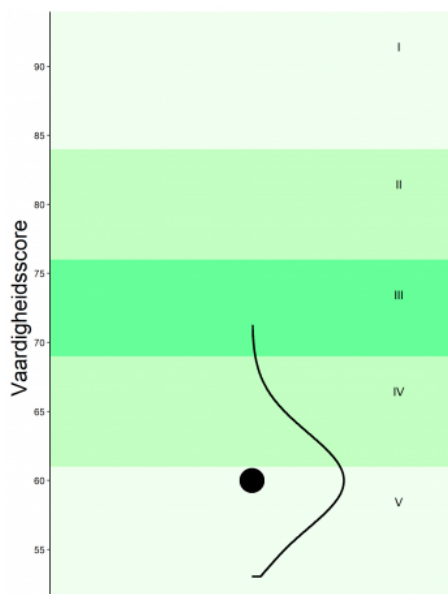
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling hoger is dan niveau V?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau V is?

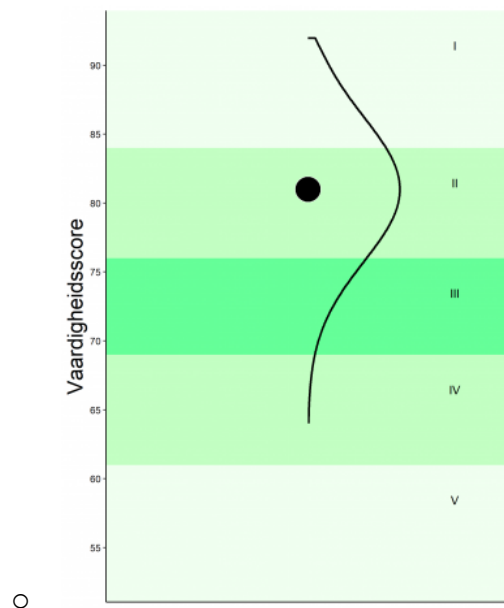
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)

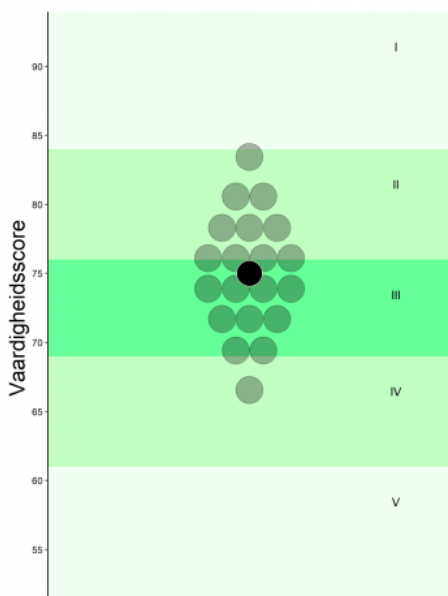
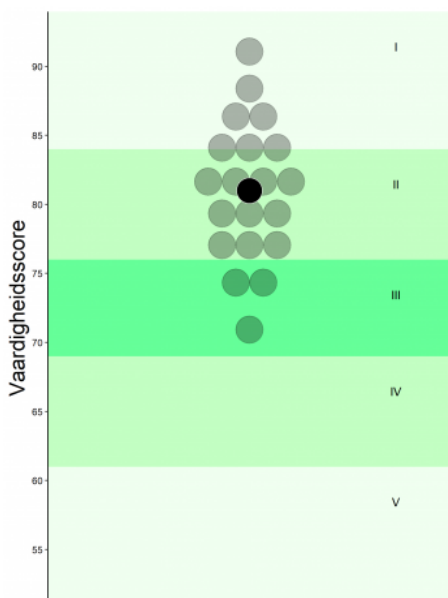


Hierboven ziet u nog een voorbeeld van de visualisatie. Kunt u toelichten waar u naar gekeken heeft in de vorige vijf visualisaties om een passende instructievorm te kiezen?

Deel 2. Interpretatie onzekerheid en keuze instructie (Visualisatie Quantile Dot Plot)

In dit deel van de vragenlijst krijgt u 5 keer een geobserveerde score van een leerling met afgebeelde foutenmarge te zien. We stellen u per visualisatie twee vragen over uw inschatting van het niveau van de leerling. Deze kansinschatting loopt van 0 tot 100%. Vervolgens wordt er steeds gevraagd wat voor keuze u zou maken met betrekking tot een passende instructie voor de betreffende leerling. Aan het einde van dit onderdeel wordt u gevraagd om een toelichting te geven over deze keuzes.

Onderstaand figuur toont een voorbeeld van een leerling rapport met foutmarge. De zwarte punt geeft de geobserveerde score van een leerling weer. De doorzichtige bolletjes in de grafiek geven de foutenmarge voor deze score aan. Elk bolletje staat voor een waarschijnlijkheid van 5%. De niveaus I t/m V zijn aangegeven door middel van gekleurde gebieden, rechts in de grafiek ziet u telkens de niveaus weergegeven.



Hoe schat u de kans in dat het werkelijke niveau van deze leerling lager is dan niveau III?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau II is?

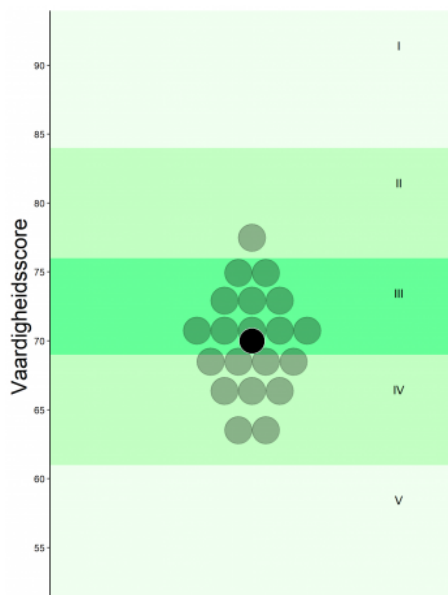
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling lager is dan niveau III?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau III is?

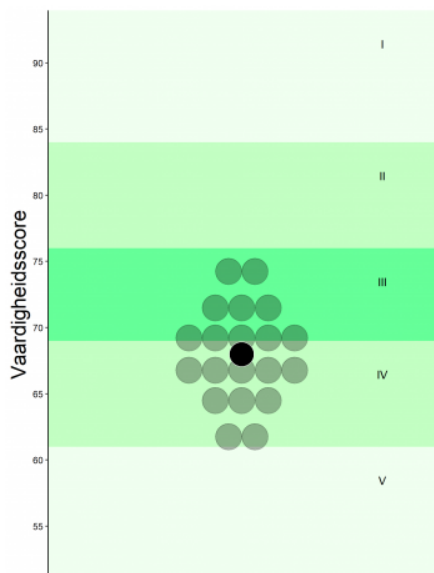
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling hoger is dan niveau IV?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau IV is?

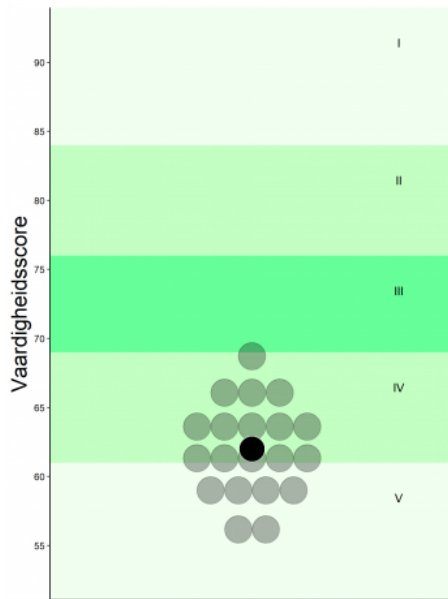
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling lager is dan niveau IV?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau IV is?

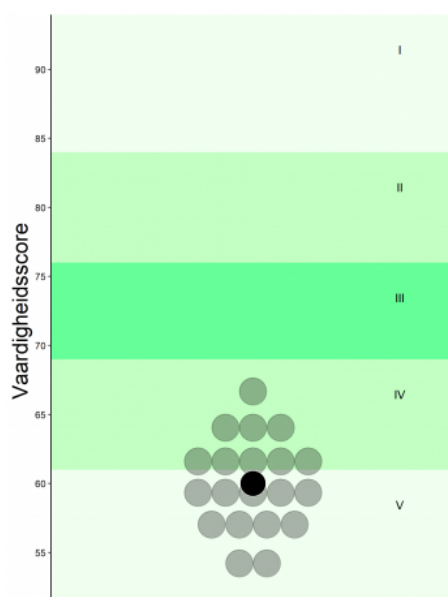
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)



Hoe schat u de kans in dat het werkelijke niveau van deze leerling hoger is dan niveau V?

0 10 20 30 40 50 60 70 80 90 100

Procent



Hoe schat u de kans in dat het werkelijke niveau van deze leerling niveau V is?

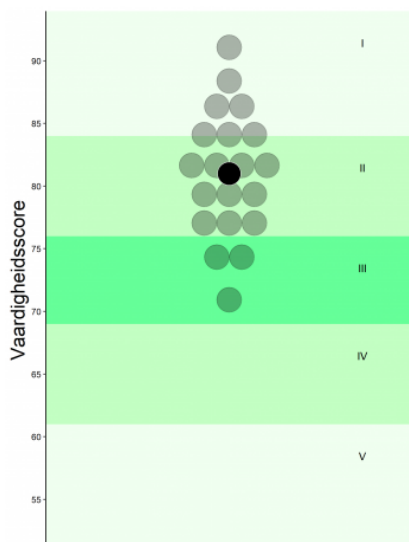
0 10 20 30 40 50 60 70 80 90 100

Procent



In welke instructiegroep zou u deze leerling plaatsen op basis van deze gegevens?

- Deze leerling plaats ik in de basisondersteuning
- Deze leerling plaats ik in een instructiegroep met extra ondersteuning (zoals een verlengde instructie)

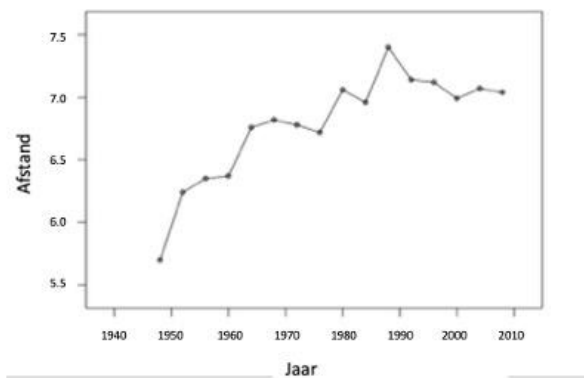


Hierboven ziet u nog een voorbeeld van de visualisatie. Kunt u toelichten waar u naar gekeken heeft in de vorige vijf visualisaties om een passende instructievorm te kiezen?

Deel 3. Statistische kennis

Vraag 1.

Het verspringen is sinds 1948 een onderdeel voor vrouwen op de Olympische zomerspelen. De onderstaande plot laat de lengte van de versprong (in meters) van de winnaressen sinds 1948 zien.

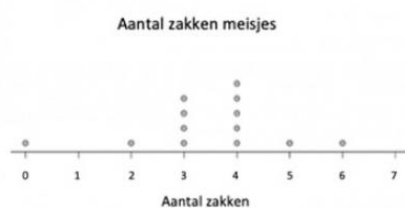
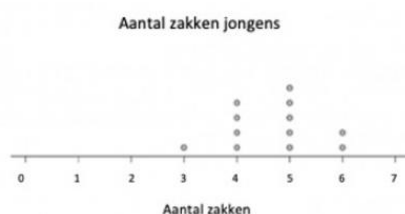


Welke van de onderstaande uitspraken geeft de beste beschrijving van de trend van de lengte van de winnende sprong weer?

- De lengte van de sprong neemt iedere 4 jaar toe, omdat meer vrouwen mee gaan doen aan de sport.
- De lengte van de langste sprong nam toe van 1948 tot ongeveer 1988, maar nam na 1988 niet meer toe.
- Er is een stabiele groei te zien in de lengte van de winnende sprong van 1948 tot 2008.
- Er is te veel onregelmatigheid in de plot om een duidelijke uitspraak te doen over de trend.

Vraag 2.

De onderstaande dotplots laten een verdeling zien van de hoeveelheid zakken op kleding van 12 jongens en 13 meisjes.

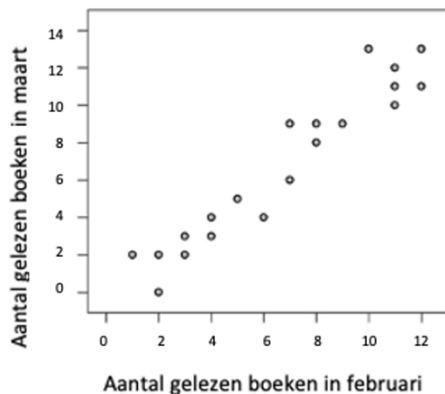


Welke van de onderstaande uitspraken is waar?

- De data voor jongens zijn meer gevarieerd.
- De mediaan voor zakken bij jongens is hoger dan die voor meisjes.
- De modus voor aantal zakken bij jongens en meisjes is 4.
- Het maximaal aantal zakken voor jongens is hoger dan dat van meisjes.

Vraag 3.

Een leerkracht heeft data verzameld over de hoeveelheid boeken die de leerlingen uit haar klas in februari en maart hebben gelezen. Een scatterplot van deze data staat hieronder weergegeven.



Gebaseerd op de grafiek, wat is de best passende conclusie?

- De modus van het aantal boeken gelezen in februari is 10.
- Leerlingen die meer boeken lezen in februari zijn geneigd om minder boeken te lezen in maart.
- Leerlingen die meer boeken lezen in februari zijn geneigd om meer boeken te lezen in maart.
- Er is geen relatie tussen het aantal boeken gelezen in februari en maart.

Vraag 4.

Er is een vragenlijst afgenomen onder 400 volwassenen. Er is data verzameld over gender en de support voor een bepaalde kandidaat. Dit is samengevat in de onderstaande tabel

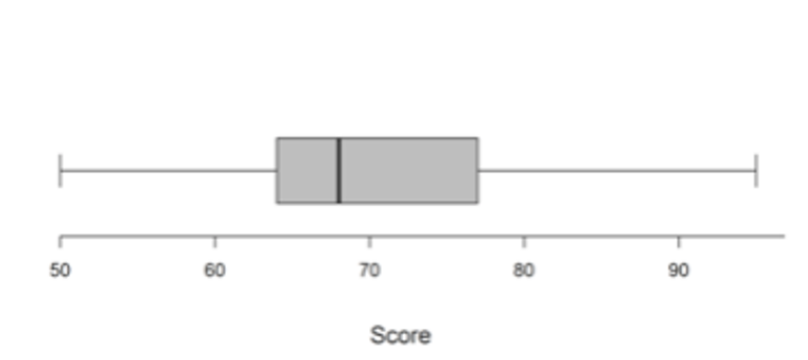
	Man	Vrouw	Totaal
Support	112	102	214
Geen support	70	116	186
Totaal	182	218	400

Welke proportie van de respondenten die 'support' hebben geantwoord zijn man?

- 112/400
- 182/400
- 112/214
- 112/182

Vraag 5.

Onderstaande boxplot beschrijft toetsscores van 1031 leerlingen.



Welke van de onderstaande uitspraken is in ieder geval goed?

- Ten minste 60% van de leerlingen scoorde lager dan 70 op de toets.
- Ten minste 60% van de leerlingen scoorde hoger dan 70 op de toets.
- Ten minste 30% van de leerlingen scoorde lager dan 70 op de toets.
- Ten minste 30% van de leerlingen scoorde hoger dan 70 op de toets.t gebied, voordat we de vragenlijst afronden.

Bedankt voor uw deelname aan ons onderzoek.

Mocht u nog vragen hebben over het onderzoek of de vragenlijst, dan kunt u contact opnemen met:

Merle van Tuinen (m.e.van.tuinen.1@student.rug.nl)

Femke Drost (f.m.drost@student.rug.nl)