

KLIMAATSCHAAL

Deze beoordeling betreft alleen de leerlingversie. Waar nodig worden twee beoordelingen gegeven, één voor gebruik op groepsniveau en één voor gebruik op individueel niveau.

Theoretische Uitgangspunten

De Klimaatschaal beoogt het pedagogisch klimaat te meten. De theoretische uitgangspunten hiervoor zijn vooral gebaseerd op het werk van Moos (1979) die drie basisdimensies onderscheidt; de relatiedimensie (met medeleerlingen en leerkracht), de persoonlijke groei- of doelgerichtheidsdimensie en de systeemhandhavingsdimensie. Deze worden bij de Klimaatschaal 'vertaald' in de volgende vier dimensies: 'kwaliteit onderlinge leerlingrelaties' (relatiedimensie), 'kwaliteit interactie leerkracht-leerling' (relatiedimensie), 'sfeer in de klas' (groeidimensie) en 'orde in de klas' (systeemhandhavingsdimensie). Overigens wordt het model van Moos vrij summier besproken.

De operationalisatie van de dimensies naar de items is voor verbetering vatbaar. Zo worden er geen inhoudelijke analyses van de gepostuleerde dimensies gegeven. Via diverse itempools en statistische analyses komt de uiteindelijke vragenlijst tot stand, maar niet duidelijk is onder andere wat mogelijke subdimensies zijn van de dimensies, hoe de oorspronkelijke items daarbij pasten en wat voor type items zijn afgevallen. Ter illustratie, Pianta (2001) onderscheidt drie subdimensies binnen de leerling-leerkracht relatie, namelijk 'Nabijheid', 'Conflict' en 'Afhankelijkheid'. De items in de betreffende schaal in de Klimaatschaal lijken zich voornamelijk te richten op 'Nabijheid', terwijl items over 'Conflict' en 'Afhankelijkheid' ontbreken. De inhoudelijke representatie van het begrip leerling-leerkracht relatie is daardoor beperkt. De items in de schaal 'Sfeer in de klas' gaan vooral over ruzie, pesten, lastig vallen, schoppen en uitschelden. Dit is mogelijk één van de voorwaarden om tot groei te komen, maar komt inhoudelijk niet overeen met de persoonlijke groei- of doelgerichtheidsdimensie waarvan deze schaal een operationalisatie zou moeten zijn. Het lijkt eerder een negatieve versie van de schaal 'Kwaliteit onderlinge leerlingrelaties'. De persoonlijke groei- of doelgerichtheidsdimensie is daardoor beperkt gerepresenteerd in de vragenlijst.

De Klimaatschaal kan gebruikt worden vanaf groep 4 van zowel regulier als speciaal basisonderwijs, speciaal onderwijs, VO en MBO. Dit met het (hoofd)doel om het pedagogisch klimaat in kaart te brengen en om handvatten te geven voor verbeteringen in het pedagogisch klimaat. Het instrument is zowel bedoeld voor gebruik op individueel niveau (bijvoorbeeld het signaleren van zorgleerlingen) als voor gebruik op groepsniveau (bijvoorbeeld het gebruik in persoonlijke ontwikkelingsgespreken van leerkrachten of het evalueren van de effectiviteit van de interventie die gericht is op het verbeteren van het klimaat). In het eerste deel van de handleiding krijgt de eerste (individuele) functie weinig aandacht, waardoor het lijkt alsof het instrument voornamelijk bedoeld is voor gebruik op groepsniveau; dit zet de testgebruiker mogelijk op het verkeerde been.

Kwaliteit van het Testmateriaal

Papier-en-potloodversie

N.v.t.

Computerversie

De test is gestandaardiseerd en er wordt een automatisch scoringsstelsel gebruikt. De items zijn vrij van kwetsende inhoud en correct geformuleerd. Het kind vult zelf de vragenlijst in onder supervisie van de leerkracht. De instructie voor het kind is duidelijk beschreven in bijlage 6 van de handleiding. In de instructie zijn geen voorbeeldvragen opgenomen (volgens de auteurs is uitleg door de leerkracht ruim voldoende) en geen informatie over de beschikbare tijd (of het feit dat er geen tijdslimiet geldt). Door de simpele en overzichtelijke interface van de website waarop de vragenlijst wordt ingevuld, lijkt het er niet op dat er iets mis zou kunnen gaan met het invullen en het gebruik van de vragenlijst. Ook jongere kinderen zullen hun weg, zo nodig met enige hulp, in de vragenlijst kunnen vinden. Bij het overslaan van een item kan men niet naar de volgende bladzijde. Er verschijnt dan bij het betreffende item in het rood een kleine waarschuwing. Wanneer dit het eerste item van een pagina is en men is inmiddels aan het eind van die pagina gekomen, valt dit niet meteen op. Dit had wat duidelijker kunnen worden aangegeven. De gebruikersinterface ziet er

overzichtelijk uit met steeds een beperkt aantal items op een pagina, voldoende ruimte tussen de verschillende items en een duidelijk lettertype. Toegang tot de Klimaatschaal voor zowel leerling als leerkracht geschiedt met een inlognaam en een wachtwoord. De resultaten zijn alleen door de eigen leerkracht in te zien en er wordt gebruik gemaakt van een beveiligde https-verbinding.

Kwaliteit van de Handleiding

De instructie voor de testleider is duidelijk (bijlage 7) en verwijst naar bijlage 6 waarin de instructie voor de geteste uitgeschreven staat. Specifieke instructies over de manier van afnemen zouden nog kunnen worden opgenomen, zoals richtlijnen over de werkruimte, het licht en/of het kind de vragenlijst moet invullen buiten de klas of in de klas. In bijlage 7 wordt tevens informatie gegeven over de bediening en mogelijkheden van de software c.q. website. Na inloggen zijn er voor de leerkracht met name onder de kopjes ‘handelingswijzer Klimaatschaal’ en ‘SIGA’ allerlei links naar documenten die nuttig en relevant zijn. Echter, hier wordt in de handleiding bijna niets over gezegd. Het gebruik hiervan zou nog beter tot zijn recht kunnen komen wanneer hier in de handleiding meer uitleg over wordt gegeven en er een soort beslisboom wordt opgenomen over wat je wanneer kunt gebruiken als leerkracht. In hoofdstuk 9 worden vier verschillende typen casus besproken als ook een schoolanalyse. Buiten de invloed van geslacht en normgroep op de scores van de klimaatschaal (tabel 11), wordt er niet gesproken over andere soorten informatie die belangrijk zijn bij de interpretatie. De klimaatschaal kan worden afgenomen door leerkrachten. Interpretatie dient te geschieden door professionals, zoals IB'ers, orthopedagogen, psychologen, ambulante begeleiders, mentoren of zorgcoördinatoren. Voor technische ondersteuning is een helpdesk per mail bereikbaar.

Normen

De normgegevens zijn verzameld in de periode 2010-2012. Per normgroep (in totaal zeven verschillende normgroepen) varieert het aantal kinderen van 558 tot 25254. Voor gebruik op individueel niveau zijn de normgroepen qua grootte ‘goed’ voor alle normgroepen. Voor gebruik op groepsniveau zijn de normgroepen qua grootte ‘goed’ voor de meeste normgroepen (1 t/m 5; zie tabel 2 op p. 22), ‘voldoende’ voor de normgroep REC-scholen (n tussen 750 en 1000), maar ‘onvoldoende’ voor de normgroep MBO ($n < 750$). Bij deze aantallen wordt uitgegaan van een gemiddelde groeps grootte van 25 leerlingen. Of dit klopt is niet te achterhalen, omdat over het gemiddeld aantal personen per klas en het aantal klassen geen informatie wordt verstrekt.

De steekproef voor de normen is samengesteld door het benaderen van diverse onderwijsopleidingscentra. Van de centra die mee wilden werken hebben de studenten die in opleiding zijn tot leerkracht of specialist in het onderwijs, de Klimaatschaal bij hun klas met kinderen afgenomen. Hiermee is geen sprake van een aselechte steekproef, maar een gelegenheidssteekproef. De steekproef van scholen wordt vergeleken met de populatiegegevens wat betreft urbanisatiegraad, regio en religie. Grote steden zijn in de meeste normgroepen oververtegenwoordigd. De regio Noord ontbreekt vrijwel geheel. Gegevens over etniciteit (of over mogelijke proxy's zoals leerlinggewicht of SES) en sekse ontbreken. Ook wordt niets vermeld over de samenstelling naar leerjaar van de diverse normgroepen of over het mogelijke effect van leerjaar op de scores. De normgroepen ‘Algemeen’ en ‘Specifiek’ zijn verschillend. Eigenlijk zou dus voor beide onderdelen apart de representativiteit moeten worden gecheckt, maar dit zal vermoedelijk geen grote afwijkingen opleveren.

De normscores voor individueel gebruik worden uitgedrukt in septielen die respectievelijk 5%, 10%, 20%, 30%, 20%, 10% en 5% van de scoreverdeling omvatten. Septielen hebben een gemiddelde van 4 en een standaarddeviatie van 1.5. Dit (ongebruikelijke) type normschaal had mogelijk wat meer kunnen worden toegelicht, bijvoorbeeld door middel van een grafiekje van de normaalverdeling en een vergelijking met stanines.

De gemiddelden, standaardafwijkingen, gegevens over de scoreverdeling, de standaardmeetfout en de standaard schattingsfout die worden vermeld hebben betrekking op de scores op individueel niveau, dergelijke gegevens op groepsniveau worden niet vermeld. De scores zijn op alle schalen behoorlijk scheef naar links, met name die voor de twee specifieke schalen.

TOELICHTING BIJ DE BEOORDELING

Verschillen tussen jongens en meisjes worden besproken als ook verschillen tussen de verschillende normgroepen c.q. schoolniveaus (tabel 11). Verschillen met betrekking tot bijvoorbeeld etniciteit zijn niet onderzocht.

Voor de normen op groepsniveau zou het ideaal zijn om normen te hebben per groepsgrootte, echter dit is zeer bewerkelijk. Daarom is gekozen voor het toepassen van een rekenregel om zodoende het verschil in nauwkeurigheid van het groepsgemiddelde op te vangen. Hiervoor wordt gebruik gemaakt van het gemiddelde en de standaarddeviatie van de steekproevenverdeling waarbij voor de populatieverdeling wordt uitgegaan van de septielverdeling op individueel niveau (d.w.z. $N(4; 1.5)$). Vervolgens worden met behulp hiervan de groepsgemiddelden omgezet naar standaardnormaal scores om de gemiddelden te normeren en dus te vergelijken. Waar bij deze rekenregel echter aan voorbij lijkt te worden gegaan, is dat de steekproevenverdeling die gebruikt wordt om een uitspraak te doen over de nauwkeurigheid van het gevonden gemiddelde, uitgaat van een 'populatie' van groepen met *hetzelfde* aantal leerlingen. Het kan niet gebruikt worden om het gemiddelde van een individuele groep van 6 leerlingen te vergelijken met het gemiddelde van een individuele groep van 30 leerlingen. Dit leidt tot onjuiste (en ook verwarrende) conclusies, zoals bij het voorbeeld op pagina 44 van de handleiding waarin wordt aangetoond dat eenzelfde groepsgemiddelde bij een verschillend aantal leerlingen in de groep leidt tot verschillende normscores. Dat is niet correct; een geobserveerd gemiddelde blijft een geobserveerd gemiddelde, alleen een gemiddelde gebaseerd op minder leerlingen in de groep is onnauwkeuriger (d.w.z. heeft een breder betrouwbaarheidsinterval) dan een gemiddelde gebaseerd op meer leerlingen in een groep. Hier komt verder bij dat de gebruikte rekenregel is gebaseerd op een *gekozen* normaal verdeling van de individuele normscores (d.w.z. septielverdeling), terwijl de geobserveerde en hier gebruikte verdeling een verdeling betreft van ruwe scores op individueel niveau die niet normaal verdeeld zijn. Het lijkt juister om uit te gaan van de ruwe scores en van daaruit normen te construeren gebaseerd op de groepsgemiddelden waarbij rekening wordt gehouden met de groepsgrootte.

Betrouwbaarheid

In Hoofdstuk 5 worden de resultaten van het betrouwbaarheidsonderzoek besproken. Cronbach's alfa is berekend voor alle zeven normgroepen en voor de vier schalen van de Klimaatschaal. De waarden variëren van .75 voor de schaal 'Orde' en de normgroep speciaal basisonderwijs tot .94 voor de schaal 'Interactie' en de normgroepen VMBO en HAVO/VWO. Deze waarden zijn, met uitzondering van drie waarden, alle boven de .80 en daarmee 'goed' voor een instrument voor minder belangrijke beslissingen op individueel en/of groepsniveau. Voor het berekenen van alfa is gebruik gemaakt van de normgegevens.

De gemiddelde test-hertest correlaties na vier en zeven dagen (groepen samengevoegd) voor de vier schalen zijn respectievelijk .85, .86, .71 en .78. Dit betekent 'voldoende/goed' voor 'minder belangrijke beslissingen op individueel niveau' en 'goed' voor 'onderzoek op groepsniveau'. Het tijdsinterval is relatief kort (vier en zeven dagen) en de vraag is in hoeverre leerlingen hierdoor niet hetzelfde antwoord invullen zoals ze zich dat wellicht herinneren. Opmerkelijk is dat er redelijke verschillen zitten tussen de correlaties berekend na vier dagen en na zeven dagen. Bijvoorbeeld de test-hertest correlatie voor de schaal 'Interactie' is .86 na vier dagen maar .70 na zeven dagen. Voor de schaal 'Kwaliteit onderlinge leerlingrelaties' is dat .83 na vier dagen en .91 na zeven dagen. Dit roept vragen op over de stabiliteit over langere intervallen. Dit onderzoek is uitgevoerd bij een steekproef van 129 kinderen uit vier van de zeven normgroepen.

Op basis van SEM is omega berekend per normgroep voor de vier schalen van de Klimaatschaal. De waarden variëren van .81 voor de schaal 'Orde' voor de normgroep speciaal basisonderwijs tot .96 voor de schaal 'Interactie' voor de normgroepen VMBO en HVAO/VWO. Voor zowel minder belangrijke individuele beslissingen als beslissingen op groepsniveau is dat 'goed'. Voor dit onderzoek is gebruik gemaakt van de normgegevens.

Het is echter nodig een kanttekening bij bovenstaande resultaten te plaatsen voor de toepassing van bovengenoemde coëfficiënten op groepsniveau. Hierbij is het niet relevant of het antwoord van een individu op het ene item in een schaal correleert met dat op een ander item in een schaal (wat vervolgens leidt tot een hoge alfa of omega), maar of het oordeel van de ene leerling over het klimaat in de klas correleert met het

oordeel van een andere leerling (vergelijk bijvoorbeeld de wijze waarop de betrouwbaarheid van organisatieklimaat wordt vastgesteld in de A&O-literatuur). Wanneer er een grote variatie zou zijn tussen leerlingen in een klas in hun oordeel over het klimaat, heeft het gemiddelde in die klas immers geen betekenis. Het zegt dan meer iets over de leerlingen dan over de leerkracht. Met andere woorden, de betrouwbaarheid voor gebruik op groepsniveau moet met een andere maat worden vastgesteld, bijvoorbeeld met Intra Class Correlaties. De gepresenteerde gegevens zeggen niets over de overeenstemming van leerlingen over 'het' klimaat in de klas.

Begripsvaliditeit

In hoofdstuk 6 worden de resultaten van het (begrips)validiteitsonderzoek besproken. De factoranalyse (CFA) bevestigt dat de vier schalen onafhankelijke factoren zijn, waarbij de fit voor het algemene gedeelte goed is en voor het specifiek gedeelte acceptabel tot goed. Deze factorstructuur wordt eveneens teruggevonden in de zeven normgroepen, waarbij de fit in de MBO normgroep een uitzondering is. Voor de eendimensionaliteit van de vier factoren wordt een CFA uitgevoerd op het algemeen en specifiek gedeelte samen voor een subgroep van de steekproef. Deze resultaten laten zien dat items hoog laden op de bedoelde factor en laag op de andere factoren. Ook een analyse waarbij de fit voor alle vier de factoren in één analyse werd bepaald (in een subgroep van 885 personen) laat een goede passing zien. Bij bovenstaande onderzoeken dient vermeld te worden dat hier geen rekening gehouden is met het feit dat de resultaten op groepsniveau worden geïnterpreteerd. Een multi-level aanpak zou daar bijvoorbeeld geschikt voor zijn.

De item-rest correlaties variëren van .44 tot .74 en zijn daarmee 'goed' te noemen. Verder is de invariantie van de factorstructuur onderzocht voor de normgroepen en voor geslacht. De resultaten laten zien dat de factorstructuur vergelijkbaar is over deze subgroepen, waarbij de fit-waarden voor het algemene gedeelte goed zijn en voor het specifieke gedeelte acceptabel tot goed. De invariantie van de factorstructuur is helaas niet onderzocht voor etniciteit.

Voor de convergente validiteit zijn de relaties onderzocht met schalen van de SVL, de NPV-J-2, de PMT-K-2 en de SIGA. De Klimaatschaal subschaal 'Kwaliteit onderlinge leerlingrelaties' hangt hoog ($r = .64$) samen met de subschaal 'Sociaal aanvaard voelen' van de SVL. De subschaal 'Kwaliteit interactie leerkracht-leerling' hangt matig ($r = .41$) samen met de subschaal 'Relatie met leerkrachten (in het algemeen)' van de SVL. In tegenstelling tot de SVL zijn bij de andere gebruikte vragenlijsten (NPV-J, PMT-K en SIGA) geen directe 'soortgenoot'-schalen aanwezig. Tevens speelt bij deze vragenlijsten mee dat het om persoonlijkheids(/gedrags)dimensies en bij de Klimaatschaal om kenmerken van de leeromgeving gaat. De zin van de validatie aan deze vragenlijsten is daardoor betrekkelijk. Wat betekent bijvoorbeeld de correlatie van .31 tussen Recalcitrantie (persoonlijkheidsdimensie) en Orde in de klas (klimaatdimensie)? Dat de leerling gedeeltelijk zelf verantwoordelijk is voor mogelijke wanorde? Het zou beter zijn geweest om werkelijke soortgenoten te kiezen, zoals de LLRV. Voor het onderzoeken van de convergente validiteit is gebruik gemaakt van subgroepen. Deze subgroepen bestaan altijd uit meerdere normgroepen waarbij de aantallen wisselend zijn van 83 tot 885. De divergente validiteit is niet onderzocht.

Tevens is gekeken naar verschillen tussen geslacht en verschillen tussen normgroepen (zie Tabel 11). De resultaten zijn in grote lijnen conform de beschreven verwachtingen. De verschillen tussen de normgroepen blijken significant. Volgens de auteurs zijn deze klein, maar de verschillen tussen de extreem scorende normgroepen bedragen toch ruim een halve standaarddeviatie op alle vier de variabelen. Er worden geen gegevens vermeld over verschillen tussen leerjaren binnen de normgroepen en tussen etnische groepen. Mogelijkerwijs zouden hier verschillen aanwezig kunnen zijn, zeker gezien de sociaal-emotionele ontwikkeling gedurende kindertijd en adolescentie.

Net als bij *Betrouwbaarheid* dient ook hier een kanttekening te worden geplaatst. De analyses zijn alle uitgevoerd op individueel niveau. Er worden geen gegevens gepresenteerd op groepsniveau en/of bij de analyses is geen rekening gehouden met de groepsstructuur.

Criteriumvaliditeit

Er is geen onderzoek uitgevoerd.