

COTAN Beoordeling 2021 van de AMN Eindtoets 2022
(handleiding uit 2021)

1. Theoretische uitgangspunten:	goed
2. Kwaliteit van het testmateriaal:	goed
3. Kwaliteit van de handleiding:	goed
4. Normen:	n.v.t. ¹
5. Betrouwbaarheid:	goed
6. Begripsvaliditeit:	voldoende
7. Criteriumvaliditeit:	voldoende

¹ De AMN Eindtoets kent twee normeringen: 1. normering t.a.v. toetsadvies, 2. normering t.a.v. referentieniveaus. Beide normeringen worden na de echte afname van de AMN Eindtoets 2022 vastgesteld door de Expertgroep Toetsen PO. Derhalve liggen deze normeringen buiten de scope van de COTAN beoordeling van de AMN Eindtoets 2022 en geldt de beoordeling ‘n.v.t.’.

*Reden bij ‘onvoldoende’:

1.	
2.	
3.	
4.	
5.	
6.	
7.	

N.B. De AMN Eindtoets 2022 is aanvullend beoordeeld op het aspect ‘Computer adaptief toetsen’; dit aspect is als ‘voldoende’ beoordeeld.

AMN EINDTOETS 2022

Uitgangspunten van de Testconstructie

De AMN Eindtoets 2022 is een adaptieve digitale eindtoets voor leerlingen van groep 8 in het primair onderwijs, specifiek voor afname in de periode van 15 april tot 29 april 2022. De toets bestaat uit de volgende drie wettelijk verplichte onderdelen c.q. deelttoetsen; ‘Lezen’, ‘Taalverzorging’ en ‘Rekenen’. Alle deelttoetsen worden via een adaptieve procedure afgenomen, met uitzondering van ‘Opzoeken’ bij de deelttoets ‘Lezen’.

De toetsmatrijs is samengesteld en ontwikkeld in overeenstemming met de volgende documenten; ‘Algemeen deel toetswijzer voor Eindtoets PO’ (2014), ‘Toetsbesluit PO’ (2014), ‘Referentiekader taal en rekenen’ (2009) en ‘Beoordelingskader Eindtoetsen’ (2020). De opgaven zijn geselecteerd op basis van de analyse van gegevens die zijn verzameld met de proeftoets in de periode van 1 november 2020 tot 13 maart 2021. Daarnaast zijn per deelttoets de verplichte 20 opgaven van de gezamenlijke ankertoets opgenomen. De toetsmatrijzen laten zien dat bij elk van de drie deelttoetsen een brede dekking is gerealiseerd overeenkomstig het referentiekader. Bij ‘Rekenen’ zijn items uit alle domeinen opgenomen (d.w.z. ‘getallen’, ‘verhoudingen’, ‘meten & meetkunde’, ‘verbanden’). Voor elk domein zijn items opgenomen die te plaatsen zijn onder één van de drie subdomeinen (d.w.z. ‘notatie, taal en betekenis’, ‘met elkaar in verband brengen’ en ‘gebruiken’). Ook is voldaan aan de voorschriften met betrekking tot de verhouding tussen contextloze en contextrijke items. De items zijn van de niveaus ‘onder 1F’, ‘1F’, ‘1S’ en ‘2F’. Bij ‘Lezen’ wordt onderscheid gemaakt naar het soort tekst (zakelijke teksten - waarbij worden onderscheiden informatieve, instructieve en betogende teksten - en fictionele, narratieve en literaire teksten). De items behorend bij de teksten, toetsen verschillende taken: techniek en woordenschat, begrijpen, interpreteren, samenvatten en opzoeken. De items zijn van de niveaus ‘onder 1F’, ‘1F’ en ‘2F’. Bij ‘Taalverzorging’ zijn items opgenomen met betrekking tot categorieën van spellingsregels, regels voor lettergreepgrenzen, regels voor woordgrenzen, morfologische spelling, overige regels en leestekens. De items zijn van de niveaus ‘1F’, ‘2F’ en ‘3F’. Bij de constructie van de opgaven en de samenstelling van de proeftoets is gebruik gemaakt van ervaren toetsconstructeurs en experts. Daarnaast biedt het inhoudelijk kader voldoende structuur en controles om ervan te mogen uitgaan dat de items inhoudelijk aan de eisen voldoen.

De AMN Eindtoets 2022 kent twee functies, namelijk 1. een niveauadvies met betrekking tot het vervolgonderwijs en 2. een niveaubepalende functie in relatie tot de referentieniveaus Rekenen en Nederlandse taal (onderverdeeld in ‘Lezen’ en ‘Taalverzorging’). Het schooladvies van de leerkracht wordt alleen heroverwogen wanneer de toetsuitslag wijst op een hoger doorstroomniveau dan door de leerkracht aangegeven.

De toets is geschikt voor alle leerlingen uit groep 8 inclusief leerlingen binnen het speciaal basisonderwijs, met uitzondering van de leerlingen die voldoen aan de omschrijving zoals vermeld in de wet hierover (leerlingen die uitstromen naar het voortgezet speciaal onderwijs binnen de profielen arbeidsmarkt of dagbesteding; leerlingen met een IQ lager dan 75; leerlingen die 4 jaar of korter in Nederland zijn en de Nederlandse taal onvoldoende beheersen; leerlingen die zeer moeilijk leren of meervoudig gehandicapt zijn en waarvan zeer moeilijk leren een van de handicaps is).

Kwaliteit van het Testmateriaal

Papier-en-potloodversie

N.v.t.

Computerversie

De toetsopgaven zijn vrij van kwetsende inhoud en zijn gestandaardiseerd qua inhoud en vorm. Omdat het adaptieve toetsen betreft is de volgorde niet voor iedereen hetzelfde, maar de keuze van de vervolgitems is gespecificeerd. De opgaven zijn tevens correct geformuleerd.

De scoring staat beschreven in een apart document. De scoring van de gesloten vragen is objectief. Bij ‘Rekenen’ en ‘Taalverzorging’ komen ook open vragen voor. Bij ‘Rekenen’ is in het systeem gespecificeerd welke antwoorden goed worden gerekend (bijv. zowel € 5,00 als € 5). De open vragen

TOELICHTING BIJ DE COTAN BEOORDELING

kunnen bestaan uit vragen met één antwoordveld dat ingevuld moet worden of meerdere antwoordvelden die ingevuld moeten worden. Wanneer er meerdere antwoorden moeten worden ingevuld, wordt de vraag alleen goed gerekend als alle antwoorden correct zijn. De open vragen bij ‘Taalverzorging’ betreffen alleen de items met betrekking tot werkwoordspelling. De leerlingen moeten het antwoord intypen. Hierbij is altijd één antwoord goed, waarbij niet wordt gelet op hoofdletters en kleine letters. Ook de scoring van de open vragen is derhalve objectief.

De instructie voor de leerlingen is duidelijk en compleet. De afname wordt gestart met een algemene uitleg en bij elk onderdeel krijgt de leerling eerst een instructie, waarin uitgelegd wordt wat het onderdeel inhoudt en wat voor soort vragen er zijn (inclusief een voorbeelditem). Overigens kunnen de instructies tijdens het maken van de opgaven niet meer worden geraadpleegd. In de instructie wordt gemeld dat de toets niet tijdgebonden is. Leerlingen wordt de mogelijkheid geboden om opgaven over te slaan; deze worden in eerste instantie fout gerekend en als het item bij het terugbladeren goed wordt beantwoord, wordt theta opnieuw geschat. Voor dyslectische leerlingen bestaat een voorleesoptie voor zowel de instructie als de opgaven. Bij elk scherm kunnen zij daartoe op het luidsprekericoontje klikken. De voorgelezen teksten doen wat gekunsteld aan door pauzes op ongebruikelijke momenten en vreemde klemtonen.

De software lijkt robuust en er zijn diverse maatregelen ingebouwd om het openen en gebruik van andere applicaties te voorkomen. De gegeven antwoorden worden altijd direct opgeslagen, zodat na bijvoorbeeld een technische storing, de leerling automatisch verder gaat waar hij/zij gebleven was in de toets. De vormgeving van de gebruikersinterface (d.w.z. één vraag per scherm) draagt er ook aan bij dat fouten door onjuist gebruik kunnen worden vermeden. Verder ziet de gebruikersinterface er duidelijk en prettig uit voor de leerlingen. Op het scherm staat niet meer en niet minder dan wat er moet staan en het lettertype is goed gekozen. De rekenopgaven zien er tevens aantrekkelijk uit door het gebruik van kleuren.

Afname van de AMN Eindtoets gaat via de AMN afname app. De beveiliging van de toets en (de versturing van) de data wordt uitgebreid besproken in de Verantwoording. AMN heeft onder andere het ‘Convenant Digitale Onderwijsmiddelen en Privacy – Leermiddelen en Toetsen 3.0’ onderschreven. In dit convenant is tussen aanbieders en de onderwijssectorraden vastgelegd dat een (onderwijs)instelling in juridische zin de ‘verwerkersverantwoordelijke’ is voor de verwerking van persoonsgegevens met alle rechten en plichten van dien.

Kwaliteit van de Handleiding

Alle informatie aangaande de AMN Eindtoets 2022 is opgenomen in verschillende documenten: (inhoudelijke en psychometrische) Verantwoording, Praktische Handleiding, Documentatie Automatische Scoring, Toetsreglement en Evaluatie AMN Eindtoets 2021. Deze gezamenlijke documentatie wordt beschouwd als de handleiding.

In de Praktische Handleiding en het Toetsreglement wordt informatie gegeven over de gebruiksmogelijkheden bij leerlingen met diverse beperkingen. Voor leerlingen met dyslexie is een versie van de AMN Eindtoets 2022 met auditieve voorleesondersteuning beschikbaar. Er is rekening gehouden met kleurenblindheid, door in de toets niet te verwijzen naar slecht te onderscheiden kleuren. De gehele toets kan gelezen en gemaakt worden in zwart-wittinten. Er worden geen auditieve onderdelen aangeboden, waardoor leerlingen met een auditieve beperking geen hinder ondervinden bij het maken van de toets. Wanneer leerlingen moeite hebben met het lezen van de schermen, is het toegestaan om de schermresolutie aan te passen. Het is tevens toegestaan om een aangepast beeldscherm te gebruiken, waarmee het scherm kan worden vergroot. De AMN Eindtoets is niet geschikt voor meervoudig gehandicapte leerlingen en leerlingen die korter dan 4 jaar in Nederland zijn en daardoor de Nederlandse taal niet goed beheersen. Wanneer bij een leerling sprake is van een specifieke ondersteuningsbehoefte die de AMN Eindtoets niet ondersteunt, wordt aangeraden om voor deze leerling gebruik te maken van de Centrale Eindtoets.

In de Praktische Handleiding en het Toetsreglement wordt uitgebreide en duidelijke informatie gegeven over allerlei praktische zaken met betrekking tot de afname. In de Praktische Handleiding staat bijvoorbeeld de letterlijke instructie vermeld die de toetsleider voorafgaand aan de afname moet geven. Ook wordt hierin aangegeven dat de toetsleider (d.w.z. bij voorkeur de eigen leerkracht)

TOELICHTING BIJ DE COTAN BEOORDELING

tijdens de toets geen inhoudelijke vragen mag beantwoorden en dat een woordenboek en een rekenmachine niet mogen worden gebruikt, maar kladpapier wel. De uitgebreide informatie ten aanzien van de afname zorgt er tevens voor dat de leerkracht goed op zijn/haar taak wordt voorbereid. In de Praktische Handleiding worden drie voorbeelden van een leerlingrapport weergegeven en in de Verantwoording wordt nog een vierde rapport gegeven. De scores en een korte interpretatie van de scores worden gegeven. De diverse gebruikte typen scores (d.w.z. totaalscore, percentielscores en referentieniveaus) worden kort uitgelegd en er wordt aangegeven wat de relatie is tussen de eindscore en het toetsadvies. Van een echte gevalsbeschrijving is echter geen sprake. Bij alle voorbeelden ontbreekt een typering van de leerling als persoon, met zijn/haar schoolgeschiedenis, persoonlijke omstandigheden en bijzonderheden, interesses, werkhouding en zelfstandigheid. Door een dergelijke toevoeging zou de plek van de eindtoets in het geheel van informatie over een leerling en bij de totstandkoming van de schoolkeuze levendiger en duidelijker kunnen worden. Daarmee zou tevens een beschrijving gegeven kunnen worden van soorten informatie die bij de interpretatie van belang kunnen zijn. Naast het leerlingrapport zijn een groepsrapport en een managementrapport beschikbaar. In het groepsrapport worden per leerling de referentieniveaus per deelttoets, de totaalscore en het toetsadvies gegeven. Het managementrapport toont tabellen en grafieken waarin de scores van de school worden vergeleken met de landelijke scores.

De toetsafname wordt begeleid door een toetsleider die door de directeur van de school is aangewezen. Dit is bij voorkeur de groepsleerkracht van groep 8. Mocht deze niet aanwezig zijn, dan kan de rol van toetsleider vervuld worden door de directeur of ander personeel met een onderwijsbevoegdheid voor het basisonderwijs. Er wordt geen specifieke deskundigheid vereist van de toetsleider. De interpretatie van het leerlingrapport kan eenvoudig gedaan worden door de bevoegde leerkracht. De rapportage bevat ook een algemene toelichting op gebruikte begrippen, zodat ook ouders het rapport zelfstandig kunnen lezen en scores kunnen interpreteren.

De leerling start de eindtoets op via het testafname programma 'AMN Eindtoets Client'. De instructie voor het installeren van dit programma wordt beschreven in Bijlage 5a en 5b van de Praktische Handleiding voor leerkrachten. Alle gangbare browsers voldoen. De systeemeisen worden uitvoerig beschreven. De AMN Portal is de administratieve omgeving waar de school een overzicht heeft van gegevens van de leerlingen, de aanmelding en de voortgang van de toetsen en waar de rapportages beschikbaar komen. In Bijlage 6 van de Praktische Handleiding voor leerkrachten wordt informatie gegeven over het afnemen van de toets via AMN Portal. Informatie over hoe te navigeren tussen de verschillende type rapportages is te vinden in de online help-functionaliteit van de AMN Portal. Wanneer de rapportages gereed zijn worden scholen hierover geïnformeerd en krijgen zij ook een link naar de betreffende informatiepagina binnen de online help-functie.

De servicedesk van AMN is op werkdagen tijdens kantooruren telefonisch en per e-mail bereikbaar. Tijdens de afnameperiode van de eindtoets is er een speciaal telefoonnummer beschikbaar voor vragen gerelateerd aan de eindtoets.

Normen

Op de leerlingrapportage worden het toetsadvies en het behaalde referentieniveau per deelttoets weergegeven. Bij zowel de normering ten aanzien van het toetsadvies als de normering ten aanzien van de behaalde referentieniveaus is sprake van criteriumgerichte interpretatie (d.w.z. de behaalde scores worden vergeleken met vastgestelde grensscores). Bij deze normeringen zijn/worden de te hanteren grensscores vastgesteld door de Expertgroep Toetsen PO met behulp van het gezamenlijk anker. Deze normeringen liggen daardoor buiten de scope van de COTAN beoordeling van de AMN Eindtoets 2022. Hieronder wordt wél de totstandkoming van die normeringen beschreven.

De normering ten aanzien van het toetsadvies is gebaseerd op de eindscore, die een gewogen (theta)somscore is van de deelttoetsscores. De gewichten van de deelttoetsen zijn vastgesteld aan de hand van een advies van de Expertgroep Toetsen PO voor alle eindtoetsaanbieders. Voor de AMN Eindtoets gelden de volgende gewichten: 45% voor 'Rekenen', 30% voor 'Lezen' en 25% voor 'Taalverzorging'. De gewogen (theta)somscore wordt middels een lineaire transformatie omgezet naar een eindscore met waarden tussen 300 en 500 (de percentielscore $\times 2 + 300$).

TOELICHTING BIJ DE COTAN BEOORDELING

De te gebruiken grensscores voor de toetsadviezen worden door de Expertgroep Toetsen PO vastgesteld op basis van de afnamegegevens van alle eindtoetsen in april en mei 2022. Dit kan doordat sinds 2019 ter bevordering van de vergelijkbaarheid van de eindtoetsen gebruik wordt gemaakt van het gezamenlijk anker dat in alle eindtoetsen is opgenomen.

Ook voor de normering ten aanzien van de referentieniveaus geldt dat de te hanteren grensscores vastgesteld worden door de Expertgroep Toetsen PO op basis van het gezamenlijk anker en na afname van alle eindtoetsen in april en mei 2022.

Betrouwbaarheid

In Hoofdstuk 6 van de Verantwoording worden de betrouwbaarheidsgegevens gerapporteerd.

De uitkomsten van de AMN Eindtoets zijn het toetsadvies en de behaalde referentieniveaus op de deelttoetsen (zoals gerapporteerd op het leerlingrapport). Deze uitkomsten worden bepaald door middel van grensscores. De lokale betrouwbaarheid rondom deze grensscores is derhalve het meest relevant voor het gebruik van de AMN Eindtoets en is als uitgangspunt genomen bij de beoordeling van *Betrouwbaarheid*. Daarbij wordt het geven van advies over het type vervolgonderwijs op basis van de totaalscore op de AMN Eindtoets beschouwd als een belangrijke beslissing op individueel niveau en het bepalen van de behaalde referentieniveaus op de deelttoetsen als minder belangrijke beslissing op individueel niveau.

Het betreffende onderzoek staat beschreven in paragraaf 6.1.6 van de Verantwoording. Tabel 41 rapporteert de lokale betrouwbaarheid van de grensscores voor het toetsadvies. De lokale betrouwbaarheid is .98 voor alle grensscores en wordt daarmee als ‘goed’ beoordeeld. Tabel 38 rapporteert de lokale betrouwbaarheid van de grensscores voor de referentieniveaus per deelttoets. Alle waarden liggen boven .90 en worden als ‘goed’ beoordeeld. De gehanteerde cesuren in dit onderzoek zijn de cesuren van de operationele toets van 2021 en niet van de toets van 2022 (die zijn nog niet bekend). Aangenomen mag worden dat de resultaten zeer vergelijkbaar zijn voor de cesuren van de operationele toets van 2022.

Om een indicatie te krijgen van de betrouwbaarheid van de operationele adaptieve deelttoetsen is de globale betrouwbaarheid geschat op basis van afnames van 10.000 gesimuleerde leerlingen. De waarden zijn .97 voor ‘Rekenen’, .87 voor ‘Lezen’ en .91 voor ‘Taalverzorging’. De globale betrouwbaarheid is tevens geschat op basis van de beschikbare afnamegegevens van de operationele adaptieve AMN eindtoets 2021 en de lineaire proeftoets 2022; waarbij populatie 1 ($n = 2706$) de AMN Eindtoets 2021 heeft gemaakt, populatie 2 ($n = 1440$) zowel de AMN Eindtoets 2021 als de proeftoets 2022 heeft gemaakt en populatie 3 ($n = 77$) de proeftoets 2022 heeft gemaakt. Dus de betrouwbaarheidsgegevens van populatie 1 zijn alleen gebaseerd op de adaptieve afname van de toets. De geschatte betrouwbaarheidswaarden voor deze groep zijn .97 voor ‘Rekenen’, .88 voor ‘Lezen’, .95 voor ‘Taalverzorging’ en .98 voor de totale toets. Deze waarden zijn vrijwel gelijk aan de waarden die gevonden zijn voor populatie 2 en geven waarschijnlijk een goede voorspelling van de betrouwbaarheid voor de operationele adaptieve AMN Eindtoets 2022.

Tevens vermeldt de Verantwoording de marginale betrouwbaarheid voor de deelttoetsen en de totale toets, apart voor de lineaire proeftoets ($n = 1517$) en voor de totale itembank (d.w.z. items uit de operationele toets van 2021 en de proeftoets samen; $n = 4223$). Op basis van de lineaire proeftoets zijn de waarden .95 voor ‘Rekenen’, .88 voor ‘Lezen’, .89 voor ‘Taalverzorging’ en .97 voor de totale toets. Op basis van de totale itembank zijn de waarden .99 voor ‘Rekenen’, .97 voor ‘Lezen’, .98 voor ‘Taalverzorging’ en .99 voor de totale toets. Ook worden de waarden van Guttman’s lambda2 vermeld voor de lineaire proeftoets ($n = 1517$); .95 voor ‘Rekenen’, .90 voor ‘Lezen’, .88 voor ‘Taalverzorging’ en .97 voor de totale toets. Bij deze onderzoeken dient benadrukt te worden dat ze betrekking hebben op de totale itembank of de gehele lineaire proeftoets (logischerwijs niet op de operationele adaptieve toets) en daardoor zeer waarschijnlijk een overschatting zijn van de betrouwbaarheid van de adaptieve operationele toets met minder items.

Begripsvaliditeit

In de Verantwoording worden de resultaten van onderzoeken gerapporteerd ter ondersteuning van de begripsvaliditeit van de AMN Eindtoets 2022.

TOELICHTING BIJ DE COTAN BEOORDELING

De unidimensionaliteit van elke deelttoets is onderzocht door te kijken naar de modelpassing van het gehanteerde 2-parameter logistisch model. De verschillende passingsmaten laten zien dat er voor ‘Rekenen’ en ‘Lezen’ gesproken kan worden van goede passing en voor ‘Taalverzorging’ (op basis van de waarde van de TLI en de CFI) van adequate modelpassing. In de Verantwoording worden alleen gegevens verstrekt over de modelpassing van de items uit de proeftoets 2022 en niet voor de items uit de (proef)toetsen van voorgaande jaren. Deze zijn echter door AMN alsnog verstrekt voor de proeftoetsen van 2018, 2019 en 2021. Voor deze proeftoetsen is de modelpassing voor de drie deelttoetsen ‘goed’ te noemen zowel op grond van de RMSEA als de TLI- en CFI-waarden. Er wordt niet ingegaan op de dimensionaliteit van de totaalscore bijvoorbeeld middels factoranalyse.

De correlaties tussen de deelttoetsen worden vermeld in paragraaf 6.2.4 van de Verantwoording voor de drie deelsteekproeven apart (d.w.z. op basis van alleen operationele toets 2021, operationele toets 2021 plus proeftoets 2022 en alleen proeftoets 2022). De correlaties tussen de drie deelttoetsen verschillen nauwelijks en variëren tussen .66 en .71. Deze waarden lijken in ieder geval te wijzen op een algemene factor, met ruimte voor specifieke bijdragen van de deelttoetsen. Overigens zijn de correlaties tussen de Taal onderdelen onderling (d.w.z. tussen ‘Lezen’ en ‘Taalverzorging’) niet duidelijk sterker zijn dan de correlaties tussen de ‘Taal’-onderdelen enerzijds en ‘Rekenen’ anderzijds.

In Bijlage 2 van de Verantwoording worden de psychometrische gegevens van de items uit de totale itembank vermeld. Vanuit de KTT werd er door de auteurs naar gestreefd dat de p -waarden liggen tussen .05 en .95 en de gemiddelde p -waarde per deelttoets tussen .60 en .70 ligt; de item-restcorrelaties zouden groter moeten zijn dan .20 en elk item zou moeten bijdragen aan de betrouwbaarheid. Aan al deze voorwaarden wordt grotendeels voldaan. Voor ‘Rekenen’ liggen de p -waarden tussen .06 en .96, de gemiddelde p -waarden voor de items uit de operationele toets en de proeftoets zijn respectievelijk .56 en .57 en de item-restcorrelaties liggen tussen .20 en .59. Voor ‘Lezen’ liggen de p -waarden tussen .21 en .94, zijn de gemiddelde p -waarden voor de items uit de operationele toets en de proeftoets respectievelijk .64 en .72 en liggen de item-restcorrelaties tussen .21 en .49. Voor ‘Taalverzorging’ liggen de p -waarden tussen .36 en .93, zijn de gemiddelde p -waarden voor de items uit de operationele toets en de proeftoets respectievelijk .65 en .67 en liggen de item-restcorrelaties tussen .25 en .49. Binnen elke deelttoets is derhalve een grote variatie in moeilijkheidsgraad en ‘Rekenen’ is de moeilijkste deelttoets. Vanuit de IRT werd er door de auteurs naar gestreefd geen items met een te lage discriminatieparameter (a -parameter) en geen extreem moeilijke items (hoge b -parameter) op te nemen. Grenswaarden voor deze parameters worden niet gegeven, wel worden van alle items de waarden van de a - en b -parameters vermeld.

Voor de proeftoets zijn DIF-analyses uitgevoerd met betrekking tot de variabelen ‘afnamevorm’ (pc vs. tablet), ‘dyslexie’ (wel vs. geen dyslexieverklaring) en ‘geslacht’. DIF is onderzocht met behulp van logistische regressie en als grenswaarde voor DIF werd $\Delta R^2 < .035$ gekozen. Geen enkel item bleek DIF te vertonen. DIF-analyses voor de operationele toets van 2021 waren al eerder verricht waarbij een klein aantal items DIF vertoonde.

Tevens is onderzocht wat de overeenstemming is tussen het advies op basis van de AMN Eindtoets 2021 en het advies van de leerkracht, dat onafhankelijk van het AMN-advies tot stand is gekomen. Er is exacte overeenstemming in 30% van de gevallen en de overeenstemming is 87% wanneer ook het schooltype één hoger en één lager als goed wordt beschouwd. Een versturende factor in dit onderzoek was dat de leerkrachtadviezen met 10 schooltypen werkten en de AMN-adviezen met 6 schooltypen. Wanneer de leerkrachten werd gevraagd hun advies om te zetten naar 6 schooltypen, dan werden de overeenstemmingspercentages respectievelijk 51% en 93%. Bij deze resultaten dient aangegeven te worden dat ze betrekking hebben op de AMN Eindtoets 2021. Aangezien 25% van de itembank wordt ververst en dus 75% van de operationele eindtoets 2021 onderdeel is van de itembank voor de eindtoets 2022, wordt aangenomen dat vergelijkbare resultaten gevonden zullen worden voor de AMN Eindtoets 2022. Eenzelfde redenering geldt voor de informatie in Bijlage 4 van de Verantwoording die ook betrekking heeft op de AMN Eindtoets 2021 en waaruit is af te leiden dat zoals verwacht mag worden de gemiddelde score per schooltype oploopt naarmate het schooltype hoger is.

Al met al is er voldoende ondersteuning voor de begripsvaliditeit van de AMN Eindtoets 2022.

Criteriumvaliditeit

Vanwege de AVG mogen middelbare scholen via DUO geen doorstroomgegevens van individuele leerlingen meer beschikbaar stellen. Dit maakt onderzoek naar de voorspellende waarde van de AMN Eindtoets 2022 in feite onmogelijk (bovendien zouden de resultaten daarvan pas over enkele jaren beschikbaar zijn). Daarom rapporteren de auteurs onderzoek met doorstroomgegevens van januari 2019 voor leerlingen die de (lineaire) AMN Eindtoets 2017 hebben gemaakt. In januari 2019 zaten die leerlingen inmiddels in de tweede klas van het voortgezet onderwijs. Hierbij zijn de gegevens van leerlingen die vallen onder de categorie ‘brugjaar’ niet meegenomen, waardoor het een betere criteriummaat is dan wanneer deze gegevens wel meegenomen zouden worden. Het zou overigens nog beter zijn om gegevens van de hogere leerjaren te gebruiken of het niveau waarop de leerlingen het voortgezet onderwijs hebben afgerond, maar zoals aangegeven kost dat tijd en is het niet gemakkelijk om dergelijke gegevens te verzamelen. Voor het onderzoeken van de criteriumvaliditeit zou het verder interessant kunnen zijn om de incrementele validiteit van de AMN Eindtoets bovenop het leerkrachtadvies te berekenen.

In het gerapporteerde onderzoek is gekeken naar de overeenstemming tussen het toetsadvies en de werkelijke plaatsing in het voortgezet onderwijs. Hierbij is sprake van enige mate van criteriumcontaminatie aangezien de plaatsing op het voortgezet onderwijs deels gebaseerd is op de uitkomst van de eindtoets. Uitgaande van de meer gedetailleerde door DUO verstrekte gegevens is 77% van de leerlingen terechtgekomen in het schooltype dat op basis van de AMN-score was geadviseerd (hierbij zijn de aangrenzende schooltypen als overeenstemming gerekend).

Om de onderzoeksresultaten te kunnen beoordelen is een beschrijving van de steekproef nodig, maar die informatie ontbreekt in de Verantwoording. Deze is echter alsnog door AMN verstrekt in reactie aan de COTAN. De AMN Eindtoets 2017 is door 557 leerlingen gemaakt en van 517 leerlingen was het niveau in het tweede leerjaar van het vo bekend (d.w.z. 92.8% overlap). Van deze 517 leerlingen zijn geen verdere gegevens bekend (door DUO worden geen gegevens op individueel niveau verstrekt), zodat strikt genomen niet kan worden bepaald of de groep van 517 leerlingen niet homogener of heterogener is dan de totale groep van 557 leerlingen. Echter, gelet op de overlap van 92.8% in beide groepen is het zeer aannemelijk dat mogelijke afwijkingen een minimale invloed zullen hebben op de resultaten.

Vanwege een continue doorlopende ontwikkeling met vastgestelde toetsmatrijzen voor de AMN Eindtoets, is het aannemelijk dat bovenstaande resultaten van de AMN Eindtoets 2017 kunnen worden generaliseerd naar de AMN Eindtoets van 2022.

Tevens worden doorstroomgegevens van januari 2019 verstrekt voor leerlingen die de AMN Eindtoets 2018 hebben gemaakt. Echter, omdat er slechts een paar maanden zitten tussen dit advies en het bepalen van het werkelijk gekozen schooltype en het dan nog niet duidelijk is of een leerling het betreffende niveau aankan, zijn deze gegevens niet relevant en derhalve buiten beschouwing gelaten.

Computer Adaptief Toetsen

Sinds 2018 is de AMN Eindtoets een adaptieve toets. De stappen in het ontwikkeltraject worden beschreven in paragraaf 2.3 van de Verantwoording. Een uitgebreide beschrijving van het algoritme dat gebruikt wordt bij het adaptief toetsen staat in hoofdstuk 4 van de Verantwoording. Hierbij wordt tevens ingegaan op de zogeheten ‘constraints’ waardoor de inhoudelijke dekking per deelttoets is gewaarborgd.

De kwaliteit van de nieuw ontwikkelde items wordt onderzocht middels een lineaire proeftoets, waarbij iedereen de volledige proeftoets maakte. Vanwege het grote aantal items wordt de proeftoets op meerdere momenten afgenomen (minimaal twee afnamemomenten). De proeftoets is door 1517 leerlingen gemaakt. De itemparameters van de operationele (adaptieve) toets van 2021 worden gefixeerd om zodoende de itemparameters van de items in de itembank voor de AMN Eindtoets 2022 te schatten. Uit het bijgeleverde Excel-bestand waarin per item de itemparameters en gegevens over in welk jaar en op basis van hoeveel leerlingen het item gekalibreerd is, blijkt dat ieder item door minimaal 866 leerlingen is gemaakt, dat wil zeggen ruim boven het vereiste aantal van 400 afnamen.

De kwaliteit van de link tussen de items is bepaald door de schattingsfouten van de gemiddelden te bepalen van de groep leerlingen die de operationele AMN Eindtoets 2021 heeft gemaakt en van de

TOELICHTING BIJ DE COTAN BEOORDELING

groep leerlingen die zowel de operationele AMN Eindtoets 2021 als de proeftoets heeft gemaakt. De standaardfout van het verschil tussen deze twee groepen wordt gebruikt om de mate van vertrouwen voor de *linking* te berekenen. Met een waarde van .999 kan geconcludeerd worden dat er een zeer hoge kwaliteit is van de link in het design.

In paragraaf 3.4.6 van de Verantwoording worden gegevens verstrekt over itemoverlap. Bij de AMN Eindtoets wordt gesproken van overlap wanneer een tweetal items aan minimaal 20% van steekproef wordt aangeboden. Bij een itemoverlap van minimaal 20% is het mogelijk om beide items ten opzichte van elkaar te kalibreren. De meeste items zijn met tientallen andere items verbonden.

Het *2-parameter logistic model* (2PLM) met de schattingsmethode *marginal maximum likelihood* (MML) is gebruikt. Een beschrijving van en een toelichting op de keuze voor dit model en niet een ander model ontbreekt in de Verantwoording. Deze is echter door AMN alsnog verstrekt in reactie aan de COTAN. Het gebruik van het 2PLM heeft de voorkeur van de Expertgroep Toetsen PO. Tevens staat dit model uitgebreid beschreven in de literatuur, onder andere voor de toepassing van CAT. Bovendien is de passing van dit model voor de AMN Eindtoets voldoende (zie hiervoor de toelichting bij *Begripsvaliditeit* m.b.t. unidimensionaliteit en modelpassing).

Op pagina 44 van de Verantwoording wordt een tabel gepresenteerd waarin per deeltoets voor drie verschillende stopcriteria de percentages worden gegeven voor de AMN Eindtoets 2021 (en logischerwijs niet voor de AMN Eindtoets 2022). De stopcriteria zijn: 1. SE klein genoeg, 2. maximale lengte bereikt, en 3. geschatte theta meer dan 2xSE onder of boven nul. Het laatste criterium komt bij 'Lezen' en 'Taalverzorging' bijna niet voor, bij 'Rekenen' is het verantwoordelijk voor 15%. Het is opvallend dat bij 'Lezen' en 'Rekenen' de toets vaker gestopt is vanwege het behalen van de maximale lengte - respectievelijk in 57.7% en 43.4% van de gevallen - dan doordat de SE klein genoeg is - respectievelijk in 42.3% en 41.1% van de gevallen. Deze percentages suggereren dat het adaptieve algoritme mogelijk minder optimaal werkt voor deze twee deeltoetsen. Bij 'Taalverzorging' lijkt het adaptieve algoritme beter te werken, aangezien in circa 85% van de gevallen de toets gestopt is vanwege het behalen van de vereiste betrouwbaarheid van de meting (d.w.z. SE klein genoeg). In het document 'Evaluatie AMN Eindtoets 2021' wordt daarover met betrekking tot 'Rekenen' aangegeven: "Het feit dat relatief veel leerlingen het maximale aantal items krijgen aangeboden komt voornamelijk doordat de SE nog te groot is en daardoor wordt het stopcriterium niet bereikt. Gemiddeld genomen zitten deze leerlingen echter wel heel dichtbij het stopcriterium, de afstand tot het criterium is gemiddeld namelijk 0,017." (p. 12). Het aantal items voor 'Lezen' is vaak hoog, omdat per leestekst alle items aangeboden worden.

Paragraaf 2.4 beschrijft de verversingsstrategie, waarbij gelet wordt op items die 1. niet gebruikt zijn, 2. heel vaak geselecteerd worden, 3. verschillend functioneren in de proeftoets en de operationele toets, 4. bekend zijn geworden en 5. een lage informatiewaarde hebben. Elk jaar wordt minimaal 20% van de itembank verversd. De verversingsgraad voor de AMN Eindtoets 2022 is 24%. Gedetailleerde informatie over welke items uit de proeftoets zijn afgevallen en waarom is te vinden in Bijlage 1 van de Verantwoording. Informatie over welke items verwijderd zijn uit de itembank op basis van de afname van de AMN Eindtoets 2021 staat in het document 'Evaluatie AMN Eindtoets 2021'.

N.B. De aantallen items voor de onderdelen 'Rekenen' en 'Taalverzorging' in het Excel-bestand komen niet exact overeen met de aantallen vermeld op pagina 27 van de Verantwoording.

FAIRNESS MATRIJS BIJ DE COTAN BEOORDELING

<i>Kenmerk</i>	Geslacht	Herkomst (etnisch- cultureel)	Leeftijd	Beperkingen	Regio	Urbanisatiegraad
<i>Fairness aspecten m.b.t.</i>						
Uitgangspunten testconstructie		1		2		
Testmateriaal (computer)				3		
Normen	4				5	6
Begripsvaliditeit	7			8		
Criteriumvaliditeit						

Toelichting:

1	De toets is niet geschikt voor leerlingen die korter dan 4 jaar in Nederland zijn en daardoor de Nederlandse taal onvoldoende beheersen.
2	De toets is niet geschikt voor meervoudig gehandicapte leerlingen.
3	Voor leerlingen met dyslexie is auditieve ondersteuning mogelijk. Bij de ontwikkeling van de toets is rekening gehouden met dyslectische leerlingen. Voor leerlingen met dyslexie en andere problematiek zoals dyscalculie kunnen andere voorzieningen worden getroffen.
4	Meisjes scoren significant hoger op de deoltoetsen 'Lezen' en 'Taalverzorging' en jongens scoren significant hoger op de deoltoets 'Rekenen'. Gemiddelden en spreidingen van de subgroepen worden echter niet verstrekt.
5	Leerlingen uit regio Oost scoren het laagst en leerlingen uit regio Zuid scoren het hoogst.
6	Leerlingen uit de vier grootste gemeenten scoren het laagst en leerlingen uit de middelgrote gemeenten scoren het hoogst.
7	Geen enkel item vertoont DIF op basis van 'geslacht'.
8	Geen enkel item vertoont DIF op basis van 'dyslexie'.

MATERIAALVERANTWOORDING BIJ DE COTAN BEOORDELING

Deze COTAN beoordeling heeft betrekking op:

AMN (2021). *AMN Eindtoets 2022. Verantwoording. Versie juli 2021.* Arnhem: AMN.

AMN (2021). *AMN Eindtoets 2022. Praktische handleiding voor leerkrachten. Versie juli 2021.* Arnhem: AMN.

AMN (2021). *AMN Eindtoets 2022. Toetsreglement. Versie juli 2021.* Arnhem: AMN.

AMN (2021). *AMN Eindtoets 2021. Evaluatie: vergelijking pre-test en officiële afname. Versie juli 2021.* Arnhem: AMN.

AMN (2021). *AMN Eindtoets 2022. Documentatie automatische scoring. Versie juli 2021.* Arnhem: AMN.

Deze materialen zijn door de secretaris van de Expertgroep toetsen PO bij de COTAN aangeleverd op 3 augustus 2021.

Copyright COTAN documentatie

<http://www.cotandocumentatie.nl/>

© 2009-2022, NIP/COTAN/de auteurs, p/a Boom test uitgevers, Amsterdam. Alle auteursrechten en databankrechten ten aanzien van de inhoud van deze uitgave worden uitdrukkelijk voorbehouden. Deze rechten berusten bij Boom test uitgevers c.q. de desbetreffende makers.

Voor het overnemen, opslaan en verspreiden van (delen van) de inhoud en gebruik van de vormgeving, op welke wijze dan ook, dient u vooraf schriftelijke toestemming te hebben verkregen van de uitgever, behoudens beperkingen bij de wet gesteld. *No part of this publication may be reproduced in any way whatsoever without the written permission of the publisher*