

1. Uitgangspunten van de toetsconstructie

Bij onderstaande beoordeling van de kwaliteitsaspecten met bijbehorende codes van het voornoemde beoordelingskader worden passages uit de Handleiding en Verantwoording veelal letterlijk vermeld. De Handleiding en Verantwoording bespreken de uitgangspunten van de toets, waaronder de afname, scoring, normering, interpretatie, rapportage en foutenanalyse en de psychometrische kenmerken van de toets.

Algemeen

De Schoolvaardigheidstoets (SVT) Technisch Lezen is onderdeel van het leerlingvolgsysteem van Boom test uitgevers. Het leerlingvolgsysteem van Boom test uitgevers Onderwijs bestaat uit de Schoolvaardigheidstoetsen Technisch Lezen, Begrijpend Lezen, Hoofdrekenen, Rekenen-Wiskunde en Spelling. Met de schoolvaardigheidstoetsen worden de leerlingen vanaf groep 3 tot en met 8 op de belangrijkste vaardigheden gevolgd.

De SVT Technisch Lezen is een tempotest en kent twee versies: Leeskaart A 'De kat' en Leeskaart B 'De mug'.

Meetpretentie

De SVT Technisch Lezen meet de vaardigheid in het technisch lezen van leerlingen, zoals wordt aangeleerd in het Nederlandse basisonderwijs. Daarbij gaat het zowel om de nauwkeurigheid als het tempo van het technisch lezen.

Doelgroep

De SVT Technisch Lezen is bestemd voor leerlingen in groep 3 tot en met 8 van het Nederlandse basisonderwijs.

Gebruiksdoel en functie

Het gebruiksdoel van de SVT Technisch Lezen is vierledig:

- De toets kan worden ingezet voor niveaubepaling van de leerlingen, zowel individueel als per groep.
- De toets kan worden ingezet voor de voortgangsbepaling van de leerlingen.
- De toets kan tevens een rol spelen bij het tijdig signaleren van leeszwakke leerlingen in de eerste schooljaren, waarin het technisch lezen systematisch wordt aangeleerd en inge oefend.
- De toets kan ten slotte dienen als indicator van een minimumvereiste van 'functionele geletterdheid'.

Inhoudelijke theoretische inkadering:

De (leestechische) opbouw van de tekst is gebaseerd op de drie fundamentele principes die aan ons schrift ten grondslag liggen. Dit zijn:

1. het alfabetische of het fonemische principe;
2. het morfologische principe;
3. het etymologische principe of het principe van de woordherkomst.

Inhoud van het toetspakket

Het toetspakket bestaat uit:

- Handleiding en Verantwoording
- Instructiekaart

Beoordeling van LVS toets SVT Technisch Lezen – Boom test uitgevers

- Leeskaart A 'De kat'
- Leeskaart B 'De mug'
- Scoreformulier A 'De kat'
- Scoreformulier B 'De mug'
- Informatieblad voor ouders/verzorgers
- Voorbeeld van een Individueel rapport (3x)
- Voorbeeld van een Individueel overzicht (2x)
- USB-stick met excelbestand 'Normeringsmodel SVT Technisch Lezen 2018'. Hierin staan de gegevens over het normeringsmodel van de toets voor de nieuwe normen.

2. Beoordeling van de kwaliteitsaspecten

De beoordeling vindt plaats volgens het 'Beoordelingskader voor de psychometrische aspecten van (reeksen van) toetsen uit leerlingvolgsystemen (LVS)', zoals opgesteld door de Expertgroep Toetsen PO. De Expertgroep Toetsen PO wordt gevormd door Prof. Dr. Cees Van der Vleuten (voorzitter), Prof. dr. Cees Glas (psychometrisch expert), Dr. Desiree Joosten-Ten Brinke (onderwijskundig expert) en mevrouw Pauly K. Berding-Oldersma MSc (secretaris).

De kwaliteit van de steekproef

S1.1. Is de steekproef representatief?

Bevindingen:

Informatie over de normeringssteekproef wordt in Hoofdstuk 6 van de handleiding gepresenteerd. In Tabel 6.1 staat dat de normeringssteekproef in totaal 15 scholen, 93 klassen en 2168 leerlingen telt. De representativiteit van de steekproef is op schoolniveau onderzocht met betrekking tot schoolgrootte, regionale spreiding, stedelijkheid, statuscode postcode, denominatie en gemiddelde eindtoetscore. De representativiteit van de steekproef is op leerlingniveau onderzocht met betrekking tot sekse, leeftijd, achtergrond en leerlinggewichten. De representativiteit voor zowel de variabelen op school- als leerlingniveau is beoordeeld door middel van Cramers V en Cohen's *d*.

Omdat er bij de verdelingen van de didactische leeftijden (DL's) enige afwijkingen waren ten opzichte van de verdelingen in de populatie voor de regio (Noord, West, Oost en Zuid) en achtergrond (Nederlandse; Migratie, westers; Migratie, niet-westers) van de leerlingen, is voor deze variabelen gewogen. Tabel 6.18 geeft de populatie en de gewogen normeringssteekproef voor de variabele regio en Tabel 6.19 geeft de populatie en de gewogen normeringssteekproef voor de variabele achtergrond van de leerlingen. Door middel van weging kan de oververtegenwoordiging van leerlingen met een migratieachtergrond redelijk goed worden gecorrigeerd, hetgeen ook geldt voor de oververtegenwoordiging van de regio West. Maar ook na weging blijft de regio Zuid in de steekproef enigszins ondervertegenwoordigd voor de hogere DL's (groep 7 en groep 8) ten gunste van leerlingen uit de regio West. Hierdoor zijn ook de leerlingen met een migratieachtergrond (Westers en Niet-Westers) voor de normeringsmomenten in de gewogen steekproef lichtelijk oververtegenwoordigd ten opzichte van de populatie. Voor het overige worden de streefwaarden van de percentages na weging goed benaderd.

In paragraaf 6.7.3 wordt terecht aandacht besteed aan het feit dat de steekproef gestratificeerd is, waardoor er in feite sprake is van een multilevel probleem in de vorm van schoolafhankelijkheid en er mogelijke substantiële design effecten kunnen ontstaan. Er zal dus gecontroleerd moeten worden op schoolafhankelijkheid. Om deze reden worden er ICC's (Intra Klasse Correlaties) berekend per DL (Didactische Leeftijd) en over de totale normeringssteekproef in Tabel 6.24. Over de totale normeringssteekproef bleek er geen sprake van schoolafhankelijkheid ($ICC < .05$). Voor sommige normeringsmomenten bleken de ICC's weliswaar hoger, maar nog steeds laag. De ICC's dalen met de leerjaren; in de onderbouw is nog sprake van een lichte schoolafhankelijkheid terwijl deze in de bovenbouw nagenoeg verdwijnt. Waarschijnlijk is dit te verklaren doordat scholen verschillen in de onderwijsmaand waarin ze in groep 3 het technisch leesonderwijs gaan intensiveren.

Bij het opstellen van het continue normeringsmodel werd gecontroleerd voor een eventueel effect van de school en de groep. Daartoe werd op de gegevens van de normeringssteekproef het normeringsmodel toegepast, zoals beschreven in paragraaf 6.7.1, maar uitgebreid met de variabelen van basisschool, interactie basisschool en normeringsmoment (DL) en interactie basisschool en vaardigheidsscore om te bekijken of deze variabelen een toegevoegde waarde hebben bij de voorspelling van de z-score. Dit bleek echter niet het geval te zijn, waaruit de conclusie kan worden getrokken dat schoolafhankelijkheid geen relevante factor is bij het opstellen van het normeringsmodel. Ook hoeven de standard errors van de gemiddelden, zoals vermeld in Tabel 6.20 en Tabel 6.21, daarom niet te worden gecorrigeerd.

Conclusie:

Op aspect S1.1 wordt aan de Schoolvaardigheidstoets (SVT) Technisch Lezen groep 3-8 het oordeel '**voldoende**' toegekend.

S1.2. In geval van een onvolledig dataverzamelingsdesign: is het design adequaat?

Bevindingen:

Bij de SVT Technisch Lezen is er sprake van twee parallelle toetsen, Leeskaart A 'De kat' en Leeskaart B 'De mug'. Op pagina 59 van de handleiding staat dat de twee versies binnen een periode van twee weken zijn afgenomen. In Tabel 5.1 staan de aantallen leerlingen uit de groepen 3-8 waarbij de twee versies afgenomen zijn. Groep 8 bestaat uit het grootste aantal leerlingen, namelijk 401, en groep 4 uit het kleinste aantal leerlingen, namelijk 203. In totaal werden de twee toetsen bij 1598 leerlingen afgenomen.

De twee versies meten dezelfde vaardigheid, maar uit Tabel 5.1 blijkt dat bij de meeste groepen Leeskaart B 'De mug' moeilijker is dan Leeskaart A 'De kat'. Dit verschil treedt met name op in de groepen 3 t/m 6. In groep 7 en 8 liggen zowel de gemiddelden als de standaarddeviaties van beide versies zeer dicht bij elkaar. Dit blijkt ook uit de berekende effectgroottes (Cohen's *d*). Deze duiden op een klein effect in groep 3 t/m groep 6 en op een minimaal effect in groep 7 en 8. Maar de verschillen zijn klein; over de totale normgroep gaat het om een verschil van zo'n 7 punten. Met name voor de groepen 5 en 6 is dit verschil groter en gaat het om een verschil van 10 punten; Cohen's *d* is voor deze twee groepen zo'n 0.3. In Tabel 5.2 worden de correlaties tussen beide versies gepresenteerd. De correlatie over de totale steekproef afgerond is 0.97. De correlaties blijken ook op de verschillende normeringsmomenten hoog te zijn.

Via equivalering met de equipercntiel methode werd voor de verschillen tussen de beide versies gecorrigeerd. Kleine onregelmatigheden, die waarschijnlijk moeten worden toegeschreven aan steekproeffluctuaties, worden bij deze methode vereffend door middel van loglineaire pre-smoothing. De resultaten van de equivalering van de toetsscores worden besproken en vermeld in paragraaf 5.3. In Tabel 5.5 worden de gemiddelden en standaarddeviaties gegeven na equivalering met de equipercntiel methode met loglineaire pre-smoothing, uitgesplitst naar de verschillende normeringsmomenten. De gemiddelde verschillen en standaarddeviaties na equivalering van de beide versies zijn minimaal voor zowel de totale normgroep als voor de verschillende normeringsmomenten. De toetsaanbieder gaat er hierom dan ook vanuit dat versie A en versie B na equivalering als psychometrisch evenwaardig kunnen worden beschouwd.

Conclusie:

Het cross-sectionele dataverzamelingsdesign dat hier toegepast is, is adequaat. Op aspect S1.2 wordt aan de SVT Technisch Lezen groep 3-8 het oordeel '**voldoende**' toegekend.

Normering

N1.2.1. Zijn de normgroepen groot genoeg?

Bevindingen:

In Tabel 5.1 staan voor de twee versies de aantallen leerlingen van groep 3-8 vermeld. De 6 groepen tellen in totaal 1598 leerlingen met een range van 203 – 401 leerlingen.

In paragraaf 6.7 van de handleiding wordt gewezen op de wenselijkheid om de toetsen ook buiten de normeringsmaand te gebruiken. Door middel van continue normering heeft men normen voor de tien verschillende maanden per leerjaar opgesteld. In grafiek 6.1 wordt de continue normering van de SVT Technisch Lezen grafisch weergegeven, waarbij de functies van de verschillende normeringsmomenten van links naar rechts oplopen (beginnend bij DL = 3 en eindigend met DL = 60).

Er zijn normen geconstrueerd voor de vaardigheidsscores en de accuratessescores. De normen voor de vaardigheidsscores betreffen genormaliseerde standaardscores met een gemiddelde gelijk aan 5 en een standaarddeviatie van 2 (zogenaamde C-schaal). Voor de accuratesseschaal is gekozen voor een zeer grove normeringsschaal van drie klassen: laag tot zeer laag (1), beneden gemiddeld (2) en gemiddeld tot zeer hoog (3). De eerste twee categorieën zijn hierbij van gelijke grootte (5%) en de restcategorie, voor leerlingen die accuraat lezen, is 90% (zie Tabel 4.2). In deze opzet zijn de grote afwijkingen van normaliteit van de verdelingen (zeer sterk scheef naar links, zodat het weinig zin heeft een onderscheid te maken tussen 'gemiddelde' accuratesse en 'hoge' accuratesse) niet bezwaarlijk, omdat de normering beperkt blijft tot een indeling in deze drie klassen en voornamelijk tot doel heeft een onderscheid te maken in leerlingen die relatief veel fouten maken en leerlingen die nauwelijks tot geen fouten maken bij technisch lezen.

Conclusie:

Op aspect N1.2.1 wordt aan de SVT Technisch Lezen groep 3-8 het oordeel '**voldoende**' toegekend.

N1.2.2. Zijn de normgroepen representatief?

Bevindingen:

De representativiteit werd in S1.1 van voorliggend document besproken en daar werd geconstateerd dat de normgroepen representatief waren voor relevante variabelen op school- en leerlingniveau.

De normen voor de SVT Technisch Lezen zijn verzameld in de periode november 2016 - juni 2018. De COTAN hanteert een periode van 15 jaar waarin normen geldig zijn. Dit betekent dat de normen geldig zijn tot november 2031. Wanneer uit een tussentijdse

steekproef significante verschuivingen blijken, kan dit echter leiden tot eerdere hernormering van de toets.

Conclusie:

Op aspect N.1.2.2 wordt aan de SVT Technisch Lezen groep 3-8 het oordeel '**voldoende**' toegekend.

Betrouwbaarheid

B1.1. Zijn of worden de betrouwbaarheidsgegevens correct berekend?

Bevindingen:

In Hoofdstuk 7 van de handleiding worden gegevens over de betrouwbaarheid vermeld. Zowel parallelle als test-hertest betrouwbaarheden zijn berekend (betrouwbaarheid aan de hand van interne consistentie kon niet worden berekend, omdat de SVT Technisch Lezen een tempotest is). De betrouwbaarheden zijn correct berekend, d.w.z. de parallelle betrouwbaarheid (in 2017 opnieuw berekend nadat deze in 2002 ook al was berekend) als zijnde de correlatie tussen Leeskaart A 'De kat' en Leeskaart B 'De mug' (in Hoofdstuk 5 was immers al geconcludeerd dat de beide versies als parallelle toetsvormen konden worden beschouwd) en de test-hertest betrouwbaarheid (in 2002 berekend) als zijnde de correlatie tussen Leeskaart A 'De kat' op een eerste afname en dezelfde Leeskaart A op een tweede afname (Leeskaart A werd in 2002 dus tweemaal afgenomen bij dezelfde steekproef). De tweede afname vond plaats onmiddellijk na de eerste afname. Omdat scholen vaak moeilijk te motiveren zijn om mee te doen aan een test-hertest onderzoek, is er destijds in 2002 voor gekozen het onderzoek te beperken tot de groepen 4, 6 en 8 om zodoende de toetsoverlast voor scholen te beperken. Hierbij werd er destijds van uitgegaan dat de test/hertest betrouwbaarheid van de toets voor de tussenliggende groepen niet wezenlijk zou verschillen.

Conclusie:

Op aspect B.1.1 wordt aan de SVT Technisch Lezen groep 3-8 het oordeel '**voldoende**' toegekend.

B1.2. Zijn de betrouwbaarheidsgegevens voldoende gezien de beslissingen die met de toets genomen worden?

Bevindingen:

De parallelle betrouwbaarheden in 2017 voor de zes afzonderlijke groepen zijn hoog (voor alle groepen > .91) en voor de totale steekproef van 1598 leerlingen zeer hoog, namelijk .974 (zie Tabel 7.1). De hoogte van de coëfficiënten in 2017 komen sterk overeen met de destijds berekende waarden in 2002 (varieerden destijds in 2002 tussen .88 en .94 met een parallelle betrouwbaarheid van .96 over de totale steekproef van 2591 leerlingen) en liggen daar zelfs nog iets boven. Als verklaring draagt de toetsaanbieder hiervoor aan dat de parallelle betrouwbaarheid in 2017 is gebaseerd op de correlatie tussen de vaardigheidsscores en niet, zoals destijds in 2002, op de correlatie tussen de ruwe toetsscores.

De destijds in 2002 berekende test-hertest betrouwbaarheid (stabiliteit) over de gehele onderzoeksgroep van 311 leerlingen is bijna gelijk aan die van de parallelle betrouwbaarheid in 2017, namelijk .98, en voor de afzonderlijke groepen 4, 6 en 8 respectievelijk .96, .96 en .95 (zie Tabel 7.3).

Paragraaf 7.4 bespreekt de classificatie consistentie aangaande de twee niveau-indelingen die gebruikt worden, namelijk niveau A-E en niveau I-V. In Tabel 7.4 en Tabel 7.5 worden de kruistabellen gepresenteerd voor respectievelijk de niveau-indelingen A-E en I-V voor de normeringssteekproef 2017 met in totaal 1598 leerlingen. De overeenstemmingsmaat P (gedefinieerd als de proportie van leerlingen die op consistente wijze worden geclassificeerd bij twee opeenvolgende toetsafnames) en Cohen's Kappa, waarbij gecorrigeerd wordt voor toevalsconsistentie. Aangezien in de normeringssteekproef 2017 beide versies van SVT Technisch Lezen bij een groot aantal leerlingen zijn afgenomen ($n = 1598$), zijn er voldoende gegevens beschikbaar om de classificatie consistentie te berekenen.

Op basis van Tabel 7.4 en Tabel 7.5 stelt de auteur op pagina 94 dat uit de overeenstemmingsmaat P kan worden afgeleid dat in ongeveer zeven van de tien gevallen bij een herhaalde toetsafname een leerling wordt ingedeeld in hetzelfde niveau (Niveau A-E en niveau I-V). Ondanks de hoge betrouwbaarheden van de toets (op basis waarvan mag aangenomen worden dat het percentage misclassificaties niet groot zal zijn), kent de niveau-indeling van de leerling dus nog wel enige variatie. Uit Tabel 7.4 en Tabel 7.5 blijkt dat het vrij zeldzaam is dat een leerling bij een hertest twee niveaus hoger of lager wordt ingedeeld. Een indeling van een leerling bij een hertest één niveau hoger of lager komt echter wel vrij vaak voor. Zoals bijv. uit Tabel 7.5 kan worden afgeleid, blijkt van de leerlingen die op basis van een score op Leeskaart A 'De kat' niveau-IV krijgt toegekend, bij afname van Leeskaart B 'De mug' 18% niveau V en 23% niveau III krijgt toegekend, terwijl 63% van de leerlingen hetzelfde niveau IV toebedeeld krijgen.

Conclusie:

Op aspect B1.2 wordt aan de SVT Technisch Lezen groep 3-8 het oordeel '**voldoende**' toegekend.

Validiteit

V1. Dragen de items in de toets bij aan de validiteit van de toets (hierbij gaat het om aspecten als relevantie, objectiviteit en efficiëntie van de items)?

Bevindingen:

De beoordeling van de validiteit beperkt zich hier tot het statistisch/psychometrisch onderzoek dat verricht is inzake de validiteit en dan met name de begripsvaliditeit.

In Hoofdstuk 8 van de handleiding, getiteld 'Begripsvaliditeit', wordt een onderbouwing gegeven van de begripsvaliditeit van de SVT Technisch Lezen. Argumenten die worden aangedragen hebben betrekking op de lineaire ontwikkeling van het technisch lezen gedurende de gehele basisschool (zie Grafiek 8.1). Bij groepsverschillen is gekeken naar de verschillen voor sekse en achtergrond van de leerlingen.

Op basis van eerder onderzoek (nationaal en internationaal) werd verwacht dat meisjes gemiddeld beter zouden scoren op de toets dan jongens. Wanneer we naar de totale normeringssteekproef kijken, blijkt uit Tabel 8.1 inderdaad dat meisjes hoger op de toets scoren dan jongens. De verschillen treden op bij alle groepen in het basisonderwijs; Cohen's d laat een kleine tot middelgrote effectgrootte zien. Het verschil tussen jongens en meisjes blijkt met name in de bovenbouw (groep 6-8) op te treden, terwijl er in de onderbouw (groep 3-5) geen significante verschillen werden gevonden. Geconcludeerd kan dan ook worden dat meisjes vanaf het begin van het basisonderwijs een voorsprong hebben op jongens en dit verschil lijkt zich te stabiliseren in de bovenbouw.

Verder werden er lagere scores verwacht op de SVT Technisch Lezen bij de leerlingen met een niet-westerse migratieachtergrond in vergelijking met leerlingen met een Nederlandse achtergrond, omdat deze groep vaak een leerachterstand heeft bij de schoolse vaardigheden. Opvallend genoeg laat Tabel 8.2 echter zien dat deze verwachtingen niet werden bevestigd. Leerlingen met een westerse migratieachtergrond werden ook vergeleken met leerlingen met een Nederlandse achtergrond. Tabel 8.3 laat hier wel een significant effect zien ten gunste van de leerlingen met een Nederlandse achtergrond over de totale steekproef. We dienen ons hierbij echter wel te realiseren dat de eerste onderzoeksgroep van relatief kleine omvang is en de verschillen bovendien niet groot en eenduidig zijn.

Voor de convergente validiteit is in het kader van het normeringsonderzoek van 2002 destijds de relatie onderzocht met verschillende 'soortgenoten'. Nadat eerst Leeskaart A 'De kat' was afgenomen, werd daarnaast een 'soortgenoot' aan de leerling voorgelegd. Deze soortgenoten waren toetsen voor technisch lezen die destijds het meest gebruikelijk waren in het basisonderwijs, namelijk de AVI, de DMT en de EMT. Tabel 8.4 toont de correlaties voor de afzonderlijke groepen en de verschillende 'soortgenoten'. Ondanks de relatief kleine onderzoeksgroepen, waren de correlaties zoals verwacht hoog (variërend tussen .78 en .94) en vormden als zodanig een bevestiging van de soortgenotvaliditeit (convergente validiteit) van de SVT Technisch Lezen.

Om de effectiviteit (i.e., behandel-effecten) van een dyslexiebehandeling te meten, is onlangs de SVT Technisch Lezen ook afgenomen bij een groep dyslectische leerlingen. In een eerste onderzoek bleek er inderdaad na gemiddeld 20 behandelingen een klein effect in de groei van de vaardigheidsscore van de SVT Technisch Lezen op te treden (zie Tabel 8.8). De 'winst' in de behandeling ligt op ongeveer een halve standaarddeviatie in vaardigheidsscore, hetgeen een middelmatig effect is. Nader onderzoek moet uitwijzen of dit effect na meerdere behandelingen over een langere periode aanhoudt (retentie-effect) en eventueel toeneemt.

Bevindingen vanuit onderwijskundig perspectief

Aan bovenstaande beoordeling van validiteit van de toets kan vanuit onderwijskundig perspectief worden toegevoegd dat er een voldoende passende onderbouwing geleverd wordt van gemaakte keuzes m.b.t. vorm en inhoud van het materiaal. De gekozen operationalisering is tevens herkenbaar en voldoende onderbouwd.

Conclusie:

Op aspect V1 wordt aan de SVT Technisch Lezen groep 3-8 het oordeel '**voldoende**' toegekend.

Het volg-aspect

VA1.1. Is er een voldoende empirische onderbouwing van de schaal waarop de groei van een leerling wordt uitgedrukt? Wordt groei op een adequate manier gemeten?

Bevindingen:

Onderscheiden worden, zie Hoofdstuk 4 van de handleiding, de toetsscore (= ruwe score = aantal binnen de minuut goed gelezen woorden) en de accuratessescore (= percentage correct gelezen woorden). De toetsscore is voor de interpretatie leidend.

In Grafiek 4.1 wordt op basis van empirische gegevens, verkregen met behulp van het eerdere vernoemde cross-sectionele dataverzamelingsdesign, het ontwikkelingsverloop aangegeven van de toetsscore en de accuratessescore. Uit de grafiek blijkt dat de gemiddelde toetsscore zich door de leerjaren heen nagenoeg lineair ontwikkelt. Verder laat Grafiek 4.1 zien dat de accuratesse in groep 3 (DL = 3 en DL = 10) aanvankelijk laag is, maar dat de accuratesse zich zeer sterk verbetert in groep 4 (DL = 13). Daarna blijft de accuratesse van het technisch lezen op een zeer hoog niveau. Met andere woorden, de gemiddelde leerling maakt vanaf groep 5 (DL = 23 e.v.) weinig tot geen fouten meer in het technisch lezen. Elke toetsscore die op een bepaalde toets wordt behaald, correspondeert met een vaardigheidsscore die op een vaardigheidsschaal ligt met een gemiddelde van 5 en een standaarddeviatie van 2 (zogenaamde C-schaal). De toetsaanbieder geeft als argument om voor de C-schaal te kiezen dat andere toetsen van Boom ook deze schaal gebruiken en dus goed kan worden vergeleken met de overige schoolvaardigheidstoetsen van Boom. De motivatie voor deze transformatie is dus vooral een praktische motivatie en heeft geen psychometrische redenen.

Conclusie:

Op aspect VA1.1 wordt aan de SVT Technisch Lezen groep 3-8 het oordeel '**voldoende**' toegekend.

VA1.2. Worden er gegevens verstrekt over hoe groei geïnterpreteerd dient te worden? Wordt de betrouwbaarheid van de groei op die schaal adequaat weergegeven?

Bevindingen:

Informatie over groei en de interpretatie van groei staat in de paragrafen 4.3 en 4.4 plus bijbehorende losbladige informatie. Paragraaf 4.3 betreft het individueel rapport waarin van de leerling (numeriek en grafisch) vermeld staan: Score, DLE, Accuratessescore, Accuratesse norm (deze kent zoals in paragraaf 4.2.2 en in Tabel 4.2 aangegeven, de drie verschillende klassen 'laag tot zeer laag', 'beneden gemiddeld' en 'gemiddeld tot zeer hoog'), AVI-niveau (daarmee wordt het niveau van de leesvaardigheid aangegeven, dat voor het kunnen lezen van de tekst is vereist; zie Tabel 4.1), Percentiel, Niveau I - V, Niveau A - E, vaardigheidsscore met 90% betrouwbaarheidsinterval. Paragraaf 4.4 betreft het individueel overzicht, waarmee de ontwikkeling van een leerling beter in beeld kan worden gebracht en een eventuele achterstand tijdig kan worden gesignaleerd. In dit overzicht staat van de leerling: Versie van de toets, Afnamedatum, DL, Score, Percentiel, DLE, vaardigheidsscore met 90% betrouwbaarheidsinterval. Daarnaast bevat het overzicht van de SVT Technisch Lezen een grafiek met op de X-as de DL en op de Y-as de vaardigheidsscore met bijbehorend betrouwbaarheidsinterval.

Conclusie:

Op aspect VA1.2 wordt aan de SVT Technisch Lezen groep 3-8 het oordeel '**voldoende**' toegekend.

Inzicht in leervorderingen

I1. Levert de toetsaanbieder een format voor een geschreven toelichting bij de leervorderingen van de leerling die (ook) voor ouders/voogden/verzorgers begrijpelijk is?

Bevindingen:

In de handleiding zijn vijf casussen opgenomen waarin individuele rapporten en individuele overzichten besproken worden. Historische overzichten kunnen bij het Boom testcentrum opgevraagd worden. Hoe deze overzichten gecommuniceerd worden met ouders/verzorgenden/voogden wordt niet vermeld. Bij het testcentrum kunnen ook groepsrapporten, vergelijkbaar met het individuele rapport, aangemaakt worden. De normering van de groep is gebaseerd op dezelfde normering, die ook voor de leerlingen is gehanteerd. Daarbij wordt uitgegaan van het groepsgemiddelde.

Naast de gebruikelijke percentielscore is in de groepsrapportage ook een T-score opgenomen. De T-score is in de groepsrapportage opgenomen om daarmee de gemiddelde groepsscores te kunnen vergelijken, maar uiteraard kan hiervoor ook gebruik worden gemaakt van de percentielscore. Bij gebruik van de percentielscore dient men zich echter wel te realiseren dat het van belang is in ogenschouw te nemen dat deze niet gebaseerd is op een intervalschaal. Dit betekent bijvoorbeeld dat het verschil tussen een percentielscore van 50% en 60% (d.w.z. 60% van de normeringssteekproef heeft een lagere score en 40% een score hoger of gelijk eraan) voor de gebruiker relatief groot kan lijken, terwijl het verschil feitelijk berust op slechts enkele punten verschil in de ruwe score. Van belang is het daarom dan ook om te kijken waar de ruwe toetsscore zich in de normaalverdeling bevindt.

Conclusie:

Op aspect I.1 wordt aan de SVT Technisch Lezen groep 3-8 het oordeel '**voldoende**' toegekend.

3. Verzamelstaat

Kwaliteitsaspect	Code	Oordeel
De kwaliteit van de steekproef	S1.1	Voldoende
	S1.2	Voldoende
Normering	N1.1	Voldoende
	N1.2	Voldoende
Betrouwbaarheid	B1.1	Voldoende
	B1.2	Voldoende
Validiteit	V1.1	Voldoende
Volg-aspect	VA1.1	Voldoende
	VA1.2	Voldoende
Inzicht in leervorderingen	I1.1	Voldoende

4. Literatuurlijst

Vos de, T. (2018). *Handleiding en Verantwoording Schoolvaardigheidstoets Technisch Lezen*. Amsterdam: Teije de Vos & Boom uitgevers.