

Onderzoek Discrepantie Rekentoets vo 2F en CE wiskunde vmbo bb en kb

Eindrapportage juni 2015

Franziska van Dalen

Met medewerking van: Hendrik Straat en Paul Drijvers

Studentnummer: 3693562

Cito, Arnhem

Supervisor: Paul Drijvers

30 ECTS



Universiteit Utrecht

Freudenthal Instituut
voor Didactiek van Wiskunde en Natuurwetenschappen



Inhoudsopgave

SAMENVATTING	3
1. INLEIDING EN ONDERZOEKSVRAGEN	6
2. METHODE	8
2.1. METHODE ONDERZOEKSVRAAG 1	8
2.2. METHODE ONDERZOEKSVRAAG 2	8
2.3. METHODE ONDERZOEKSVRAAG 3	10
2.3.1. Tijdnood.....	10
2.3.2. Terugbladeren	10
2.3.3. Complexiteit en moeilijkheid	10
2.3.4. Motivatie	11
2.3.5. Taligheid	12
2.3.6. Cesuur.....	12
2.3.7. Afnamejaar 2015	12
3. RESULTATEN	13
3.1.1. Vergelijking rekentoetswijzer 2F en syllabi wiskunde examen vmbo bb en kb	13
3.1.2. Vorm en afnamecondities	14
3.2. RESULTATEN ONDERZOEKSVRAAG 2	15
3.3. RESULTATEN ONDERZOEKSVRAAG 3	16
3.3.1. Tijdnood.....	16
3.3.2. Terugbladeren	17
3.3.3. Complexiteit en moeilijkheid	17
3.3.4. Motivatie	21
3.3.5. Taligheid	23
3.3.6. Cesuur.....	23
4. CONCLUSIE & DISCUSSIE	25
5. REFERENTIES	27
BIJLAGEN	29
BIJLAGE 1: CRITERIA VOOR ITEMS	29
BIJLAGE 2: VOORBEEDEN VAN ITEMS	30
BIJLAGE 3: CONSTRUCTIE VAN EEN OPLM-SCHAAL	33
BIJLAGE 4: WAT IS EEN REKENSTAP?	38
BIJLAGE 5: OPDRACHT EXPERTS VOOR BEPALING MOEILIKHEID.....	41
BIJLAGE 6: VERGELIJKING REKENTOETSWIJZER 2F EN DE SYLLABI CE WISKUNDE VMBO BB EN KB.....	42
BIJLAGE 7: GEGEVENS REKENTOETS VO EN CE WISKUNDE VMBO BB EN KB.....	44
BIJLAGE 8: SCORES PER KOPPEL ITEMS MET GELIJKE INHOUDELIJKE COMPLEXITEIT.....	48
BIJLAGE 9: ITEMKOPPELS VOOR ANALYSE VAN TALIGHEID	50
BIJLAGE 10: UITKOMST ANALYSE VAN TALIGHEID	55

Samenvatting

Vanaf het schooljaar 2015-2016 zal de rekentoets vo een verplicht eindexamenonderdeel zijn, dat meetelt in de slaag-/zakregeling. De toets is bedoeld om ervoor te zorgen dat kandidaten een basisniveau bij rekenen bereiken, de zogeheten referentieniveaus 1F, 2F en 3F. De commissie Bosker (2014) heeft onderzocht of de operationalisering van de referentieniveaus goed is en hoe deze verder verbeterd kan worden. Deze commissie adviseerde onder andere om de discrepantie tussen de rekentoets vo en het CE wiskunde vmbo nader te onderzoeken en na te gaan wat de resultaten betekenen voor de rekentoets vo en de referentiecesuur, de slaag-/zakgrens. Om een te groot verschil tussen de rekentoets vo en het CE wiskunde te voorkomen, is inzicht in de verschillen en overeenkomsten tussen beide toetsen van belang.

Het doel van het hier beschreven onderzoek is om de volgende onderzoeksvragen te beantwoorden:

1. In welke opzichten verschillen de rekentoets vo en het CE wiskunde vmbo bb en vmbo kb?
2. Zijn er significante verschillen tussen de scores van kandidaten op de rekentoets vo en het CE wiskunde vmbo bb en vmbo kb?
3. Kunnen de bij 2. geconstateerde scoreverschillen verklaard worden uit de bij 1. gevonden verschillen?

Om de verschillen tussen de rekentoets vo en het CE wiskunde vmbo bb en kb te bepalen (onderzoeksvraag 1) zijn de rekentoetswijzer 2F en de syllabi wiskunde vmbo bb en kb geanalyseerd. Tevens zijn verschillen in vorm en afnamecondities in kaart gebracht en is er onderzocht welke items in beide toetsen kunnen worden opgenomen. Ten slotte zijn er sets van items uit beide toetsen gemaakt die in principe dezelfde rekenvaardigheid betreffen. De analyses van de scoreverschillen van kandidaten die beide toetsen hebben gemaakt (onderzoeksvraag 2) zijn gebaseerd op de scores van ongeveer 47.000 kandidaten verdeeld over 2013, 2014 en 2015 en over de leerwegsectoren vmbo bb en kb. Tot slot is de derde onderzoeksvraag beantwoord door de resultaten van de eerste en tweede onderzoeksvraag aan elkaar te koppelen.

Met betrekking tot de eerste onderzoeksvraag leidt het onderzoek tot de volgende resultaten. De gevonden verschillen tussen de rekentoets vo en het CE wiskunde vmbo bb en kb hebben met name betrekking op de vorm en afnamecondities van de toetsen. De toetsen verschillen in toetsduur, het aantal items, het al dan niet aanwezig zijn van contextloze items, het gebruik van contexten, de soort rekenmachine die beschikbaar is, de aanwezigheid van een rekenmachine bij de items en het al dan niet kunnen terugbladeren. Het al dan niet kunnen terugbladeren speelt in 2015 geen rol meer omdat het vanaf dan mogelijk is om tussen de eerste 15 contextloze items terug te bladeren en tussen de laatste 30 contextitems. De genoemde factoren hebben een rol gespeeld in de analyse van de scoreverschillen. Er is ook een verschil wat betreft de inhoud van de twee toetsen. Alle vaardigheden die bij de rekentoets vo getoetst worden, worden ook getoetst in het CE wiskunde vmbo bb en kb behalve de basisvaardigheden. Er zijn echter wel meerdere vaardigheden die op het CE wiskunde vmbo bb of kb getoetst worden die niet voorkomen bij de rekentoets vo.

Ten aanzien van de tweede onderzoeksvraag concluderen we dat er inderdaad significante verschillen zijn tussen de scores van kandidaten op de rekentoets vo en het CE wiskunde vmbo bb en vmbo kb. Als we ons beperken tot de contextitems van de rekentoets vo en de items van de centrale

examens wiskunde vmbo bb en kb die ook bij de rekentoets vo zouden kunnen voorkomen, dan blijkt dat de rekentoets vo slechter is gemaakt dan de centrale examens wiskunde vmbo bb en kb in 2013, 2014 en 2015. Dit verschil is statistisch significant. Het verschil in 2014 tussen de scores op de rekentoets vo en het centrale examen wiskunde vmbo kb is overigens erg klein (in de orde van 0,1 standaardafwijking). In 2013 en 2014 is er een middelmatige samenhang tussen de scores op de rekentoets en de centrale examens.

Als antwoord op de derde onderzoeksvraag rapporteren we de volgende bevindingen:

- *Tijdnood*

Het is bekend dat op de rekentoets vo 2013 sprake geweest kan zijn van tijdnood bij de kandidaten. Dit kan een gedeeltelijke verklaring zijn voor de scoreverschillen op de toetsen in 2013, maar in 2014 en 2015 speelt dit niet.

- *Terugbladeren*

Doordat er in 2013 mogelijk sprake was van tijdnood, en omdat juist bij tijdnood terugbladeren nodig kan zijn, heeft het al dan niet kunnen terugbladeren mogelijk effect gehad op de scores van de kandidaten. In 2014 was er echter geen tijdnood waardoor het al dan niet kunnen terugbladeren naar verwachting geen effect heeft gehad. Vanaf 2015 is het in de rekentoets vo mogelijk om terug te bladeren en dus zal deze eventuele verschilfactor dan niet meer aan de orde zijn.

- *Inhoudelijke complexiteit*

De contextitems in de rekentoets vo zijn inhoudelijk complexer – in termen van het aantal benodigde rekenstappen – dan de rekenitems in het examen wiskunde vmbo bb in 2013, 2014 en 2015 en vmbo kb in 2013. Het centrale examen wiskunde vmbo kb in 2014 en 2015 is inhoudelijk vrijwel even complex als de rekentoets vo.

- *Motivatie*

De prestatieverschillen van kandidaten op de beide toetsen kunnen in 2013 en 2014 mede verklaard worden door verschillen in motivatie aangezien (a) een aantal kb-examenitems inhoudelijk complexer is dan rekentoetsitems die door de leerlingen even goed zijn gemaakt en (b) een aantal items van de examens van vmbo bb en kb beter wordt gemaakt dan inhoudelijk even complexe items van de rekentoets:

- a) Van items van de rekentoets vo en de centrale examens wiskunde vmbo bb in 2013 en 2014 die even goed worden gemaakt, schatten experts de inhoudelijke complexiteit even hoog in. Bij kb schatten experts de inhoudelijke complexiteit van de items van het CE wiskunde vmbo kb hoger in dan die van items die even goed gemaakt zijn op de rekentoets vo.
- b) Items van de rekentoets vo en de centrale examens wiskunde vmbo bb en kb die inhoudelijk even complex zijn worden in 2013 even goed gemaakt. In 2014 geldt er dat de items van de centrale examens wiskunde vmbo bb en kb beter worden gemaakt dan de vergelijkbare items van de rekentoets vo. Hierbij valt echter wel op dat zowel in 2013 als 2014 items met veel rekenstappen in de rekentoets vo beter worden gemaakt dan items met evenveel rekenstappen in de centrale examens wiskunde vmbo bb en kb.

- *Verschillende scoring*

Als methode van goed-fout scoring van het eindantwoord, zoals die bij de rekentoets vo wordt gebruikt, ook bij de wiskunde-examens zou worden toegepast, dan worden de verschillen in gemiddelde score tussen de rekentoets vo en de centrale examens wiskunde in 2013 en 2014

kleiner. Het centrale examen wiskunde vmbo kb 2014 blijkt dan zelfs moeilijker dan de rekentoets vo.

- *Taligheid*

Op basis van een kleinschalige meting blijkt dat de items van de rekentoets vo minder talig zijn dan de items van de centrale examens wiskunde vmbo bb en kb. Omdat de taligheid van het CE wiskunde eerder hoger lijkt te zijn dan bij de rekentoets vo, lijkt dit geen factor te zijn die het gevonden scoreverschil kan verklaren.

- *Cesurbepaling*

Voor de centrale examens wiskunde vormt een referentie-examen met een bijbehorend percentage onvoldoende het uitgangspunt voor de cesuur. Voor de rekentoetsen is de cesuur bepaald door het te behalen referentieniveau. Daardoor is het lastiger om aan het referentieniveau rekenen te voldoen dan aan de prestatie-eis die gekoppeld is aan het referentie-examen wiskunde. Hoewel het CE wiskunde vmbo kb moeilijker is, is het daardoor toch eenvoudiger om een voldoende te behalen. Dit is voor de afnamejaren 2013 en 2014 onderzocht.

1. Inleiding en onderzoeksvragen

Al jaren wordt geconstateerd dat studenten die het mbo of hoger onderwijs instromen moeite hebben met lezen, spellen en rekenen. De vraag is hoe het niveau op de gebieden taal en rekenen kan worden verbeterd. Om deze vraag te beantwoorden is in 2007 de commissie Meijerink in het leven geroepen. Deze commissie heeft referentieniveaus opgesteld voor zowel taal als rekenen. Deze referentieniveaus zijn bedoeld om de beheersing van taal en rekenen op verschillende momenten in de schoolcarrière te beschrijven. Voor rekenen zijn er drie fundamentele referentieniveaus: 1F, 2F en 3F (commissie Meijerink, 2008). Hierbij wordt 2F gezien als het referentieniveau dat het algemene maatschappelijke gewenste niveau voorstelt. Dit niveau zou iedereen moeten behalen om goed te kunnen functioneren in de maatschappij. Vervolgens is de rekentoets vo door de regering ingevoerd om te toetsen of elke kandidaat voldoet aan het referentieniveau 2F.

Om te onderzoeken of de rekentoets 2F de gestelde referentieniveaus op een passende wijze toetst, heeft het ministerie van OCW de commissie Bosker in het leven geroepen. Deze commissie heeft de rekentoets vo tegen het licht gehouden met onder andere het volgende resultaat:

Wij merken op dat de Expertgroep Doorlopende Leerlijnen Rekenen en Taal als oogmerk had dat het 2F-niveau voor een belangrijk deel overeen zou stemmen met de rekenonderdelen van het wiskunde-examen voor vmbo-bb en vmbo-kb. Nu niet alleen de vmbo-bb, maar ook de vmbo-kb leerlingen het relatief minder goed doen op de rekentoetsen dan op dat examen, is er volgens ons alle reden toe nader onderzoek te laten doen naar de oorzaken van deze discrepantie. Op basis daarvan zal bekeken moeten worden wat dit betekent voor de rekentoetsen en/of de referentiecesuur.

[...]

Onderzoek de discrepantie tussen de vmbo wiskunde-examens en de rekentoets (en de verschillen in resultaten) en bezie wat dit betekent voor de rekentoetsen en/of de referentiecesuur.

(Commissie Bosker, 2014, p. 23)

Het bovenstaande advies is de aanleiding geweest voor dit onderzoek. Omdat deze schooltypen nadrukkelijk in het bovenstaande citaat worden genoemd, beperken we ons hierin tot kandidaten van vmbo bb en vmbo kb. Andere redenen voor deze keuze zijn dat rekenen en wiskunde in deze schooltypen dicht bij elkaar staan en dat rekenen in het centraal examen (CE) wiskunde vmbo bb en kb een grote rol speelt. Verder worden de rekentoets 2F en het CE wiskunde vmbo bb en kb hoofdzakelijk digitaal afgenomen, wat niet het geval is bij de andere schooltypen van het voortgezet onderwijs en vergelijking beter mogelijk maakt dan bij toetsing met verschillende media.

Het doel van dit onderzoek is het inventariseren van verschillen tussen de resultaten van kandidaten op de rekentoets 2F en het digitale centraal examen (CE) van wiskunde bij vmbo bb en vmbo kb en het identificeren van mogelijke oorzaken van deze verschillen. Als er vanaf nu over de rekentoets vo wordt geschreven, dan wordt hiermee de rekentoets vo niveau 2F bedoeld; met het CE wiskunde wordt de digitale versie van het CE wiskunde vmbo bb en kb bedoeld.

Dit onderzoek is van belang omdat de rekentoets vo vanaf het schooljaar 2015-2016 een verplicht eindexamenonderdeel zal zijn en deel uitmaakt van de slaag-/zakregeling, waar de toets sinds 2013-2014 slechts verplicht is en het cijfer als bijlage op de cijferlijst wordt vermeld. Een te groot verschil tussen de resultaten op de rekentoets vo en op het CE wiskunde is ongewenst.

Centraal staan de volgende onderzoeksvragen:

1. In welke opzichten verschillen de rekentoets vo en het CE wiskunde vmbo bb en vmbo kb?
2. Zijn er significante verschillen tussen de scores van kandidaten op de rekentoets vo en het CE wiskunde vmbo bb en vmbo kb?
3. Kunnen de bij 2. geconstateerde scoreverschillen verklaard worden uit de bij 1. gevonden verschillen?

2. Methode

2.1. Methode onderzoeksvraag 1

Om de eerste onderzoeksvraag te beantwoorden heeft er een inventarisatie plaatsgevonden. Deze inventarisatie was bedoeld om een overzicht te krijgen van de inhoud, vorm en afnamecondities van de rekentoets vo en het CE wiskunde vmbo bb en kb in de jaren 2013, 2014 en 2015. Er zijn verschillende documenten, bronnen, toetsen en examens bestudeerd. Het gaat om de volgende activiteiten.

- Een analyse en beschrijving van overeenkomsten en verschillen in de toetswijzer van de rekentoets vo en de syllabi van het CE wiskunde vmbo bb en kb en in bijbehorende achtergronddocumenten;
- Een analyse van overeenkomsten en verschillen wat betreft de getoetste rekenvaardigheid en de presentatie van de items van de rekentoets vo en de items van het CE wiskunde vmbo bb en kb. Elementen die hierin betrokken kunnen worden, zijn:
 - De verschillen in het al dan niet kunnen terugbladeren;
 - De verschillen in contextgebruik: bijvoorbeeld elk item een nieuwe context dan wel clusteritems met meerdere items rond één context. Of dit verschil invloed heeft op de resultaten van de kandidaten wordt in een ander onderzoek naar aanleiding van het advies van de commissie Bosker onderzocht;
 - De verschillen in complexiteit aan de hand van het tellen van rekenstappen;
 - De verschillen in het soort rekenmachine en wanneer de rekenmachine toegestaan is. Dit aspect is in het kader van dit onderzoek niet nader onderzocht;
 - De verschillen qua taligheid;
 - De verschillen in de beoordeling (goed/fout versus deelscores).
- Een inventarisatie van contextitems uit de rekentoets vo die inhoudelijk ook in het CE wiskunde vmbo bb en kb zouden kunnen worden opgenomen en, andersom, opgaven van het CE wiskunde die rekendoelen toetsen. Hiervoor zijn eerst criteria opgesteld met hulp van een rekentoetsexpert;
- De ontwikkeling van een instrument om de inhoudelijke complexiteit van items te bepalen. Dit instrument kan worden gebruikt om sets van items te identificeren die bestaan uit één item uit de rekentoets vo en één uit het CE wiskunde vmbo bb en kb, die dezelfde rekenvaardigheid toetsen en vergelijkbare complexiteit hebben. Uitgangspunt voor deze instrumentontwikkeling is het aantal rekenstappen dat nodig is om het eindantwoord van het item te vinden. Om de betrouwbaarheid van het ontwikkelde instrument te waarborgen, is de interbeoordelaarsbetrouwbaarheid bepaald. Hier wordt verder op ingegaan in de discussie.

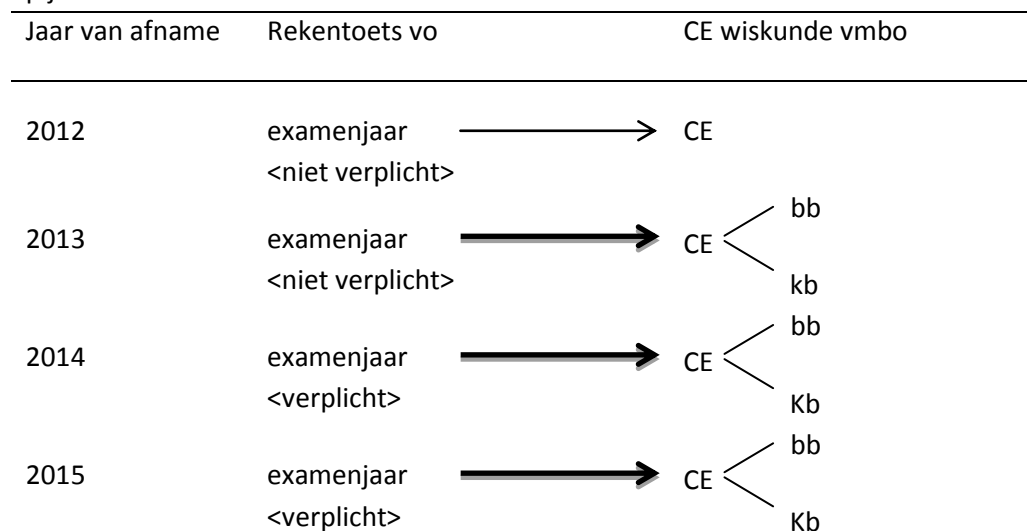
2.2. Methode onderzoeksvraag 2

Om de tweede onderzoeksvraag te beantwoorden zijn scores van kandidaten geanalyseerd voor de afnamejaren 2013, 2014 en 2015. Drie groepen kandidaten zijn in dit onderzoek betrokken. De eerste groep bestaat uit de vmbo bb en kb kandidaten die in 2013 zowel de rekentoets vo als het CE wiskunde vmbo bb of kb hebben gemaakt. Deze eerste groep bestaat in totaal uit 12407 kandidaten, waarvan 5426 vmbo bb kandidaten en 6981 vmbo kb kandidaten. De tweede groep bestaat uit de vmbo bb en kb kandidaten die in 2014 zowel de rekentoets vo als het CE wiskunde vmbo bb of kb hebben gemaakt. Deze tweede groep bestaat in totaal uit 22802 kandidaten, waarvan 9764 vmbo bb

kandidaten en 13038 vmbo kb kandidaten. De derde groep bestaat uit de vmbo bb en kb kandidaten die in 2015 zowel de rekentoets vo als het CE wiskunde vmbo bb of kb hebben gemaakt. Deze derde groep bestaat in totaal uit 11987 kandidaten, waarvan 5534 vmbo bb kandidaten en 6453 vmbo kb kandidaten. Figuur 1 geeft een overzicht van de betrokken groepen.

Figuur 1

Overzicht van de cohorten kandidaten; het onderzoek richt zich op de kandidaten bij de vetgedrukte pijlen



Om de scores van kandidaten te analyseren is er een selectie gemaakt van items die meegenomen worden in de analyse. Deze selectie is gebaseerd op inhoudelijke en toetstechnische criteria waaraan items uit het CE wiskunde vmbo bb en kb moeten voldoen om in de rekentoets vo te kunnen functioneren en andersom (zie bijlage 1). Deze criteria zijn opgesteld bij het beantwoorden van onderzoeksvraag 1. In bijlage 2 zijn enkele items weergegeven die al dan niet voldoen aan deze criteria. Doordat er een selectie is gemaakt van items die meegenomen worden in de analyse zijn niet alle items uit een toets bekeken. Dit betekent bijvoorbeeld voor het CE wiskunde vmbo kb 2014 variant 1h dat er slechts 6 items zijn meegenomen in de analyse. Over deze 6 items is de totale score per kandidaat berekend en de maximaal te behalen score. Vervolgens is de behaalde score van een kandidaat gedeeld door de maximaal te behalen score. Op deze manier zijn de scores van de kandidaten op alle toetsen geschaald naar een schaal van 0 tot 1. Met t-toetsen is onderzocht of deze geschaalde resultaten van de kandidaten op de rekentoets vo en het CE wiskunde vmbo bb en kb significant verschillen. Deze toets is zes maal uitgevoerd, namelijk op de scores van vmbo bb kandidaten in 2013, 2014 en 2015 en op scores van vmbo kb kandidaten in 2013, 2014 en 2015. Doordat de leerwegsectoren vmbo bb en kb afzonderlijk zijn bekeken, kunnen er ook uitspraken worden gedaan over de verschillen tussen deze sectoren.

Tevens is er op kleine schaal gekeken naar de scores van leerlingen op sets van items. Deze sets van items zijn gekoppeld aan de hand van het aantal benodigde rekenstappen. Van elk leerjaar en elke leerwegsector is van acht itemkoppels, met behulp van een t-toets, nagegaan of de scores op de items uit de rekentoets vo en de items uit het CE wiskunde vmbo bb en kb significant verschillen en zo ja, in welke richting.

2.3. Methode onderzoeksvraag 3

Om de derde onderzoeksvraag te beantwoorden is er een antwoord op onderzoeksvraag 2 nodig. Als er geen significante scoreverschillen gevonden zijn, dan komt onderzoeksvraag 3 te vervallen. Omdat er significante scoreverschillen gevonden zijn, zijn verklaringen voor de scoreverschillen gezocht door deze te koppelen aan de bevindingen van de analyse die gedaan is bij onderzoeksvraag 1. Er is nagegaan in hoeverre de gevonden item- en presentatiefactoren, waaronder de afnamecondities, de gevonden verschillen kunnen verklaren.

Eerst is onderzocht in hoeverre de rekentoets vo en het CE wiskunde vmbo bb en kb samenhangen. Hierdoor is bepaald of de toetsen hetzelfde meten. Dit is onderzocht voor de afnamejaren 2013 en 2014. In het bepalen van de samenhang zijn enkel de items uit de rekentoets vo en het CE wiskunde vmbo bb en kb meegenomen die in beide toetsen kunnen worden opgenomen. Om de samenhang te bepalen is er allereerst een vaardigheidsschaal, in dit geval een zogeheten One-Parameter Logistic Model of OPLM-schaal, gemaakt. In bijlage 3 is toegelicht hoe deze OPLM-schaal tot stand is gekomen. Vervolgens is de correlatie tussen de toetsen berekend. Als de rekenitems in het CE wiskunde hetzelfde concept meten als de items in de rekentoets vo dan zal de correlatie groot positief zijn.

2.3.1. Tijdnood

Er is onderzocht of de kandidaten tijdnood ondervonden bij de rekentoets vo of het CE wiskunde vmbo bb of kb. Dit is geen factor die bij onderzoeksvraag 1 ter sprake komt, maar tijdnood kan een effect hebben op de scores van kandidaten. Als indicatie van tijdnood is het percentage kandidaten beschouwd dat voortijdig klaar was met een van de toetsen. Als kandidaten niet de hele beschikbare tijd benutten, kan ervan uitgegaan worden dat ze genoeg tijd hadden. Tevens zijn er gesprekken gevoerd met toetsexperts van Cito die betrokken zijn bij de ontwikkeling van deze toetsen. Zij hebben de afnames van de toetsen meegemaakt en als er sprake was van tijdnood hebben zij dit gehoord of gezien. Tot slot zijn de afnameverslagen meegenomen in de analyse. In een dergelijk verslag worden in het algemeen tijdnood of klachten over de lengte van de toets vermeld.

2.3.2. Terugbladeren

Om het effect op de scores van kandidaten te onderzoeken van het niet kunnen terugbladeren tijdens de rekentoets vo is er gebruik gemaakt van het onderzoek van Straat (2014) naar het effect hiervan op de scores van kandidaten op het Centraal Ontwikkelde Examen (COE) 2F in het mbo. Met behulp van de resultaten uit dit onderzoek is gezocht naar een soortgelijk resultaat voor de rekentoets vo.

2.3.3. Complexiteit en moeilijkheid

Er worden twee soorten complexiteit van items onderscheiden. Als eerste is er de inhoudelijke complexiteit van een item. Deze is uitgedrukt in het aantal benodigde rekenstappen om het eindantwoord te vinden. Dit betekent dat een item waarbij vier rekenstappen nodig zijn inhoudelijk complexer is dan een item waarbij twee rekenstappen nodig zijn. Het aantal benodigde rekenstappen om een item op te lossen is een eerste indicatie van de complexiteit van de items en, door te middelen over de items, van de toets als geheel. Op deze manier is er een algemene eerste indruk gekregen van de complexiteit van de toetsen. Om het aantal stappen per item te kunnen tellen, is er bepaald wat als een rekenstap wordt gezien; criteria voor het stappen tellen zijn opgesteld en besproken met enkele toetsdeskundigen van Cito. De criteria zijn te vinden in bijlage 4

en in bijlage 2 zijn enkele voorbeelden uitgewerkt. Er is gekozen om naar het aantal benodigde rekenstappen te kijken omdat dit de inhoudelijke complexiteit van de items weergeeft en het een eenvoudig middel is om items te koppelen. Ten tweede is de complexiteit van items bepaald door toetsdeskundigen. Een item is complexer als het lastiger is om alle punten voor dat item te behalen.

De moeilijkheid van een item is een empirisch kenmerk op basis van de responsen van kandidaten en is eveneens op twee manieren gedefinieerd. Een eerste maat voor moeilijkheid is de gestandaardiseerde score van het item. De tweede maat voor moeilijkheid is de positie van het item op de OPLM-schaal. Hierbij is een OPLM-schaal ontwikkeld door items van het CE wiskunde vmbo bb en kb te vergelijken met de items van de rekentoets vo op psychometrische kenmerken, volgende de redenering: "Als de items uit het CE wiskunde in de rekentoets vo zouden zijn opgenomen, hoe zouden deze items dan functioneren?" Zodoende kan er een uitspraak gedaan worden of de items uit het CE wiskunde vmbo bb en kb moeilijker zijn dan de items uit de rekentoets vo en welke prestatie op het CE wiskunde bb en kb geleverd zou moeten worden om aan het referentieniveau 2F rekenen te voldoen. Er is aangenomen dat voor de CE wiskunde items net als bij de rekentoets vo geen deelscores worden toegekend, maar alleen goed-fout scores.

2.3.4. Motivatie

De motivatie van de kandidaat voor de verschillende toetsen kan een rol spelen bij de resultaten. Het is namelijk zo dat de rekentoets vo en het CE wiskunde vmbo bb en kb verschillend meetelden in 2013, 2014 en 2015. De toetsen waren in 2013, 2014 en 2015 wel verplicht maar een onvoldoende op de rekentoets had geen consequenties. Het cijfer van het CE wiskunde vmbo bb of kb telde echter wel mee in de slaag-/zakregeling. In het onderwijsverslag 2012-2013 van de Inspectie van het Onderwijs (2014) komt naar voren dat kandidaten aangeven dat hun motivatie daalt als een toets niet meetelt. Motivatie is mogelijk dus een factor die een rol speelt.

De eerste manier waarop de motivatiefactor onderzocht is, is door de complexiteit te vergelijken van itemparen bestaande uit één item van de rekentoets vo en één item van het CE wiskunde, die vergelijkbaar zijn qua moeilijkheid op de OPLM-schaal. De redenering is dat bij gelijke motivatie van kandidaten voor beide toetsen de complexiteit van items met gelijke moeilijkheid ook ongeveer gelijk zal zijn. Als daarentegen de motivatie van kandidaten voor bijvoorbeeld de rekentoets vo lager is dan voor het CE wiskunde, dan valt te verwachten dat binnen itemparen met gelijke moeilijkheid de complexiteit van het CE-item groter is. Om dit te onderzoeken, hebben vijf toetsexperts van Cito voor elk leerjaar en elke leerwegsector van 10 itemkoppels met gelijke moeilijkheid bepaald welk item van het koppel zij als complexer beschouwen. In bijlage 5 is de opdracht te vinden.

De tweede manier om de motivatiefactor te onderzoeken werkt andersom. Hierbij is er voor elke op basis van rekenstappen gekoppelde set van items gekeken naar hun schaling op de OPLM-schaal. Voor vmbo bb en kb uit 2013 en vmbo bb uit 2014 is er naar het verschil in OPLM-schaling van 54 itemkoppels gekeken en voor vmbo kb 2014 naar 40 itemkoppels. Het kleinere aantal itemkoppels voor vmbo kb 2014 komt doordat er minder items uit het CE wiskunde ook in de rekentoets vo opgenomen kunnen worden en hierdoor zijn er minder itemkoppels te maken. Vervolgens is met een binomiaaltoets onderzocht of de verschillen van de moeilijkheidsgraad volgens deze schaling significant van 0 verschillen. Er is immers door middel van het tellen van de benodigde rekenstappen gesteld dat de gekoppelde items inhoudelijk even complex zijn, bovendien wordt er van een gelijke

vaardigheid van de kandidaten uitgegaan. Als de verschillen niet significant zijn, wijst dit niet op een verschil in motivatie van de kandidaten. Wat er wel geconcludeerd kan worden is dat het tellen van het aantal benodigde rekenstappen de moeilijkheid, zoals blijkt uit prestaties van kandidaten, goed weergeeft. Als de verschillen wel significant zijn, zijn er twee mogelijke verklaringen. De eerste is dat het tellen van het aantal benodigde rekenstappen de complexiteit, zoals blijkt uit prestaties van kandidaten, niet goed weergeeft. De andere mogelijke verklaring is dat de kandidaten op een van de toetsen minder gemotiveerd zijn en deze daarom slechter maken.

2.3.5. Taligheid

Volgens Evers-Vermeul (2015) leggen lange stukken tekst een extra druk op het werkgeheugen waardoor de prestaties van kandidaten omlaag kunnen gaan. De taligheid van de items uit een toets kan dus een rol spelen bij de scores van een kandidaat op die toets en is dus een factor die scoreverschillen mogelijk kan verklaren.

De taligheid van de rekentoets vo en het CE wiskunde is kleinschalig onderzocht. Er zijn vijf itemkoppels gemaakt aan de hand van het stappen tellen. Deze vijf koppels zijn door Jacqueline Evers-Vermeul geanalyseerd. In deze analyse is steeds bepaald welk item van het koppel qua taal complexer is. Dit geeft een eerste inschatting van de taligheid van de toetsen. De geanalyseerde koppels bestaan uit één item van de rekentoets vo tijdvak 1 2013 en van het CE wiskunde vmbo bb 2013 en hierdoor kunnen enkel uitspraken gedaan worden over de taligheid van deze twee toetsen vergeleken met elkaar. Voor het CE wiskunde vmbo kb 2013 of voor heel 2014 en 2015 zijn nog geen uitspraken mogelijk.

2.3.6. Cesuur

Voor zowel de rekentoets vo als het CE wiskunde wordt er een cesuur, dat wil zeggen de grens tussen een voldoende en een onvoldoende, gehanteerd. De cesuur wordt op basis van statistische analyses overgebracht van een vorige afname op de huidige afname. Bij het CE wiskunde en de rekentoets vo gebeurt het overbrengen van de cesuur door middel van verschillende procedures. Een uitgebreide beschrijving van deze procedures is te vinden op www.toetswijzer.nl. Op basis van deze procedures kan worden bepaald hoeveel kandidaten een voldoende hebben gehaald voor het examen of de toets. In de resultaten is kort besproken hoe de gehanteerde cesuren worden vastgelegd en is onderzocht hoe het verschil in moeilijkheid tussen de rekentoets vo en het CE wiskunde vmbo bb en kb zich verhoudt tot de percentages voldoende voor de afnamejaren 2013 en 2014.

2.3.7 Afnamejaar 2015

Gezien de beperkte tijd tussen de afname van de toetsen en de deadline van dit onderzoek was het niet mogelijk om alle hierboven genoemde factoren voor het afnamejaar 2015 te onderzoeken. Dit kwam met name doordat het in de korte tijd niet mogelijk was om een OPLM-schaal voor 2015 te maken en om toetsexperts in te schakelen. Doordat er geen OPLM-schaal voorhanden was, was het niet mogelijk om de samenhang tussen de toetsen te bepalen en de moeilijkheid met behulp van de OPLM-schaal te bepalen. Enkel de inhoudelijke complexiteit van de toetsen is bepaald, de complexiteit aan de hand van het oordeel van de toetsexperts was niet mogelijk. Verder was het niet mogelijk om de motivatiefactor te onderzoeken omdat voor beide manieren óf de OPLM-schaal óf de hulp van toetsexperts nodig is. Tot slot is de bespreking van de cesuur in 2015 achterwege gelaten.

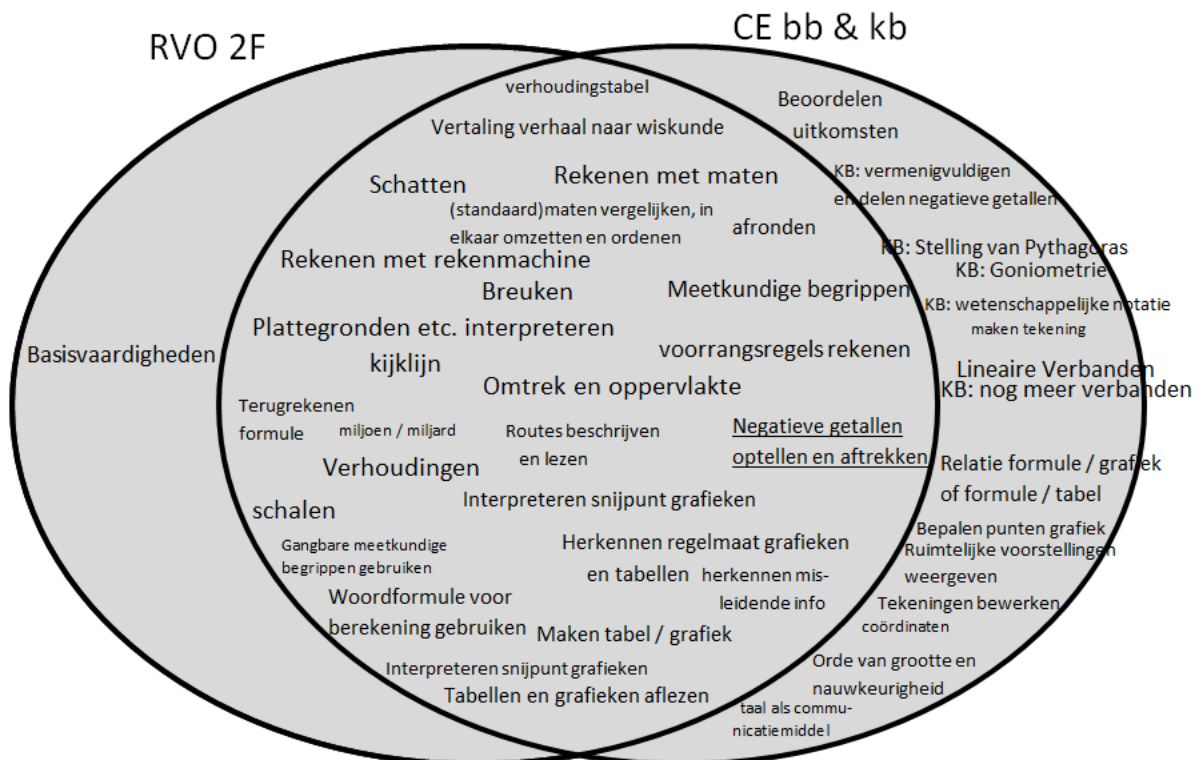
3. Resultaten

3.1.1. Vergelijking rekentoetswijzer 2F en syllabi wiskunde examen vmbo bb en kb

Hier worden de overeenkomsten en verschillen tussen de rekentoetswijzer 2F van 2013 en 2014 (Dekker & Schmidt, 2011), de rekentoetswijzer 2F van 2015 (College voor Examens, 2014) en de syllabi wiskunde examen vmbo bb en kb uit 2013 (College voor Examens, 2011), 2014 (College voor Examens, 2012) en 2015 (College voor Examens, 2013) in kaart gebracht. Voor 2013 en 2014 was de rekentoetswijzer 2F exact hetzelfde en er hoeft dus geen rekening gehouden te worden met onderlinge verschillen. Voor 2015 is er een nieuwe rekentoetswijzer gemaakt. De verschillen met de rekentoetswijzer van 2013 en 2014 zijn dat er naar verhouding meer contextloze opgaven in de rekentoets vo komen en dat er geen negatieve getallen meer opgeteld en afgetrokken moeten worden. De syllabi wiskunde examen vmbo bb en kb uit 2013, 2014 en 2015 zijn identiek voor deze leerwegsectoren. In Figuur 2 wordt in een diagram weergegeven wat de overeenkomsten en verschillen tussen de rekentoetswijzer 2F en de syllabi wiskunde examen vmbo bb en kb zijn. Hierbij is het onderstreepte gedeelte geen onderdeel meer van de rekentoetswijzer 2F in 2015. In bijlage 6 zijn de gevonden overeenkomsten en verschillen in meer detail beschreven.

Figuur 2

Overeenkomsten en verschillen tussen de rekentoetswijzer en de syllabi CE



In Figuur 2 is duidelijk te zien dat alle vaardigheden die bij de rekentoets VO getoetst worden ook getoetst worden in het CE wiskunde vmbo bb en kb behalve de basisvaardigheden. Er zijn echter vaardigheden die op het CE wiskunde vmbo bb of kb getoetst worden die niet getoetst worden op de rekentoets VO.

3.1.2. Vorm en afnamecondities

Om de rekentoets vo en het CE wiskunde vmbo bb en kb goed te kunnen vergelijken moet er ook gekeken worden naar de vorm en de afnamecondities van de toetsen. Deze kunnen namelijk de resultaten van de kandidaten op de toetsen beïnvloeden. De manier waarop de vorm en afnamecondities zijn onderzocht, is vergelijkbaar met die van Hoogland, Vliegthart, Schoonen, Van der Mark en Van Reeuwijk (2014) voor de rekentoets vo 2012 en CE wiskunde vmbo bb 2012. De feitelijke gegevens van de vergeleken toetsen zijn te vinden in bijlage 7. Tabel 1 geeft de verschillen tussen de rekentoets vo 2013 en het CE wiskunde vmbo bb en kb 2013 weer, Tabel 2 en 3 analoog voor respectievelijk 2014 en 2015.

Tabel 1

Vorm- en afnameverschillen rekentoets vo 2013 en CE wiskunde vmbo bb en kb 2013

Vershil	Rekentoets vo 2013	CE wiskunde vmbo bb 2013	CE wiskunde vmbo kb 2013
Toetsduur	90 minuten	90 minuten	120 minuten
Aantal items	60 items	22, 23 of 24 items	23 of 24 items
Contextloze items	Aanwezig	Niet aanwezig	Niet aanwezig
Context	Nieuwe context per item	Meerdere items over één context	Meerdere items over één context
Soort rekenmachine	Computer rekenmachine	Eigen rekenmachine	Computer rekenmachine met toolbox
Beschikbaarheid rekenmachine	Enkel bij contextitems	Altijd beschikbaar	Altijd beschikbaar
Terugbladeren	Niet toegestaan	Toegestaan	Toegestaan

Tabel 2

Vorm- en afnameverschillen rekentoets vo 2014 niveau 2F en CE wiskunde vmbo bb en kb 2014

Vershil	Rekentoets vo 2014	CE wiskunde vmbo bb 2014	CE wiskunde vmbo kb 2014
Toetsduur	90 minuten	90 minuten	120 minuten
Aantal items	51 items	22 tot 26 items	20 tot 22 items
Contextloze items	Aanwezig	Niet aanwezig	Niet aanwezig
Context	Nieuwe context per item	Meerdere items over één context	Meerdere items over één context
Soort rekenmachine	Computer rekenmachine	Eigen rekenmachine	Computer rekenmachine met toolbox
Beschikbaarheid rekenmachine	Enkel bij contextitems	Altijd beschikbaar	Altijd beschikbaar
Terugbladeren	Niet toegestaan	Toegestaan	Toegestaan

Tabel 3

Vorm- en afnameverschillen rekentoets vo 2014 niveau 2F en CE wiskunde vmbo bb en kb 2015

Verskil	Rekentoets vo 2015	CE wiskunde vmbo bb 2015	CE wiskunde vmbo kb 2015
Toetsduur	90 minuten	90 minuten	120 minuten
Aantal items	45 items	21 tot 25 items	21 tot 23 items
Contextloze items	Aanwezig	Niet aanwezig	Niet aanwezig
Context	Nieuwe context per item	Meerdere items over één context	Meerdere items over één context
Soort rekenmachine	Computer rekenmachine	Eigen rekenmachine	Computer rekenmachine met toolbox
Beschikbaarheid rekenmachine	Enkel bij contextitems	Altijd beschikbaar	Altijd beschikbaar
Terugbladeren	Toegestaan binnen de twee blokken	Toegestaan	Toegestaan

3.2. Resultaten onderzoeksvraag 2

De tweede onderzoeksvraag luidt: Zijn er significante verschillen tussen de scores van kandidaten op de rekentoets vo en het CE wiskunde vmbo bb en vmbo kb?

Uit de resultaten van de zes uitgevoerde t-toetsen blijkt dat de scores van de rekentoets vo significant lager zijn dan die van het CE wiskunde voor zowel vmbo bb als kb en voor de jaren 2013, 2014 en 2015 met een redelijke tot hoge effectgrootte met uitzondering van vmbo kb 2014. In Tabel 4 staan de gevonden resultaten.

Tabel 4

Uitkomst t-toetsen op scores CE wiskunde vergeleken met scores rekentoets vo

Leerwegsector en afnamejaar	M(RVO) (SD(RVO))	M(CE) (SD(CE))	t-waarde	Significantie	Cohen's d
Vmbo bb 2013	.294 (.147)	.533 (.171)	t(5362) = 111.263	.000	1.50
Vmbo kb 2013	.445 (.169)	.576 (.195)	t(6941) = 63.319	.000	0.72
Vmbo bb 2014	.354 (.002)	.504 (.178)	t(9566) = 91.334	.000	1.19
Vmbo kb 2014	.467 (.158)	.488 (.210)	t(12778) = 12.254	.000	0.11
Vmbo bb 2015	.419 (.163)	.567 (.173)	t(5532) = 66.633	.000	0,88
Vmbo kb 2015	.575 (.158)	.672 (.178)	t(6452) = 44.825	.000	0.58

In Tabel 4 staat M voor de gemiddelde score per item op de toets en SD staat voor de standaarddeviatie, M(RVO) is dus de gemiddelde score per item op de rekentoets vo.

Uit Tabel 4 komt naar voren dat de vmbo kb kandidaten in alle drie de jaren hoger scoren op de rekentoets vo dan de vmbo bb kandidaten. Een trend die tevens te zien is, is dat de rekentoets vo elk jaar beter wordt gemaakt dan het jaar ervoor, dit geldt voor zowel de vmbo bb als de vmbo kb kandidaten. De betere prestaties op de rekentoets zijn toe te schrijven aan een combinatie van twee factoren. Enerzijds is de toets eenvoudiger geworden wat zich uit in een hoger percentage correcte antwoorden dat nodig is om een voldoende te halen. In 2013 kon een vmbo-kb kandidaat bij 60% van de opgaven correct een voldoende halen, terwijl in 2015 66,7% van de opgaven correct beantwoord moeten worden door een vmbo-kb kandidaat om een voldoende te halen. Daarnaast geldt dat bij het

eerste tijdvak een steeds groter deel van de kandidaten een voldoende heeft behaald. De prestaties zijn dus ook in absolute zin verbeterd. Met name in 2015 is er een grote stijging in de prestaties ten opzichte van voorgaande jaren waargenomen. Bij de resultaten van de vergelijking van de prestaties op de rekentoets vo en het CE wiskunde vmbo kb moet echter een kanttekening geplaatst worden. Het aantal items uit het CE wiskunde van vmbo kb dat ook in de rekentoets kan voorkomen, is in sommige varianten klein (soms maar 6). Er is hier dus maar een klein aantal items bekeken.

Het globale beeld dat het CE wiskunde vmbo bb en kb beter wordt gemaakt dan de rekentoets vo, wordt geïllustreerd door de vergelijking van specifieke itemkoppels. In Tabellen 15, 16, 17, 18, 19 en 20 in bijlage 8 is te zien dat de items van het CE wiskunde in de meeste gevallen beter worden gemaakt dan de bijbehorende items van de rekentoets vo, met uitzondering van vmbo kb in 2015. Dit onderbouwt het eerder gevonden resultaat dat de scores van de rekentoets vo significant lager zijn dan die van het CE wiskunde voor zowel vmbo bb als kb. De verschillen blijken echter niet altijd groot zijn en in sommige itemparen wordt het item van de rekentoets vo veel beter gemaakt dan het bijbehorende item van het CE wiskunde.

3.3. Resultaten onderzoeksvraag 3

Deze onderzoeksvraag luidt: Kunnen de bij 2. geconstateerde scoreverschillen verklaard worden uit de bij 1. gevonden verschillen?

Allereerst is er gekeken naar de samenhang tussen de scores op het CE wiskunde vmbo bb en kb en de scores op de rekentoets vo in 2013 en 2014. De Pearson correlatiecoëfficiënt voor het CE wiskunde bb 2013 en 2014 en de rekentoets vo is respectievelijk 0,75 en 0,68. Voor het CE wiskunde kb 2013 en 2014 zijn deze waarden 0,65 en 0,53. We concluderen dus dat de scores voor het CE wiskunde en de rekentoets vo een middelmatige samenhang vertonen, die voor bb sterker is dan voor kb en in 2013 sterker dan in 2014.

3.3.1. Tijdnood

Over de tijdnood van kandidaten op de rekentoets vo kan enkel iets gezegd worden over alle kandidaten van alle leerwegsectoren van het vmbo. Dit houdt in dat alle uitspraken gaan over de kandidaten van vmbo tl, gl, kb en bb samen. In 2013 was er mogelijk sprake van enige tijdnood omdat de rekentoets toen nog uit 60 items bestond. In 2014 en 2015 is er echter geen indicatie van tijdnood omdat de meeste kandidaten niet de hele toetsduur hebben benut.

Na een gesprek met toetsdeskundigen bij Cito kan er ook geconstateerd worden dat er bij het CE wiskunde van 2013, 2014 en 2015 bij zowel vmbo bb als vmbo kb geen sprake is geweest van tijdnood. Ook zijn er geen reacties van docenten geweest over de beschikbare tijd. Tevens is er inzage geweest in de verslagen van de afnames. Hierin wordt niet gemeld dat er tijdnood is geobserveerd of is gemeld.

Hieruit kan geconcludeerd worden dat de kandidaten enkel bij de rekentoets vo 2013 wellicht tijdnood hebben ervaren. Voor de rekentoets vo 2014 en 2015 en het CE wiskunde vmbo bb en kb uit 2013, 2014 en 2015 was er geen sprake van tijdnood. Tijdnood kan dus enkel een rol spelen bij het verklaren van de scoreverschillen tussen de rekentoets vo en het CE wiskunde vmbo bb en kb uit 2013. In 2014 en 2015 was tijdnood geen factor van betekenis.

3.3.2. Terugbladeren

Bij de rekentoets vo was het in 2013 en 2104 niet mogelijk om terug te bladeren naar eerdere items en bij het CE wiskunde vmbo bb en kb wel. Straat (2014) heeft onderzoek gedaan naar het effect van het al dan niet kunnen terugbladeren op de prestaties van kandidaten. In dat onderzoek heeft hij gekeken naar het Centraal Ontwikkelde Examen (COE) 2F rekentoets die wordt afgenomen in het mbo. Straat concludeert hierover het volgende:

De conclusie is dat het niet kunnen terugbladeren op zichzelf geen negatief of positief effect heeft op de leerlingprestaties. Andere omstandigheden zouden er echter wel voor kunnen zorgen dat terugbladeren een effect heeft. Bijvoorbeeld, bij COE 2F rekenen is er geen indicatie dat kandidaten tijdnood ervaren. Mogelijk dat niet kunnen terugbladeren in combinatie met tijdnood wel een effect zou hebben op het functioneren van de items en de leerlingprestaties. Het blijft daarom de vraag of dit resultaat kan worden gegeneraliseerd naar andere toetsen zoals 3F Rekenen en de Rekentoets vo.

(Straat, 2014, p.2)

Straat (2014) geeft aan dat zijn conclusie niet zonder meer naar andere toetsen gegeneraliseerd kan worden. De belangrijkste reden hiervoor is dat het ervaren van tijdsdruk bij het maken van de toets een sterk negatief effect kan hebben wanneer de kandidaat niet terug kan bladeren. Voor de rekentoetsen 2F geldt dat er voor zowel het mbo als het vo sprake is van weinig tijdsdruk. Daarom is de verwachting dat de resultaten van het onderzoek in mbo gegeneraliseerd kunnen worden naar het vo. Hieruit kan dus geconcludeerd worden dat het niet kunnen terugbladeren geen effect heeft op de prestaties van kandidaten op de rekentoets vo. Bovendien was het vanaf 2015 wel mogelijk om terug te bladeren op de rekentoets vo en speelt deze factor vanaf dat jaar dus geen rol meer.

3.3.3. Complexiteit en moeilijkheid

Als eerste is er gekeken naar de inhoudelijke complexiteit van de items van de rekentoets vo en het CE wiskunde. Dit is gedaan met behulp van het tellen van het aantal benodigde rekenstappen per item. Voor het CE wiskunde vmbo bb en kb en de rekentoets vo van zowel 2013 als 2014 als 2015 zijn de stappen van alle items geteld. De uitkomsten hiervan staan in Tabel 5 voor 2013 en in Tabel 6 en Tabel 7 voor 2014 en 2015.

Tabel 5

Aantal items per aantal rekenstappen CE wiskunde vmbo bb en kb en rekentoets vo 2013

Aantal items per rekenstap	CE wiskunde vmbo bb 2013	CE wiskunde vmbo kb 2013	Rekentoets vo 2013
Aantal items met 1 rekenstap	15 (21.1%)	2 (6.7%)	12 (5.4%)
Aantal items met 2 rekenstappen	24 (33.8%)	6 (20.0%)	47 (21.2%)
Aantal items met 3 rekenstappen	15 (21.1%)	14 (46.7%)	81 (36.5%)
Aantal items met 4 rekenstappen	7 (9.9%)	7 (23.3%)	49 (22.1%)
Aantal items met 5 rekenstappen	9 (12.7%)	1 (3.3%)	26 (11.7%)
Aantal items met 6 rekenstappen	1 (1.4%)		7 (3.2%)
Totaal aantal items	71	30	222

Tabel 6

Aantal items per aantal rekenstappen CE wiskunde vmbo bb en kb en rekentoets vo 2014

Aantal items per rekenstap	CE wiskunde vmbo bb 2014	CE wiskunde vmbo kb 2014	Rekentoets vo 2014
Aantal items met 1 rekenstap	15 (21.1%)	1 (4.8%)	5 (3.3%)
Aantal items met 2 rekenstappen	25 (35.0%)		32 (21.1%)
Aantal items met 3 rekenstappen	21 (29.6%)	8 (38.1%)	52 (34.2%)
Aantal items met 4 rekenstappen	8 (11.3%)	8 (38.1%)	37 (24.3%)
Aantal items met 5 rekenstappen	1 (1.4%)	4 (19.0%)	20 (13.2%)
Aantal items met 6 rekenstappen			4 (2.6%)
Aantal items met 7 rekenstappen	1 (1.4%)		
Aantal items met 8 rekenstappen			1 (0.7%)
Aantal items met 9 rekenstappen			1 (0.7%)
Totaal aantal items	71	21	152

Tabel 7

Aantal items per aantal rekenstappen CE wiskunde vmbo bb en kb en rekentoets VO 2015

Aantal items per rekenstap	CE wiskunde vmbo bb 2015	CE wiskunde vmbo kb 2015	Rekentoets VO 2015
Aantal items met 1 rekenstap	13 (24,1%)	3 (15,8%)	11 (10,5%)
Aantal items met 2 rekenstappen	16 (29,6%)	4 (21,1%)	27 (25,7%)
Aantal items met 3 rekenstappen	18 (33,3%)	7 (36,8%)	36 (34,3%)
Aantal items met 4 rekenstappen	6 (11,1%)	3 (15,8%)	21 (20%)
Aantal items met 5 rekenstappen	1 (1,9%)	1 (5,3%)	7 (6,7%)
Aantal items met 6 rekenstappen			3 (2,9%)
Aantal items met 8 rekenstappen		1 (5,3%)	
Totaal aantal items	54	19	105

Met t-toetsen is getoetst of er een statistisch significant verschil is tussen het gemiddeld aantal benodigde rekenstappen van het betreffende CE wiskunde en de rekentoets vo in dat jaar. Uit de resultaten (zie Tabel 8) blijkt dat het aantal benodigde rekenstappen van de contextitems van de rekentoets vo in 2013, 2014 en 2015 significant hoger ligt dan het aantal benodigde rekenstappen van de rekenitems van het CE wiskunde vmbo bb en kb met als uitzondering vmbo kb in 2014 en 2015. Dit wijst erop dat de rekentoets vo inhoudelijk complexer was dan het CE wiskunde vmbo bb in deze jaren en het CE wiskunde vmbo kb in 2013. In 2014 en 2015 is er geen significant verschil tussen het aantal benodigde rekenstappen van de contextitems van de rekentoets vo en de rekenitems van het CE wiskunde vmbo kb. Dit wijst erop dat de toetsen in deze jaren even complex waren.

Tabel 8

Resultaten t-toetsen op aantal benodigde rekenstappen CE wiskunde vergeleken met rekentoets vo

Leerwegsector en afnamejaar	M(RVO) (SD(RVO))	M(CE) (SD(CE))	t-waarde	significantie
Vmbo bb 2013	3.23 (1.164)	2.63 (1.344)	t(291) = -3.613	.000
Vmbo kb 2013	3.23 (1.164)	2.97 (.928)	t(250) = 1.187	.038
Vmbo bb 2014	3.39 (1.277)	2.41 (1.129)	t(221) = -5.569	.000
Vmbo kb 2014	3.39 (1.277)	3.67 (.966)	t(171) = .938	.349
Vmbo bb 2015	2.95 (1.188)	2.37 (1.033)	t(157) = -3.054	.003
Vmbo kb 2015	2.95 (1.188)	3.00 (1.633)	t(122) = .151	.880

Bij het onderzoeken van onderzoeksvraag 2 is er gevonden dat de rekentoets vo slechter is gemaakt dan zowel het CE wiskunde vmbo bb als het CE wiskunde vmbo kb. Dit suggereert dat de rekentoets vo zowel in 2013 als in 2014 als in 2015 moeilijker was dan het CE wiskunde vmbo bb en kb uit de respectievelijke jaren. De kanttkening hierbij is dat het verschil tussen het gemiddelde van de vmbo kb kandidaten op de rekentoets vo 2014 en hun CE wiskunde minimaal is.

Een belangrijk verschil tussen de rekentoets vo en het CE wiskunde vmbo bb en kb is dat bij de rekentoets vo alleen scorepunten worden toegekend aan het correcte antwoord, terwijl het CE wiskunde vmbo bb en kb toestaat om scorepunten toe te kennen aan correct uitgevoerde tussenstappen. Bij de OPLM-analyses is bij het CE wiskunde vmbo bb en kb dezelfde scoringsregel toegepast als bij de rekentoetsen vo om te onderzoeken of dit het gevonden verschil in moeilijkheid zou kunnen verklaren. Via deze analyse kan de vraag beantwoord worden hoe de resultaten eruit zouden zien als het CE wiskunde vmbo bb en kb en de rekentoets vo dezelfde scoringsregels zouden hebben. Tabel 9 geeft een samenvatting weer van de resultaten van de OPLM-analyses. Elke cel geeft het verschil in moeilijkheid tussen het CE wiskunde en de rekentoets op de gestandaardiseerde vaardigheidsschaal. Een positief getal betekent dat het CE wiskunde moeilijker was dan de rekentoets vo en een negatief getal betekent dat de rekentoets vo moeilijker was. Deze getallen zijn uitgedrukt in standaarddeviaties. Om de resultaten in perspectief te kunnen plaatsen: Het verschil van 0,28 standaarddeviaties in moeilijkheid tussen het CE wiskunde bb 2013 rekenitems en de rekentoets vo 2013 is gelijk aan iets meer dan twee scorepunten op de rekentoets.

Een aantal effecten is opvallend. Ten eerste blijkt dat de rekenitems in 2013 en 2014 gemiddeld makkelijker zijn dan het hele examen. Ten tweede blijkt dat de rekentoets vo in 2014 relatief gemakkelijker is geworden ten opzichte van het CE wiskunde bb en kb. Met name het CE wiskunde kb was in 2014 veel moeilijker dan de rekentoets vo.

Tabel 9
Overzicht van de resultaten van de OPLM-analyses

Toets	Soort items	2013	2014
CE Wiskunde bb	Alle items	-0.25	0.16
	Rekenitems	-0.28	-0.16
CE Wiskunde kb	Alle items	0.24	1.60
	Rekenitems	-0.06	1.48

Doordat bij de OPLM-analyses is uitgegaan van de scoringsregel zoals gebruikelijk bij de rekentoets is de moeilijkheid van beide toetsen goed met elkaar te vergelijken. Bij partial credits krijgen kandidaten gemiddeld meer scorepunten per item omdat het correct beantwoorden van tussenstappen wordt gehonoreerd. Wanneer alleen het correcte eindantwoord een scorepunt oplevert, wordt het dus moeilijker om scorepunten te verdienen. Bij de analyse is gecontroleerd voor het verschil in scoretoekenning en hieruit volgt dat de rekentoets vo qua moeilijkheid tussen het CE wiskunde vmbo bb en het CE wiskunde vmbo kb zit, maar dichterbij het CE wiskunde vmbo bb. Dit is onderzocht voor de afnamejaren 2013 en 2014.

Bovenstaande resultaten tonen aan dat de contextitems in de rekentoets vo inhoudelijk complexer zijn dan de rekenitems in het CE wiskunde vmbo bb in 2013, 2014 en 2015 en vmbo kb in 2013. Het CE wiskunde vmbo kb in 2014 en 2015 is inhoudelijk vrijwel even complex als de rekentoets vo van het bijbehorende jaar. De grotere inhoudelijke complexiteit van de rekentoets vo in vergelijking met het CE wiskunde vmbo bb in 2013, 2014 en 2015 en vmbo kb in 2013 kan een verklaring zijn voor de bij onderzoeksvraag 2 gevonden scoreverschillen. Als dezelfde methode van goed-fout scoring van het eindantwoord die bij de rekentoets vo wordt gebruikt ook bij de wiskunde-examens zou worden toegepast, dan worden de verschillen in gemiddelde score tussen de rekentoets vo en het CE wiskunde vmbo bb en kb kleiner. Dit is onderzocht voor de afnamejaren 2013 en 2014. Het CE wiskunde vmbo kb 2014 blijkt dan zelfs moeilijker dan de rekentoets vo en de rekentoets vo zit qua moeilijkheid tussen het CE wiskunde vmbo bb en het CE wiskunde kb in.

3.3.4. Motivatie

De eerste manier waarop de motivatiefactor onderzocht is, is door de complexiteit van itemparen bestaande uit één item van de rekentoets vo en één item van het CE wiskunde die op de OPLM-schaal gelijk scoren te vergelijken. In Tabel 10 staan de resultaten voor 2013 en 2014 weergegeven. Op de gevonden resultaten zijn binomiaaltoetsen uitgevoerd om na te gaan of de nulhypothese dat de items van de rekentoets vo en het CE wiskunde gemiddeld even complex zijn houdbaar is.

Tabel 10

Beoordeling complexiteit itemkoppels en uitkomst binomiaaltoets

Leerjaar en leerwegsector	CE wiskunde complexer	RVO complexer	Significantie
Vmbo bb 2013	27	23	.672
Vmbo kb 2013	34	16	.015
Vmbo bb 2014	30	20	.203
Vmbo kb 2014	41	9	.000

In Tabel 10 is te zien dat in zowel 2013 als 2014 de nulhypothese voor vmbo bb niet kan worden verworpen. Voor deze gekoppelde items kan dus niet gezegd worden dat de items uit de ene toets complexer zijn dan de items uit de andere toets. Dit is geen aanwijzing dat een lage motivatie van de vmbo bb kandidaten een negatief effect heeft gehad op de score. Voor vmbo kb in zowel 2013 als 2014 wordt de nulhypothese wel verworpen. Voor deze gekoppelde items kan gezegd worden dat de items uit het CE wiskunde complexer zijn dan de items uit de rekentoets vo. Dit zou veroorzaakt kunnen zijn door een lage motivatie voor de rekentoets vo bij de vmbo kb kandidaten. Deze lage motivatie kan een negatief effect gehad hebben op de score op deze toets.

De tweede manier waarop de motivatiefactor onderzocht is, behelst het vergelijken van de scores op de OPLM-schaal van itemparen bestaande uit één item van de rekentoets vo en één item van het CE wiskunde die inhoudelijk even complex zijn. Hierbij is de OPLM-schaling van het item uit de rekentoets vo verminderd met de OPLM-schaling van het bijbehorende item uit het CE wiskunde. De resultaten hiervan zijn te vinden in Tabel 11. Een – geeft hierbij aan dat het verschil negatief is, dus het CE-item is moeilijker dan het item uit de rekentoets, en een + betekent een positieve verschilwaarde.

Tabel 11

Uitkomsten OPLM-schaling rekentoets vo – OPLM-schaling CE wiskunde

Leerwegsector en afnamejaar	Aantal minnen	Aantal plussen
Vmbo bb 2013	28	26
Vmbo kb 2013	28	26
Vmbo bb 2014	17	37
Vmbo kb 2014	14	26

Uit Tabel 11 volgt dat de items van de rekentoets vo en het CE wiskunde vmbo bb en kb die inhoudelijk even complex zijn in 2013 even goed worden gemaakt. In 2014 geldt er daarentegen dat de items van het CE wiskunde vmbo bb en kb beter worden gemaakt dan vergelijkbare items van de rekentoets vo. Dit wijst erop dat de vmbo bb en kb kandidaten in 2014 mogelijk een lage motivatie hadden voor de rekentoets vo.

Naar aanleiding van dit resultaat is er ook nog gekeken naar de gemiddelde OPLM-schaling van items per toets en per aantal benodigde rekenstappen. De resultaten hiervan staan in Tabellen 12 en 13. Hierbij valt op dat zowel in 2013 als 2014 items met veel rekenstappen in de rekentoets vo beter worden gemaakt dan items met evenveel rekenstappen in het CE wiskunde vmbo bb en kb. Dit suggereert dat de afzonderlijke rekenstappen in de rekentoets vo eenvoudiger zijn dan in het CE wiskunde.

Tabel 12

Gemiddelde OPLM-schaling per aantal benodigde rekenstappen per toets van 2013

Aantal rekenstappen	Vmbo bb		Vmbo kb	
	CE wiskunde	Rekentoets vo	CE wiskunde	Rekentoets vo
1	-.811	-.542	-.252	-.546
2	-.138	-.044	-.243	-.106
3	.190	.005	-.075	-.013
4	.529	.267	.250	.303
5	-	.361	.664	.400
6	-.360	.375	-	.321

Tabel 13

Gemiddelde OPLM-schaling per aantal benodigde rekenstappen per toets van 2014

Aantal rekenstappen	Vmbo bb		Vmbo kb	
	CE wiskunde	Rekentoets vo	CE wiskunde	Rekentoets vo
1	-.917	-.351	-.313	-.428
2	-.016	-.128	-	.070
3	.114	.227	.093	.176
4	.629	.524	.668	.394
5	.967	.595	.928	.480
6	-	.940	-	.733
7	1.197	-	-	-
8	-	.695	-	.674
9	-	-.207	-	.103

Met behulp van experts is aangetoond dat van op basis van OPLM-schaling gekoppelde items het item van het CE wiskunde vmbo kb complexer is dan het bijbehorende item van de rekentoets vo in zowel 2013 als 2014. Dit wijst erop dat vmbo kb kandidaten mogelijk een lage motivatie hadden voor de rekentoets vo. Tevens is aangetoond dat op basis van het aantal benodigde rekenstappen gekoppelde items het item van het CE wiskunde vmbo bb en kb in 2014 beter wordt gemaakt dan het bijbehorende item van de rekentoets vo. Dit wijst erop dat de vmbo bb en kb kandidaten in 2014 mogelijk een lage motivatie hadden voor de rekentoets vo. De gevonden resultaten suggereren dus dat prestatieverschillen van kandidaten op de beide toetsen in 2013 en 2014 mede verklaard kunnen worden door verschillen in motivatie.

3.3.5. Taligheid

De taligheid is op basis van een kleinschalige meting onderzocht. De vijf geanalyseerde itemkoppels zijn te vinden in bijlage 9 en de uitkomst van de analyse staat in bijlage 10. Uit de analyse komt naar voren dat vier items uit het CE wiskunde vmbo bb 2013 qua taligheid complexer zijn dan de bijbehorende items uit de rekentoets vo tijdvak 1 2013 en dat het vijfde item uit de rekentoets vo tijdvak 1 2013 conceptueel makkelijker te volgen is dan het bijbehorende item uit het CE wiskunde vmbo bb 2013. Hieruit kan de voorzichtige conclusie getrokken worden dat het CE wiskunde vmbo bb 2013 qua taligheid complexer is dan de rekentoets vo tijdvak 1 2013.

Het gaat hier om een voorzichtige conclusie omdat vijf vergelijkingen te weinig zijn om te generaliseren. Bovendien ontlopen de items van twee itemkoppels elkaar niet veel qua talige moeilijkheid.

Bovenstaande analyse wekt de indruk dat het CE wiskunde vmbo bb 2013 qua taligheid complexer is dan de rekentoets vo tijdvak 1 2013. Bij onderzoeksvraag 2 is echter aangetoond dat de kandidaten significant lager scoren op de rekentoets vo tijdvak 1 2013 dan op het CE wiskunde vmbo bb 2013. Het vermoeden is dus dat de taligheid van een toets geen negatieve invloed heeft op de score van de kandidaten op die toets en de gevonden scoreverschillen tussen de rekentoets vo en het CE wiskunde vmbo bb of kb niet verklaart.

3.3.6. Cesuur

De cesuren van de rekentoets vo en het CE wiskunde zijn op verschillende manieren bepaald. De rekentoets vo is direct gerelateerd aan het referentieniveau 2F rekenen. De rekentoets vo is een operationalisatie van dit referentieniveau en de cesuur is gekoppeld aan het al dan niet beheersen van 2F rekenen. Het referentieniveau 2F rekenen staat los van of de afnamepopulatie, in dit onderzoek vmbo bb en vmbo kb, dit niveau ook daadwerkelijk beheerst. Het gaat hier om een absoluut criterium. Bij een absoluut criterium is het mogelijk dat een zeer hoog of zeer laag percentage kandidaten aan dat criterium zal voldoen. Bij het CE wiskunde vmbo bb is de referentienorm vastgelegd op een referentie-examen op basis waarvan is bepaald wat redelijkerwijs van deze groep kandidaten verwacht kan worden. Bij het vaststellen van deze norm zijn dus ook expliciet historische gegevens meegenomen op basis waarvan de cesuur is bepaald.

Het vaststellen van de cesuur voor de rekentoets vo en het CE wiskunde kent verschillende uitgangspunten. Deze uitgangspunten leiden ook tot verschillende percentages voldoende. In Tabel 14 staat een overzicht van de percentages voldoende voor vmbo bb en vmbo kb op de rekentoets vo

en het CE wiskunde. Een belangrijke opmerking hierbij is dat bij het vmbo bb in 2013 en 2014 een soepelere norm voor de rekentoets is gehanteerd. Kandidaten van vmbo kb behalen een voldoende indien zij hoger scoren dan het referentieniveau 2F rekenen. Kandidaten van vmbo bb krijgen één cijferpunt extra, zodat een 6 voor een kandidaat van vmbo bb gelijk staat aan een 5 voor een kandidaat van vmbo kb.

Tabel 14

Percentage voldoende per toets/examen en schooltype in tijdvak 1

Leerwegsector en afnamejaar	Rekentoets vo	CE wiskunde
Vmbo bb 2013	25,8%	80,1%
Vmbo kb 2013	25,3%	65,9%
Vmbo bb 2014	29,9%	83,6%
Vmbo kb 2014	27,9%	68,9%
Vmbo bb 2015	40,7%	82,0%
Vmbo kb 2015	50,7%	76,6%

De rekentoets vo en het CE wiskunde vmbo bb en kb verschillen sterk met betrekking tot de percentages voldoende. Er zijn een aantal oorzaken voor deze verschillen aan te wijzen. Niet alle kandidaten zijn bijvoorbeeld verplicht om het CE wiskunde vmbo bb of kb te maken en bij de rekentoets vo lijken kandidaten minder gemotiveerd te zijn. Desalniettemin is het niet waarschijnlijk dat deze twee effecten dit grote verschil kunnen verklaren. In 2013 was de rekentoets vo moeilijker ten opzichte van de centrale examens wiskunde dan in 2014 het geval was. Dat betekent dus dat de rekentoets vo relatief makkelijker is geworden, maar Tabel 14 laat zien dat dit niet heeft geleid tot een hoger percentage voldoende. In 2015 zien we echter een ander beeld, hier valt op dat het slagingspercentage op de rekentoets vo enorm gestegen is ten opzichte van 2013 en 2014. De betere prestaties op de rekentoets zijn toe te schrijven aan een combinatie van twee factoren. Enerzijds is de toets eenvoudiger geworden wat zich uit in een hoger percentage correcte antwoorden dat nodig is om een voldoende te halen. In 2013 kon een vmbo kb kandidaat bij 60% van de opgaven correct een voldoende halen, terwijl in 2015 66,7% van de opgaven correct beantwoord moeten worden door een vmbo kb kandidaat om een voldoende te halen. Daarnaast geldt dat bij het eerste tijdvak een steeds groter deel van de kandidaten een voldoende heeft behaald. De prestaties zijn dus ook in absolute zin verbeterd.

Indien gecontroleerd wordt voor het toepassen van een andere scoringsregel blijkt dat in 2014 het CE wiskunde vmbo kb moeilijker is dan de rekentoets vo voor dezelfde groep kandidaten. Toch blijken veel minder kandidaten voor de rekentoets vo te slagen. Het verschil in moeilijkheid gecombineerd met het verschil in percentages voldoende maakt duidelijk dat de vragen bij de rekentoets vo niet per se moeilijker zijn, maar dat het vooral moeilijker is om er een voldoende voor te halen. De cesuur ligt voor de rekentoets vo hoger dan voor de CE wiskunde; met name voor de vmbo kb kandidaten kunnen beide toetsen als maakbaar worden beschouwd.

4. Conclusie & discussie

De eerste onderzoeksvraag betreft de verschillen tussen de rekentoets vo en het CE wiskunde vmbo bb en kb. De gevonden verschillen tussen de rekentoets vo en het CE wiskunde vmbo bb en kb hebben met name betrekking op de vorm en afdnamecondities van de toetsen. De toetsen verschillen in toetsduur, het aantal items, het al dan niet aanwezig zijn van contextloze items, het gebruik van contexten, de soort rekenmachine die beschikbaar is, de aanwezigheid van een rekenmachine bij de items en het al dan niet kunnen terugbladeren. Deze factoren hebben een rol gespeeld in de analyse van de scoreverschillen. Er is ook een verschil wat betreft de inhoud van de twee toetsen. Alle vaardigheden die bij de rekentoets vo getoetst worden ook getoetst worden in het CE wiskunde vmbo bb en kb behalve de basisvaardigheden. Er zijn echter wel meerdere vaardigheden die op het CE wiskunde vmbo bb of kb getoetst worden die niet voorkomen bij de rekentoets vo. Het instrument dat is ontwikkeld om de inhoudelijke complexiteit van de onderzochte toetsen te bepalen, is in deze fase gebruikt. Er is geprobeerd om aan te tonen dat het instrument betrouwbaar is maar hier doken enkele problemen op. De interbeoordelaarsbetrouwbaarheid is bepaald maar deze was erg laag. Een verklaring hiervoor kan zijn dat het instrument erg ingewikkeld is. Er zijn erg veel verschillende rekenstappen waar aan gedacht moet worden en tevens is elk item net weer iets anders. Dit zorgt ervoor dat het voor een tweede beoordelaar lastig is om het instrument direct op de juiste manier te gebruiken. Het vermoeden is dat er een lange oefenfase nodig is om het instrument op de goede manier te kunnen gebruiken.

De tweede onderzoeksvraag is of er significante scoreverschillen zijn tussen de resultaten van de vmbo bb en kb kandidaten op de rekentoets vo en het CE wiskunde. De conclusie is dat er inderdaad significante verschillen zijn tussen de scores van kandidaten op de rekentoets vo en het CE wiskunde vmbo bb en vmbo kb. Als we ons beperken tot de contextitems van de rekentoets vo en de items van de centrale examens wiskunde vmbo bb en kb die ook bij de rekentoets vo zouden kunnen voorkomen, dan blijkt dat de rekentoets vo slechter is gemaakt dan de centrale examens wiskunde vmbo bb en kb in 2013, 2014 en 2015. Dit verschil is statistisch significant. Het verschil in 2014 tussen de scores op de rekentoets vo en het centrale examen wiskunde vmbo kb is overigens erg klein (in de orde van 0,1 standaardafwijking). Er is een middelmatige samenhang tussen de scores op de rekentoets en de centrale examens van 2013 en 2014.

Ten slotte betreft de derde onderzoeksvraag verklaringen voor de scoreverschillen die gevonden zijn bij de tweede onderzoeksvraag aan de hand van de verschillen tussen de toetsen die gevonden zijn bij de eerste onderzoeksvraag. Deze onderzoeksvraag leidt tot de volgende bevindingen:

- *Tijdnood*

Het is bekend dat op de rekentoets vo 2013 sprake geweest kan zijn van tijdnoed bij de kandidaten. Dit kan een gedeeltelijke verklaring zijn voor de scoreverschillen op de toetsen in 2013, maar in 2014 en 2015 speelt dit niet.

- *Terugbladeren*

Doordat er in 2013 mogelijk sprake was van tijdnoed, en omdat juist bij tijdnoed terugbladeren nodig kan zijn, heeft het al dan niet kunnen terugbladeren mogelijk effect gehad op de scores van de kandidaten. In 2014 was er echter geen tijdnoed waardoor het al dan niet kunnen terugbladeren naar verwachting geen effect heeft gehad. Vanaf 2015 is het in de rekentoets vo mogelijk om terug te bladeren en dus zal deze eventuele verschilfactor dan niet meer aan de orde zijn.

- *Inhoudelijke complexiteit*

De contextitems in de rekentoets vo zijn inhoudelijk complexer – in termen van het aantal benodigde rekenstappen – dan de rekenitems in het examen wiskunde vmbo bb in 2013, 2014 en 2015 en vmbo kb in 2013. Het centrale examen wiskunde vmbo kb in 2014 en 2015 is inhoudelijk vrijwel even complex als de rekentoets vo.

- *Motivatie*

De prestatieverschillen van kandidaten op de beide toetsen in 2013 en 2014 kunnen mede verklaard worden door verschillen in motivatie aangezien (a) een aantal kb-examenitems inhoudelijk complexer is dan rekentoetsitems die door de leerlingen even goed zijn gemaakt en (b) een aantal items van de examens van vmbo bb en kb beter wordt gemaakt dan inhoudelijk even complexe items van de rekentoets:

- a) Van items van de rekentoets vo en de centrale examens wiskunde vmbo bb in 2013 en 2014 die even goed worden gemaakt, schatten experts de inhoudelijke complexiteit even hoog in. Bij kb schatten experts de inhoudelijke complexiteit van de items van het CE wiskunde vmbo kb zelfs hoger in dan die van items die even goed gemaakt zijn op de rekentoets vo.
- b) Items van de rekentoets vo en de centrale examens wiskunde vmbo bb en kb die inhoudelijk even complex zijn worden in 2013 even goed gemaakt. In 2014 geldt er dat de items van de centrale examens wiskunde vmbo bb en kb beter worden gemaakt dan de vergelijkbare items van de rekentoets vo. Hierbij valt echter wel op dat zowel in 2013 als 2014 items met veel rekenstappen in de rekentoets vo beter worden gemaakt dan items met evenveel rekenstappen in de centrale examens wiskunde vmbo bb en kb.

Het is interessant om ook nog te onderzoeken of de motivatie in het afnamejaar 2015 ook een rol speelt. In dit jaar zijn er namelijk kandidaten waarvoor de rekentoets al meetelt in de slaag-/zakregeling. Doordat de rekentoets voor deze kandidaten zo zwaar meetelt, kan het zijn dat de motivatiefactor wegvalt als verklaring voor de gevonden scoreverschillen.

- *Verschillende scoring*

Als methode van goed-fout scoring van het eindantwoord, zoals die bij de rekentoets vo wordt gebruikt, ook bij de wiskunde-examens zou worden toegepast, dan worden de verschillen in gemiddelde score tussen de rekentoets en de centrale examens wiskunde in 2013 en 2014 kleiner. Het centrale examen wiskunde vmbo kb 2014 blijkt dan zelfs moeilijker dan de rekentoets vo.

- *Taligheid*

Op basis van een kleinschalige meting blijkt dat de items van de rekentoets vo minder talig zijn dan de items van de centrale examens wiskunde vmbo bb en kb. Omdat de taligheid van het CE wiskunde eerder hoger lijkt te zijn dan bij de rekentoets vo, lijkt dit geen factor te zijn die het gevonden scoreverschil kan verklaren. Om deze conclusie hard te maken, is het goed om de taligheid van de toetsen verder te onderzoeken.

- *Cesurbepaling*

Voor de centrale examens wiskunde vormt een referentie-examen met een bijbehorend percentage onvoldoende het uitgangspunt voor de cesuur. Voor de rekentoetsen is de cesuur bepaald door het te behalen referentieniveau. Daardoor is het lastiger om aan het referentieniveau rekenen te voldoen dan aan de prestatie-eis die gekoppeld is aan het referentie-examen wiskunde. Hoewel het CE wiskunde vmbo kb moeilijker is, is het daardoor toch eenvoudiger om een voldoende te behalen. Dit is onderzocht voor de afnamejaren 2013 en 2014.

5. Referenties

Bosker, R., & Vorle, R. Van de (Red.) (2014). *Advies over de uitwerking van de referentieniveaus 2F en 3F voor rekenen in toetsen en examens*. Enschede: SLO.

College voor Examens (2011). *Syllabus BB, KB, GT centraal examen 2013*. Utrecht: CvE.

College voor Examens (2012). *Syllabus centraal examen 2014*. Utrecht: CvE.

College voor Examens (2013). *Syllabus centraal examen 2015*. Utrecht: CvE.

College voor Examens (2014). *Rekentoetswijzer 2F 2015*. Utrecht: CvE.

Commissie Meijerink (2008). *Over de drempels met taal en rekenen*. Enschede: SLO.

Dekker, T., & Schmidt, V. (2011). *Rekentoetswijzer 2F voortgezet onderwijs*. Enschede: SLO.

Eggen, T. J. H. M., & Verhelst, N. D. (2011). Item calibration in incomplete testing designs. *Psicológica: International Journal of Methodology and Experimental Psychology*, 32, 107-132.

Evers-Vermeul, J. (2015). Toetsing in context vraagt doordinking van de rol van taal. *De Cascade*, 12, 17-19.

Glas, C. A. W., & Verhelst, N. D. (1989). Extensions of the partial credit model. *Psychometrika*, 54, 635-659.

Hoogland, K., Vliegthart, M., Schoonen, E., Mark, J. Van der., & Reeuwijk, M. Van (2014). *Rekenen en rekenen toetsen op het vmbo BB Twee deelonderzoeken*. Utrecht: APS.

Inspectie van het Onderwijs (2014). *De staat van het onderwijs*. Onderwijsverslag 2012/2013. Utrecht: Inspectie van het Onderwijs.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

Straat, H. (2014). *Terugbladeronderzoek COE 2F Rekenen*. Intern rapport. Arnhem: Cito.

Verhelst, N. D. (1992). *Het één parameter model (OPLM). Een theoretische inleiding en een handleiding bij het computerprogramma*. Arnhem: CITO.

Verhelst, N. D. (1993). Itemresponstheorie. In: T. J. H. M. Eggen & P. F. Sanders (red.), *Psychometrie in de praktijk* (pp. 83-178). Arnhem: Cito.

Verhelst, N. D., & Engelen, R. J. H. (1999). *An ability estimator in the two parameter logistic model based on raw scores*. Research memorandum. Arnhem: Cito.

Verhelst, N. D., & Glas, C. A. W. (1993). A dynamic generalization of the Rasch model. *Psychometrika*, 58, 395-415.

Verhelst, N. D., & Glas, C. A. W. (1995). The one parameter logistic model. In: G.H. Fischer & I.W. Molenaar (Eds.), *Rasch models: Foundations, recent developments and applications* (pp. 215-239). New York: Springer.

Verhelst, N. D., Glas, C. A. W., & Verstralen, H. H. F. M. (1993). *OPLM: One parameter logistic model. Computer program and manual*. Arnhem: Cito.

Verhelst, N. D. & Kleintjes, F. G. M. (1993). Toepassingen van itemresponsetheorie. In: T. J. H. M. Eggen en P. F. Sanders (red.), *Psychometrie in de praktijk*. Arnhem: Cito.

Verhelst, N. D. & Verstralen, H. H. F. M. (2007). *Structural analysis of a univariate latent variable (SAUL)*. Theory and a computer program. Arnhem, Cito.

Verhelst, N. D., Verstralen, H. H. F. M., & Eggen, T. H. J. M. (1991). *Finding starting values for the item parameters and suitable discrimination indices in the one-parameter logistic model*. Measurement and Research Department Reports 91-10. Arnhem: Cito.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427-450.

Bijlagen

Bijlage 1: Criteria voor items

Criteria voor items uit de rekentoets vo om gekoppeld te kunnen worden aan items uit het CE wiskunde vmbo bb en kb:

- De getoetste vaardigheden moeten voorkomen in de syllabi van het wiskunde examen vmbo bb en kb van 2013 (College voor Examens, 2011), 2014 (College voor Examens, 2012) en 2015 (College voor Examens, 2013).
- De vraagvorm moet passen in het CE wiskunde vmbo bb en kb.
 - o Contextloze items kunnen niet opgenomen worden in het CE wiskunde vmbo bb en kb omdat hier geen contextloze items instaan.
 - o De beide vormen van de contextitems uit de rekentoets vo passen wel in het CE wiskunde vmbo bb en kb. Zowel meerkeuze-items als numerieke items zitten namelijk in het CE wiskunde vmbo bb en kb.

Criteria voor items uit het CE wiskunde vmbo bb en kb om gekoppeld te kunnen worden aan items uit de rekentoets vo:

- De benodigde vaardigheden moeten voorkomen in de rekentoetswijzer 2F (Dekker & Schmidt, 2011) voor 2013 en 2014 en de rekentoetswijzer 2F (College voor Examens, 2014) voor 2015. Hieronder enkele voorbeelden van type items die voorkomen in het CE wiskunde vmbo bb of kb maar niet voorkomen in de rekentoets vo:
 - o Er is een formule gegeven met letter in plaats van woorden. Een woordformule kan wel in de rekentoets een letterformule niet.
 - o Er moet gerekend worden met graden. Dit is geen onderdeel van de benodigde vaardigheden voor de rekentoets vo.
 - o Er moet gebruik gemaakt worden van de stelling van Pythagoras. Dit is geen onderdeel van de benodigde vaardigheden voor de rekentoets vo.
 - o Er wordt een vraag gesteld over symmetrie. Dit is geen onderdeel van de benodigde vaardigheden voor de rekentoets vo niveau 2F, in 3F past dit wel maar dit is hier dus niet van toepassing.
- De vraagvorm moet passen in de rekentoets. Hieronder enkele voorbeelden van vraagvormen die voorkomen in het CE wiskunde vmbo bb of kb maar niet voorkomen in de rekentoets vo:
 - o Een vraag die bijvoorbeeld begint met "Leg uit dat...". Kandidaten moeten bij de rekentoets vo met een meerkeuze antwoord of een numeriek antwoord de vraag kunnen beantwoorden en dat is bij deze vraagvorm niet mogelijk.
 - o Een vraag waarin een grafiek getekend moet worden.
- Oppervlaktes berekenen moeten kandidaten in beide toetsen kunnen maar niet van alle figuren. Als er bijvoorbeeld met een cirkel gewerkt moet worden, kan dit alleen ook in de rekentoets vo opgenomen worden als er een formule gegeven is.
- Terugrekenen moeten kandidaten ook in beide toetsen kunnen maar niet voor alle formules. Kwadratische of exponentiële functies kunnen voor een dergelijke vraag niet opgenomen worden in de rekentoets vo.

Bijlage 2: Voorbeelden van items

In Figuur 3 staat een item uit het CE wiskunde vmbo bb 2013 dat aan de criteria voldoet. Je lost het item als volgt op: “Het plastic moet aan alle kanten 5 cm uitsteken dus er moet 5 cm = 0,05 m aan beide kanten bij. Er moet dus in totaal $2 \times 0,05 \text{ m} = 0,10 \text{ m}$ bij. Het plastic moet dus $3 \text{ m} + 0,1 \text{ m} = 3,1 \text{ m}$ breed worden. Brenda heeft minsten $3,1 \times 3,1 = 9,61 \text{ m}^2$ plastic nodig.” Bij dit item worden maten omgezet en er wordt opgeteld en vermenigvuldigd. Dit zijn allemaal vaardigheden die ook in de rekentoets verwacht worden dus dit item voldoet aan de criteria en wordt meegenomen in de analyse. Dit item kan in 4 stappen opgelost worden. Stap 1: Omzetten, 5cm omzetten naar 0,05 m. Stap 2: Vermenigvuldigen, 0,05 m met 2 vermenigvuldigen omdat het er aan beide kanten bij moet. Stap 3: Optellen, 0,10 m bij 3 m optellen dit worden de afmetingen van het vierkante stuk plastic. Stap 4: Vermenigvuldigen, 3,1 m met 3,1 m vermenigvuldigen de oppervlakte is berekend.


Figuur 3

Openbaar Item 10 CE Wiskunde Vmbo bb 2013

Preview Toets
bb-wi-13

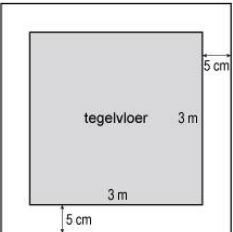
Vraag 10 van 23

Blokhut



(3p) Bereken hoeveel m^2 plastic Brenda minstens nodig heeft.
Typ je berekening in.

Onder de tegelvloer wil Brenda plastic leggen. Dit plastic moet aan alle kanten 5 cm onder de tegelvloer uitsteken.



Navigation bar: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

In Figuur 4 staat een item uit het CE wiskunde vmbo bb 2013 dat niet aan de criteria voldoet. Bij dit item wordt er gevraagd om een grafiek te tekenen. Het tekenen van een grafiek is echter geen vaardigheid uit het referentieniveau 2F en daarom wordt dit item niet meegenomen in de analyse.

Figuur 4

Openbaar Item 14 CE Wiskunde Vmbo bb 2013

Preview Toets
bb-wi-13

Vraag 14 van 23

Chipkaart

Bij aanschaf van de chipkaart staat er een tegoed op van € 75,00.
Elke keer als je met je chipkaart de speeltuin bezoekt, wordt er automatisch € 3,75 van je kaart afgeschreven.

Het bedrag dat nog op je chipkaart staat kun je berekenen met de woordformule

$$\text{bedrag} = 75,00 - \text{aantal bezoeken} \times 3,75$$

Hierbij is *bedrag* in euro.

aantal bezoeken	0	4	8	12	16	20
bedrag (€)	75					

(3p) Teken in het assenstelsel de grafiek die bij deze woordformule hoort. Vul eerst de tabel in.

In Figuur 5 staat een item uit de rekentoets vo 2013 dat niet aan de criteria voldoet. Dit is een contextloos item en kan daarom niet voorkomen in het CE wiskunde vmbo bb of kb. Dit item wordt dus niet meegenomen in de analyse.

In Figuur 6 staat een item uit de rekentoets vo 2013 dat wel aan de criteria voldoet. Je lost het item als volgt op: “Eerst moet je de overbodige informatie eruit filteren zoals de datum van het krantenartikel. Daarna moet je de term anderhalf omzetten in het getal 1,5. Er worden dus minimaal $1,5 \times 1520 = 2280$ hardlopers verwacht in 2013.” Bij dit item moet de juiste informatie geselecteerd worden, vervolgens moet er een term omgezet worden en tot slot wordt er vermenigvuldigd. Dit zijn allemaal vaardigheden die ook in het CE wiskunde vmbo bb en kb verwacht worden dus dit item voldoet aan de criteria en wordt meegenomen in de analyse. Dit item kan in 3 stappen opgelost worden. Stap 1: Bron, er staan overbodige getallen zoals de datum van het krantenartikel in het plaatje. Stap 2: Omzetten term, anderhalf moet omgezet worden naar 1,5. Stap 3: Vermenigvuldigen, 1,5 vermenigvuldigen met 1520.

Figuur 5

Openbaar Item 1 Rekenoets vo 2013

Preview Toets
rvo-2F-vb-13

1 punt
Vraag 1 van 60



779 + 341 =

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Figuur 6

Openbaar Item 2 Rekenoets vo 2013

Preview Toets
rvo-2F-vb-13

1 punt
Vraag 2 van 60



Kranenbrug

Woensdag, 28 november 2012

In 2013 veel deelnemers aan

Er worden in 2013 veel hardlopers verwacht bij de marathon van Kranenburg. Het aantal lopers zal minimaal anderhalf keer zo groot zijn als dit jaar.

In 2012 deden 1520 hardlopers mee aan deze marathon.

Hoeveel hardlopers worden in 2013 minimaal verwacht?

hardlopers



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25

Bijlage 3: Constructie van een OPLM-schaal

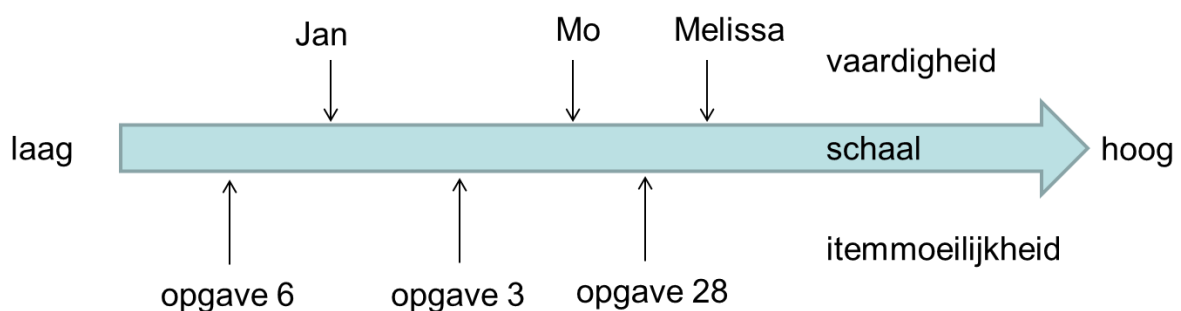
Met psychometrische technieken is het mogelijk de verschillende toetsvarianten met elkaar te vergelijken én af te beelden op een onderliggende vaardigheidsschaal. In deze bijlage geven we aan hoe we aan een vaardigheidsschaal komen en hoe we ermee werken. Het gebruik van een vaardigheidsschaal is vooral geschikt als er meerdere varianten van een toets of examen zijn die men met elkaar wilt vergelijken. De ene variant kan net iets makkelijkere opgaven hebben dan de andere variant. Dit kan gebeuren omdat de moeilijkheid van een opgave niet heel precies ingeschat kan worden voordat deze is voorgelegd aan examenkandidaten. Op de makkelijkere varianten moet een kandidaat dan iets meer vragen goed beantwoord hebben om dezelfde vaardigheid aan te tonen. Dat vinden we eerlijk.

Zodra het om meerdere varianten of versies gaat, worden deze afgebeeld op een vaardigheidsschaal. Aan de hand van een voorbeeld uit de rekentoets vo wordt uitgelegd hoe verschillende versies met elkaar vergeleken kunnen worden.

Bij de analyses is gebruik gemaakt van het *One Parameter Logistic Model* (Verhelst, Glas en Verstralen, 1994) om alle itemmoeilijkheden van overlappende toetsversies en de vaardigheden van de kandidaten af te beelden op één onderliggende schaal, zoals in Figuur A. Het OPLM is een itemresponstheorie-model (IRT-model). De vaardigheden van de kandidaten en de moeilijkheden van de opgaven zijn streepjes op eenzelfde meetlat ofwel schaal. Kandidaten met een hoge vaardigheid hebben een grotere kans om opgaven goed te maken dan kandidaten met een lagere vaardigheid. Omgekeerd, opgaven met een hoge moeilijkheid worden minder vaak goed gemaakt dan opgaven met een lage moeilijkheid. Binnen het OPLM kan ook direct de vaardigheid van een kandidaat vergeleken worden met de moeilijkheid van een opgave. Een kandidaat die even vaardig is als de moeilijkheid van een opgave heeft een kans van 50% om de opgave goed te maken. Melissa heeft in het voorbeeld in Figuur A bij alle drie de opgaven een kans die groter is dan 50% om deze goed te maken.

Figuur A

Afbeelding van personen en opgaven op één vaardigheidsschaal



De IRT beschouwt het antwoord op een item als een indicator voor de mate waarin die vaardigheid aanwezig is. Het verband tussen de vaardigheid en het antwoord op een opgave is van probabilistische aard en wordt weergegeven in de zogenaamde itemresponsfunctie. De itemresponsfunctie geeft aan hoe groot de kans is op een correct antwoord als functie van de onderliggende vaardigheid. Formeler: laat X_i de toevalsvariabele zijn die het antwoord op item i voorstelt. X_i neemt de waarde 1 aan in geval van een correct antwoord en 0 in geval van een fout

antwoord. Als symbool voor de vaardigheid wordt θ (theta) gekozen. De vaardigheid θ is niet rechtstreeks te observeren. Dat zijn alleen de antwoorden op de opgaven. Dat is de reden waarom θ een 'latente' variabele wordt genoemd¹. De itemresponsfunctie $f_i(\theta)$ is gedefinieerd als een conditionele kans:

$$\int f_i(\theta) = P(X_i = 1 | \theta) \quad (1.1)$$

Een IRT-model is een speciale toepassing van (1.1) waarbij aan de functie $f_i(\theta)$ een meer of minder specifieke functionele vorm wordt toegekend. Een eenvoudig en zeer populair voorbeeld is het zogenaamde Raschmodel (Rasch, 1960) waarin $f_i(\theta)$ gegeven is door

$$f_i(\theta) = \frac{\exp(\theta - \beta_i)}{1 + \exp(\theta - \beta_i)} \quad (1.2)$$

waarin β_i de moeilijkheidsparameter van opgave i is. Dat is een onbekende grootte die geschat wordt uit de observaties. De grafiek van (1.2) is weergegeven in Figuur B voor twee items, i en j , die in moeilijkheid verschillen. Deze figuur illustreert dat de itemresponsfunctie een stijgende functie is van θ : hoe groter de vaardigheid, des te groter de kans op een juist antwoord. Indien de latente vaardigheid precies gelijk is aan de moeilijkheidsparameter β_i , volgt

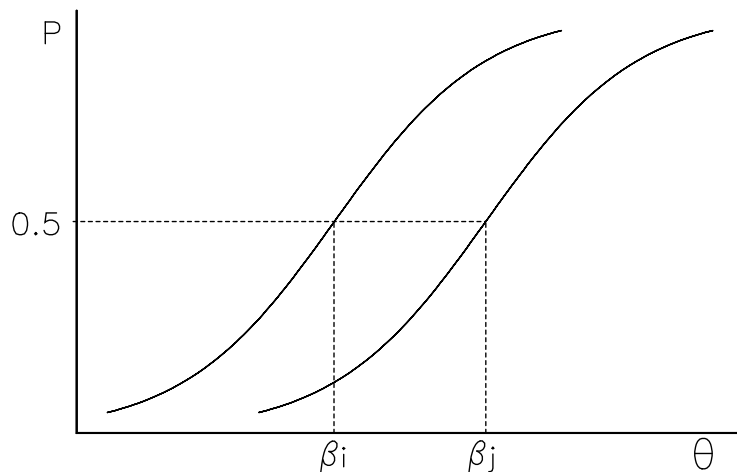
$$f_i(\beta_i) = \frac{\exp(\beta_i - \beta_i)}{1 + \exp(\beta_i - \beta_i)} = \frac{1}{1+1} = \frac{1}{2} \quad (1.3)$$

Daaruit volgt onmiddellijk een interpretatie voor de parameter β_i : het is de 'hoeveelheid' vaardigheid die nodig is voor de kans van precies een half om het item i juist te beantwoorden. Uit de figuur blijkt duidelijk dat voor item j een grotere vaardigheid nodig is om diezelfde kans te bereiken, maar dit is hetzelfde als te zeggen dat item j moeilijker is dan item i . De parameter β_i kan dus terecht omschreven worden als de moeilijkheidsparameter van item i . De implicatie van het bovenstaande is dat 'moeilijkheid' en 'vaardigheid' op dezelfde schaal liggen.

¹ Dit maakt duidelijk waarom men de modellen die ressorteren onder de IRT ook wel aanduidt met 'latente trek'

Figuur B

Twee itemresponscurven in het Raschmodel



Formule (1.2) is geen beschrijving van de werkelijkheid, het is een hypothese over de werkelijkheid die getoetst kan worden op haar houdbaarheid. Hoe zo'n toetsing grofweg verloopt, is te verduidelijken aan de hand van Figuur B. Daaruit blijkt dat, voor welk vaardigheidsniveau dan ook, de kans om item j juist te beantwoorden steeds kleiner is dan de kans op een juist antwoord op item i . Hieruit volgt de statistisch te toetsen voorspelling dat de verwachte proportie juiste antwoorden op item j kleiner is dan op item i in een willekeurige steekproef van personen. Splitst men nu een grote steekproef in bijvoorbeeld twee deelsteekproeven, een 'laaggroep', met de vijftig procent laagste scores, en een 'hooggroep', met de vijftig procent hoogste scores, dan kan men nagaan of de geobserveerde p -waarden van de opgaven in beide deelsteekproeven op dezelfde wijze geordend zijn. Daarvan kan strikt genomen alleen sprake zijn als, in termen van de klassieke testtheorie uitgedrukt, alle opgaven eenzelfde discriminatie-index hebben. Dat echter blijkt lang niet altijd zo te zijn. Ook in ons geval niet. Veel van de opgaven blijken dan ook niet beschreven te kunnen worden met het Raschmodel. Daarom wordt bij de rekentoetsen gekozen voor een ander IRT-model.

Alvorens dit gebruikte model te introduceren, is eerst een kanttekening nodig bij het schatten van de moeilijkheidsparameters in het Raschmodel. Een vaak toegepaste schattingsmethode is de 'conditionele grootste aannemelijkheidsmethode' (in het Engels: Conditional Maximum Likelihood, verder aangeduid als CML). Die maakt gebruik van het feit dat in het Raschmodel een afdoende steekproefgrootte ('sufficient statistic') bestaat voor de latente variabele θ , namelijk de ruwe score of het aantal correct beantwoorde items. Dat betekent grofweg dat, indien de parameters van de opgaven bekend zijn, alle informatie die het antwoordpatroon over de vaardigheid bevat, kan worden samengevat in de ruwe score; het doet er dan verder niet meer toe welke opgaven goed en welke fout zijn gemaakt. Hieruit vloeit voort dat de conditionele kans op een juist antwoord op item i , gegeven de ruwe score, een functie is die alleen afhankelijk is van de parameters van de opgaven en onafhankelijk van de waarde van de vaardigheid, θ . De CML-schattingsmethode maakt van deze functie gebruik. Deze methode maakt geen enkele vooronderstelling over de verdeling van de vaardigheid in de populatie, en is ook onafhankelijk van de wijze waarop de steekproef is getrokken.

2 Een gedetailleerde uiteenzetting hierover kan men vinden in Verhelst, 1992.

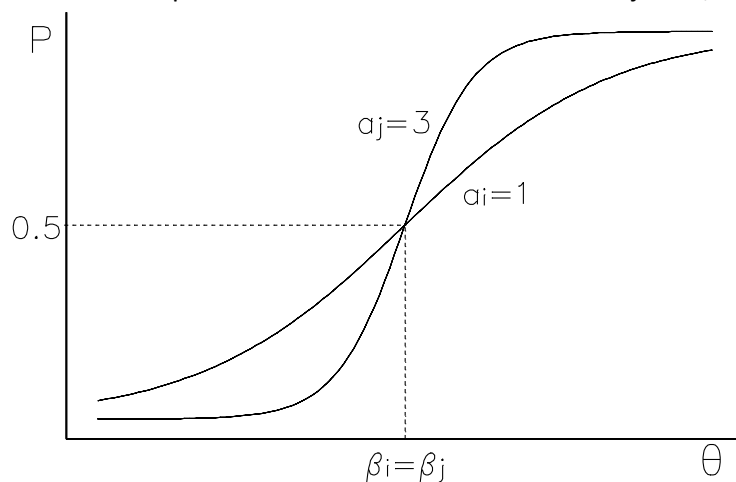
De CML-schattingsmethode is echter niet bij elk meetmodel toepasbaar. In het zogenaamde éénparameter logistisch model (*One Parameter Logistic Model*, afgekort: OPLM) is CML mogelijk. Dit model is, anders dan het Raschmodel, wel bestand tegen 'omwisseling' van 'proporties juist' in verschillende steekproeven (Glas & Verhelst, 1989, Verhelst & Glas, 1993; Eggen & Verhelst, 2011). De itemresponsfunctie van het OPLM is gegeven door

$$f_i(\theta) = \frac{\exp [a_i(\theta - \beta_i)]}{1 + \exp [a_i(\theta - \beta_i)]}, \quad (1.4)$$

waarin a_i de zogenaamde discriminatie-index van het item is. Door deze indices te beperken tot (positieve) gehele getallen, en door ze a priori als constanten in te voeren, is het mogelijk CML-schattingen van de moeilijkheidsparameters β_i te maken. In Figuur C is de itemresponscurve weergegeven van twee items i en j , die even moeilijk zijn maar verschillend discrimineren.

Figuur C

Twee itemresponscurven in het OPLM: zelfde moeilijkheid, verschillende discriminatie-index



De schattingen worden berekend met het computerprogramma OPLM (Verhelst, Glas en Verstralen, 1995). Dit programma voert ook statistische toetsen uit op grond waarvan kan worden bepaald of het model de gegevens adequaat beschrijft. Omdat een aantal van deze toetsen bijzonder gevoelig is voor een verkeerde specificatie van de discriminatie-indices, zijn de uitkomsten van deze toetsen bruikbaar als modificatie-indices: ze geven een aanwijzing in welke richting deze discriminatie-indices moeten worden aangepast om een betere overeenkomst tussen model en gegevens te verkrijgen. Kalibratie van items volgens het OPLM is dan ook een iteratief proces waarin alternerend de modelfit van items wordt onderzocht door middel van statistische toetsing en de waarden van de discriminatie-indices worden aangepast op grond van de resultaten van deze toetsen. Deze aanpassingen geschieden in de praktijk op basis van een en hetzelfde gegevensbestand. Er kan dus kanskapitalisatie optreden. Indien een steekproef een voldoende grootte heeft, is het effect van deze kanskapitalisatie echter gering (Verhelst, Verstralen en Eggen, 1991).

Hoewel het OPLM aanzienlijk flexibeler is dan het Raschmodel, heeft het met dit model toch een nadeel gemeen, waardoor het bij het kalibreren van meerkeuzeopgaven niet zonder meer bruikbaar is. Uit de formules (1.2) en (1.4) volgt dat, indien θ zeer klein is, de kans op een juist antwoord zeer dicht in de buurt van nul komt. Maar sommige opgaven in de examenvarianten zijn meerkeuze-items, zodat blind gokken een zekere kans op een juist antwoord impliceert. Er bestaan modellen die rekening houden met de raadkans (Lord & Novick, 1968), maar die laten geen CML-schattingsmethode toe. De ongeschiktheid van het Raschmodel of OPLM voor meerkeuzevragen is echter relatief: indien de items in vergelijking met de vaardigheid van de leerling niet al te moeilijk zijn, blijkt dat het effect van het raden op de overeenkomst tussen model en gegevens klein is. Door een verstandige dataverzamelingsprocedure toe te passen en met name niet te moeilijke opgaven te selecteren in de toets kan het OPLM toch toegepast worden op meerkeuzevragen, waarbij de overeenkomst tussen model en data de uiteindelijke doorslag over die geschiktheid moet geven. Ook in de normering wordt hiermee rekening gehouden.

Bijlage 4: Wat is een rekenstap?

Om het beoordelen van items, ofwel het tellen van het aantal rekenstappen die nodig zijn voor het oplossen van een rekenitem, zo objectief mogelijk te kunnen uitvoeren is dit document opgesteld. Hierin wordt beschreven wat er onder een rekenstap wordt verstaan. Vooraf is het belangrijk dat de beoordelaar zich realiseert dat het bij de beoordeling gaat om de kwaliteiten van een vmbo leerling. Voor een vmbo leerling is het bijvoorbeeld niet meteen duidelijk dat het verhogen van de prijs met 19% BTW een vermenigvuldiging met 1,19 inhoudt. De beoordelaar moet dus rekening met de parate kennis van een vmbo leerling aan het einde van zijn schooltijd.

De volgende rekenstappen worden onderscheiden:

- *Het aflezen van een bron.*
Let op: Onder bron wordt de gehele tekst en alle toegevoegde afbeeldingen, tabellen en grafieken verstaan! Als er gegevens uit een tekst of afbeelding gehaald moeten worden en deze gegevens staan duidelijk weergegeven dan wordt dit niet gezien als een rekenstap. Als de gegevens echter niet direct af te lezen zijn of als er overbodige getallen in de bron staan, wordt er één rekenstap gerekend. Het kan dus zijn dat er een jaartal in de bron staat die er niet toe doet, dit jaartal telt dan als een overbodig getal en dus wordt hiervoor één rekenstap extra gerekend. Het aflezen van gegevens uit bijvoorbeeld een grafiek wordt ook gezien als één rekenstap. Meerdere (verschillende) gegevens aflezen uit dezelfde grafiek wordt geteld als één rekenstap. Als er uit twee bronnen informatie afgelezen moet worden, zijn dit twee rekenstappen.
- *Het uitvoeren van rekenkundige bewerkingen.*
Hieronder vallen de bewerkingen optellen, aftrekken, vermenigvuldigen en delen. Als dezelfde rekenkundige bewerking bij een item meerdere keren op dezelfde manier uitgevoerd moet worden, wordt dit als één rekenstap geteld. Bijvoorbeeld voor het vermenigvuldigen van meerdere ingrediënten uit een recept, wordt één rekenstap geteld, omdat alles met hetzelfde getal vermenigvuldigt dient te worden. Een ander voorbeeld is dat de te vermenigvuldigen getallen precies op dezelfde manier gerepresenteerd zijn, dan wordt er ook maar één rekenstap geteld. Als bijvoorbeeld in een tabel de getallen uit de eerste kolom steeds met de getallen uit de tweede kolom vermenigvuldigd moeten worden, telt dit als één rekenstap omdat de representatie exact hetzelfde is. Op gelijke manier geldt het optellen van meerdere verschillende bedragen ook als één rekenstap.
- *Het begrijpen van de voorrangsregels.*
Er wordt één extra rekenstap gerekend als het voor een juiste berekening nodig is begrip te hebben van de volgorde waarin rekenkundige bewerkingen uitgevoerd moeten worden. Hierbij gaat het bijvoorbeeld om het juiste gebruik van haakjes. Het gaat hierbij dus om rekenkundige bewerkingen die in sommige rekenmachines niet goed gaan. Voorbeeld:
$$\frac{3+5}{1+3} = 2$$
. Als dit zonder haakjes in een rekenmachine ingetoetst wordt, ziet het er als volgt uit: $3+5/1+3 = 11$.
- *Het afronden van een getal.*
Hier wordt één rekenstap voor gerekend, zowel bij items waarbij het afronden expliciet in de vraag wordt benoemd, maar ook wanneer de context hierom vraagt. Hierbij gaat het bijvoorbeeld om het afronden van geldbedragen op twee decimalen, of het begrip dat er 3 blikken verf gekocht moeten worden als er 2,4 blikken verf nodig zijn om de kamer te schilderen.

- *Passingsproblemen.*
Op het moment dat er sprake is van een passingsprobleem, wordt er één rekenstap gerekend voor de berekening. Als er bijvoorbeeld gevraagd wordt hoe vaak 2 in 10.48 past, geldt dit maar als één rekenstap. Dit is anders dan het werken met een deling en een afronding waarvoor twee rekenstappen worden geteld.
- *Het toepassen van een formule.*
Hierbij wordt het invullen en uitrekenen samen als één rekenstap gezien.
- *Het werken met verhoudingen, breuken of procenten.*
Dit type item kan op veel verschillende manieren worden opgelost en kan dus leiden tot variaties in het benodigd aantal stappen. Hierom is besloten om (naast de uitvoering van de rekenkundige bewerkingen) één rekenstap extra te tellen voor het kunnen ordenen van de gegevens of het begrijpen van de verhouding. Er worden dus twee rekenstappen gerekend als er met verhoudingen, breuken of procenten gewerkt dient te worden. Een voorbeeld hiervan is het opstellen van een verhoudingstabel (stap 1) en het uitvoeren (stap 2). Een uitzondering wordt hierbij gemaakt wanneer bij het opstellen van een verhoudingstabel een van de gegeven getallen het getal 1 is. Bijvoorbeeld bij het berekenen van $1/3$ van 700 wordt toch maar één rekenstap geteld, omdat de oplossing hiervan enkel een deling vereist. Bij het berekenen van $2/3$ van 700 moet er ook nog vermenigvuldigd worden en worden er dus twee rekenstappen geteld.
- *Het omzetten van maten.*
Per omzetting wordt hierbij één rekenstap gerekend. Het gaat hierbij om omzettingen die andere bewerkingen gebruiken. Hiermee wordt bedoeld dat als kilometers naar meters omgezet moeten worden er drie maal met 10 vermenigvuldigd wordt. Dit wordt gezien als één rekenstap. Ook geldt er dat het omzetten van twee verschillende waarden in bijvoorbeeld meters naar kilometers als één rekenstap wordt geteld. Dit omdat de handeling hetzelfde blijft. Het omzetten van dm^3 naar liter wordt ook gezien als een rekenstap. Hier verandert namelijk de eenheid waarmee de hoeveelheid aangeduid wordt.
- *Het rekenen met tijd.*
Voor het omzetten van tijdseenheden gelden andere regels dan voor het omzetten van andere maten. Het omzetten van dagen naar minuten wordt gezien als twee rekenstappen. Voor deze omzetting moet er namelijk eerst vermenigvuldigd worden met 24 en vervolgens met 60, dit zijn twee verschillende bewerkingen. Zo zijn voor het omzetten van dagen naar seconden dus ook twee rekenstappen nodig. Het omzetten van 3u50min naar 210min wordt beschouwd als één rekenstap. Bij het optellen van tijden gelden in principe dezelfde regels als bij het uitvoeren van rekenkundige bewerkingen. Hierbij geldt echter de uitzondering dat er één rekenstap extra gerekend wordt als de optelling van minuten over 60 minuten heen gaat. Er moet dan namelijk beredeneerd worden dat er 1 uur bijgeteld moet worden. Ter illustratie: Het is 07:15 en Janneke heeft 50 minuten nodig om naar school te gaan. Hoe laat is ze op school? Hier geldt voor de optelling één rekenstap en voor het overschrijden van het uur geldt ook één rekenstap. Tot slot wordt de omzetting van 1 uur naar 60 minuten niet gezien als een rekenstap maar als een weetje. Net zoals men weet dat er 52 weken of 365 dagen in een jaar zitten. Een voorbeeld is het omzetten van 3 jaar naar uren. Voor deze omzetting zijn twee rekenstappen nodig: 3 vermenigvuldigen met 365 dagen en de uitkomst hiervan vermenigvuldigen met 24 uur. Echter als er met 2 uur gewerkt moet worden, is het omzetten naar 120 minuten wel één rekenstap.

- *Het omzetten van termen.*
Een term kan hierbij omgezet worden van letters naar cijfers of andersom. Per omzetting wordt één rekenstap gerekend. Bekende termen zijn o.a. miljoen, miljard, ton (zowel voor gewicht als geldbedrag), kuub, anderhalf en kwart. Er staat bijvoorbeeld in een opgave dat er 6 miljoen euro is verdiend en de kandidaat heeft die 6 miljoen euro nodig om tot het antwoord te komen. Het omzetten van de term 6 miljoen naar het getal 6.000.000 telt als één rekenstap. Bij het omzetten van termen geldt dezelfde regel als bij het omzetten van getallen: Als dezelfde omzetting meerdere keren maar met verschillende getallen uitgevoerd wordt, wordt er één rekenstap gerekend.
- *Ruimtelijk inzicht.*
Het herkennen en goed benoemen van ruimtefiguren is een onderdeel van de rekentoetswijzer 2F en dit telt dus ook als 1 rekenstap. Als er in een opgave bijvoorbeeld gevraagd wordt welk ruimtefiguur een object voorstelt, telt dit dus als 1 rekenstap.
- *Een conclusie trekken.*
Als het voor het item nodig is om een conclusie te trekken wordt dit gerekend als één rekenstap. Het gaat hier bijvoorbeeld om een vraag als: "Is abonnement A of abonnement B goedkoper?". Hierbij moet eerst berekend worden hoeveel ieder abonnement kost en vervolgens moet er een conclusie worden getrokken. Als er in de vraag staat: "Laat met een berekening zien dat...." dan hoeft er niet expliciet een conclusie getrokken worden dus dan wordt er geen rekenstap geteld.
- *Het interpreteren van gegevens.*
Hieronder verstaan we het begrip dat nodig is om een item op te kunnen lossen. Als er bijvoorbeeld met tijdsverschil wordt gewerkt, moet een kandidaat weten wat het betekent als het in een ander land 3 uur eerder is. Dit goed interpreteren geldt dan als één rekenstap. Ook als er met negatieve getallen wordt gewerkt, rekenen we één rekenstap extra. Dit komt met name voor in items waar met geld of temperatuur gerekend wordt. Als een kandidaat bijvoorbeeld moet uitrekenen wat het verschil tussen -2 en 13 is, worden er twee rekenstappen gerekend. Één rekenstap voor het begrip van het negatieve getal en één rekenstap voor de berekening.
- *Richting bepalen.*
Het bepalen van de richting telt als één rekenstap. Als er bijvoorbeeld gevraagd wordt welke stad er op de kaart ten zuiden van Arnhem ligt, telt het bepalen van het zuiden als één rekenstap, het lokaliseren van Arnhem (dus het gebruik van de bron) geldt als één rekenstap en de conclusie telt als één rekenstap.
- *Metten.*
Het meten van een plaatje telt als één rekenstap. Als er bijvoorbeeld een schaal en een plaatje gegeven is waarbij de kandidaat zelf de lengte van het plaatje moet opmeten, telt dit opmeten als één rekenstap.

Bijlage 5: Opdracht experts voor bepaling moeilijkheid.

Instructies:

Beste,

Allereerst bedankt dat je naar de complexiteit van de gekoppelde items wilt kijken. Dit zal hopelijk veel bijdragen aan het onderzoek waar ik mee bezig ben.

Op de volgende pagina vind je een invulformulier. In de eerste kolom van dit formulier staat het koppelnummer gegeven. Vervolgens zijn er twee kolommen voor de items, de eerste kolom is voor het bovenste item en de tweede kolom voor het onderste. Ik vraag je om een kruisje te zetten bij het item dat jij van het koppel het meest complex vindt. Je moet dus bepalen wat complexer is: Alle punten halen bij het item van het CE of alle punten halen (ook al is het er maar 1) bij het item van de rekentoets vo. Om te voorkomen dat er onduidelijkheid bestaat, wil ik ook graag dat je de naam van het item erbij zet. Bijvoorbeeld: bb-wi-13 vraag 19 van 24, dit is het bovenste item van het eerste koppel Tevens is er een kolom waarin je een korte toelichting kunt geven. In de toelichting ben ik met name benieuwd naar wat voor jou de doorslag heeft gegeven.

Tot slot wil ik je er nog op attenderen dat je deze bepaling van complexiteit zonder vooroordelen moet doen. Je kunt namelijk aan de items zien van welke toets het item komt, CBT of RVO. Voor het onderzoek is het echter belangrijk dat jullie hier niet naar kijken.

Groetjes,
Franziska

Bijlage 6: Vergelijking rekentoetswijzer 2F en de syllabi CE wiskunde vmbo bb en kb

Rekentoetswijzer 2F

In 2013 en 2014 werd dezelfde rekentoetswijzer 2F (Dekker & Schmidt, 2011) gebruikt. Voor het jaar 2015 is er een nieuwe rekentoetswijzer 2F (College voor Examens, 2014) geschreven. Deze toetswijzer is in grote lijnen gelijk aan de toetswijzer uit 2013 en 2014. Beide rekentoetswijzers 2F beschrijven hoe de rekendoelen uit het referentiekader (Commissie Meijerink, 2014) rekenen, niveau 2F, kunnen voorkomen in de rekentoetsen 2F.

De rekentoetswijzers beschrijven beide de domeinen waaruit items geconstrueerd worden. De vier domeinen zijn: Getallen, Verhoudingen, Meten/Meetkunde en Verbanden. De verdeling over deze domeinen is ongeveer 30% voor Getallen, 30% voor Verhoudingen, 20% voor Meten/Meetkunde en 20% voor Verbanden. Sommige items zullen betrekking hebben op meer dan één domein.

De rekentoets vo bestaat uit contextloze items en contextitems. Onder contextloze items worden items verstaan die betrekking hebben op onbenoemde (kale) getallen. Benoemde getallen zijn bijvoorbeeld geldbedragen als de eenheid is gespecificeerd. Op het moment dat er benoemde getallen in een item staan, is het een contextitem. Bij de contextloze items is het niet toegestaan om een rekenmachine te gebruiken en het niveau van deze items is van referentieniveau 1F. Omdat de rekentoets 2F moet toetsen of de basisvaardigheden uit 1F worden beheerst, zitten deze contextloze items erin. Het overgrote deel van de rekentoets vo bestaat uit contextitems omdat het functioneel gebruik van rekenvaardigheden bij niveau 2F centraal staat. Bij deze vragen moeten de kandidaten vaak gebruik maken van domeinoverstijgende vaardigheden. De domeinoverstijgende vaardigheden zoals beschreven in de rekentoetswijzer 2F zijn:

1. Relevante gegevens uit een situatie identificeren.
 2. De gegevens weergeven in een geschikte representatie (bijvoorbeeld tekening, tabel, grafiek) of in een geschikt rekenmodel.
 3. De noodzakelijke rekenkundige vaardigheden toepassen om een gewenst resultaat te verkrijgen.
 4. De resultaten interpreteren in termen van de situatie.
- (Dekker & Schmidt, 2011, p. 8)

Een groot verschil tussen de rekentoetswijzer van 2013 en 2014 en de rekentoetswijzer van 2015 is dat de contextloze items in 2015 ook het niveau van het referentieniveau 2F kunnen zijn. In 2013 en 2014 kon dit enkel het niveau van referentieniveau 1F zijn.

Syllabi CE wiskunde vmbo bb en kb

De syllabus CE wiskunde voor vmbo bb en kb uit 2013 (College voor Examens, 2011) is exact hetzelfde als de syllabus CE wiskunde voor vmbo bb en kb uit 2014 (College voor Examens, 2012) en de syllabus CE wiskunde voor vmbo bb en kb uit 2015 (College voor Examens, 2013). Met onderlinge verschillen tussen de twee jaren hoeft dus geen rekening gehouden te worden.

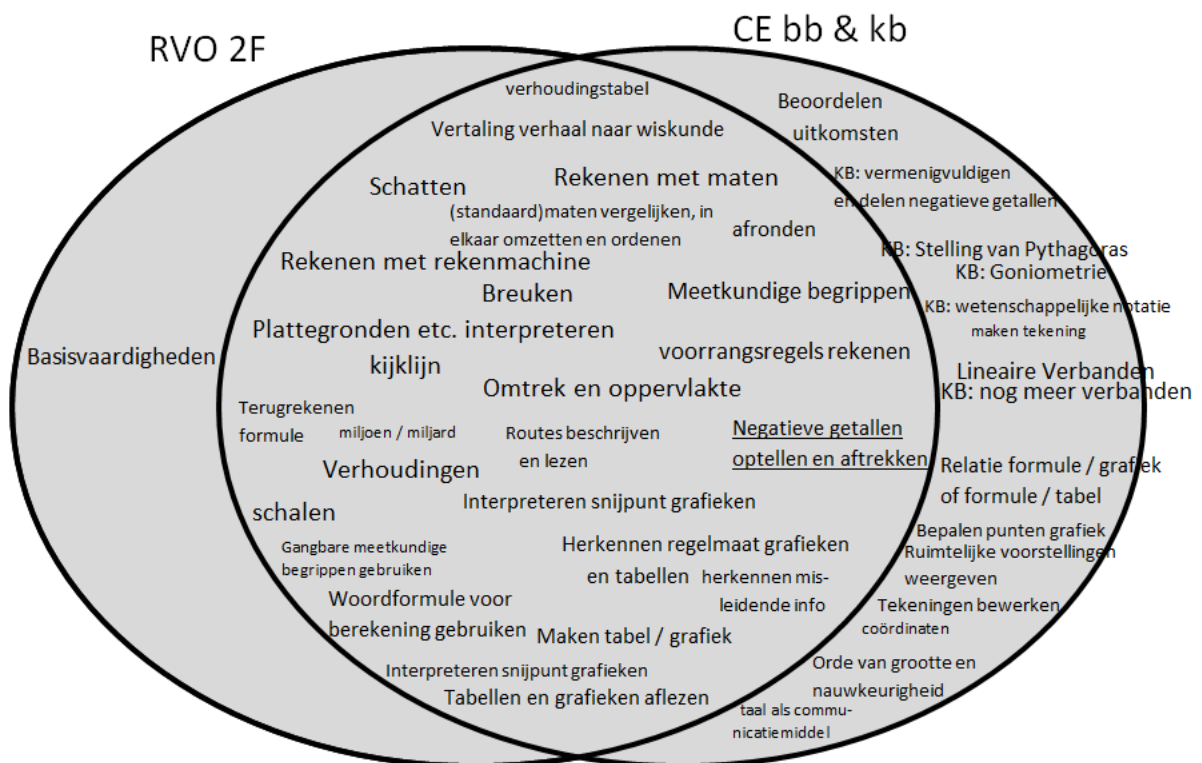
De syllabus (College voor Examens, 2011) beschrijft wat de exameneisen voor het CE zijn. In de syllabus staat verder informatie over:

... specificaties van examenstof, begrippenlijsten, bekend veronderstelde onderdelen van domeinen of exameneenheden die verplicht zijn op het schoolexamen, bekend veronderstelde voorkennis uit de onderbouw, bijzondere vormen van examinering (zoals computerexamens), voorbeeldopgaven, toelichting op de vraagstelling, toegestane hulpmiddelen. (College voor Examens, 2011, p. 4)

In het CE wiskunde vmbo bb en kb komen de globale eindtermen Leervaardigheden in het vak wiskunde, Algebraïsche vaardigheden, Rekenen, meten en schatten en Meetkunde voor. De basisvaardigheden worden echter niet meer apart getoetst op het CE. Voor de specificaties van de eindtermen kan er in de betreffende syllabus gekeken worden.

Overeenkomsten en verschillen

De overeenkomsten en verschillen tussen de rekentoetswijzer 2F (Dekker & Schmidt, 2011) en de syllabus CE wiskunde vmbo bb en kb (College voor Examens, 2011) zijn hieronder te zien in het diagram. Hierbij is het onderstreepte gedeelte geen onderdeel meer van de rekentoetswijzer 2F in 2015.



Bijlage 7: Gegevens rekentoets vo en CE wiskunde vmbo bb en kb

Rekentoets vo 2F 2013 tijdvak 1 en tijdvak 2 alle varianten

- De toets wordt afgenomen in examentester.
- De toetsduur bedraagt 90 minuten.
- De toets bestaat uit 60 items:
 - o 12 contextloze items.
 - o 48 contextitems, numeriek antwoord of meerkeuze.
- De maximumscore bedraagt 60 punten, 1 punt per item. Je antwoord is goed of fout.
- De Computer Rekenmachine (CRM) mag enkel bij de contextitems gebruikt worden.
- Bij elk item wordt er een nieuwe context gegeven.
- De uitleg aan het begin van de toets bestaat uit:
 - o Instructie over de twee beschikbare knoppen. Dit zijn de knop voor de volgende vraag en de knop voor de rekenmachine.
 - o De mededeling dat er kladpapier gebruikt mag worden.
 - o Instructie over het invoeren van getallen. Hierbij gaat het om negatieve getallen en het feit dat er geen punt in je antwoord mag staan. Bijvoorbeeld: niet 31.342,45 maar 31342,45.
 - o De waarschuwing dat het niet mogelijk is om terug te bladeren.

CE wiskunde vmbo bb 2013 alle varianten

- De toets wordt afgenomen in examentester.
- De toetsduur bedraagt 90 minuten.
- De toets bestaat uit 22 tot 24 items.
- De maximumscore bedraagt 54 punten. Er zijn 1, 2, 3 of 4 punten per item te behalen en je kan een gedeelte van de punten krijgen.
- Er is geen CRM ingebouwd. Kandidaten mogen hun eigen rekenmachine gedurende de hele toets gebruiken.
- De items zijn verdeeld over 6 contexten. Er zijn meerdere items per context.
- De uitleg aan het begin van de toets bestaat uit:
 - o De mededeling dat er kladpapier gebruikt mag worden.
 - o Instructie over de beschikbare knoppen. Dit zijn de knop voor het volume, de knop voor het invoeren van speciale tekens en de knop voor de volgende vraag.
 - o De mededeling dat het toegestaan is om je eigen rekenmachine te gebruiken.
- Aan het einde van de toets wordt er gemeld dat het toegestaan is om terug te bladeren.

CE wiskunde vmbo kb 2013 alle varianten

- De toets wordt afgenomen in examentester
- De toetsduur bedraagt 120 minuten.
- De toets bestaat uit 23 of 24 items.
- De maximumscore bedraagt 69 punten. Er zijn 1, 2, 3, 4 of 5 punten per item te behalen en je kan een gedeelte van de punten krijgen.
- De CRM is ten allen tijde beschikbaar.
- De items zijn verdeeld over 6 contexten. Er zijn meerdere items per context.
- De uitleg aan het begin van de toets bestaat uit:

- De mededeling dat er kladpapier gebruikt mag worden.
- Instructie over de beschikbare knoppen. Dit zijn de knop voor het volume, de knop voor het invoeren van speciale tekens, de knop voor de volgende vraag en de knop voor de rekenmachine.
- De mededeling dat het toegestaan is om je eigen rekenmachine te gebruiken.
- Enkele functies van de toolbox in combinatie met de CRM worden toegelicht.
- Aan het einde van de toets wordt er gemeld dat het toegestaan is om terug te bladeren.

Rekentoets vo 2F en bb variant 2014 tijdvak 1 en tijdvak 2 alle varianten

- De toets wordt afgenomen in examentester.
- De toetsduur bedraagt 90 minuten.
- De toets bestaat uit 51 items:
 - 15 contextloze items.
 - 36 contextitems, numeriek antwoord of meerkeuze.
- De maximumscore bedraagt 51 punten, 1 punt per item. Je antwoord is goed of fout.
- De Computer Rekenmachine (CRM) mag enkel bij de contextitems gebruikt worden.
- Bij elk item wordt er een nieuwe context gegeven.
- De uitleg aan het begin van de toets bestaat uit:
 - Instructie over de twee beschikbare knoppen. Dit zijn de knop voor de volgende vraag en de knop voor de rekenmachine.
 - De mededeling dat er kladpapier gebruikt mag worden.
 - Instructie over het invoeren van getallen. Hierbij gaat het om negatieve getallen en het feit dat er geen punt in je antwoord mag staan. Bijvoorbeeld: niet 31.342,45 maar 31342,45.
 - De waarschuwing dat het niet mogelijk is om terug te bladeren.

CE wiskunde vmbo bb 2014 alle varianten

- De toets wordt afgenomen in examentester.
- De toetsduur bedraagt 90 minuten.
- De toets bestaat uit 22 tot 26 items.
- De maximumscore bedraagt 54 punten. Er zijn 1, 2, 3 of 4 punten per item te behalen en je kan een gedeelte van de punten krijgen.
- Er is geen CRM ingebouwd. Kandidaten mogen hun eigen rekenmachine gedurende de hele toets gebruiken.
- De items zijn verdeeld over 6 contexten. Er zijn meerdere items per context.
- De uitleg aan het begin van de toets bestaat uit:
 - De mededeling dat er kladpapier gebruikt mag worden.
 - Instructie over de beschikbare knoppen. Dit zijn de knop voor het volume, de knop voor het invoeren van speciale tekens en de knop voor de volgende vraag.
 - De mededeling dat het toegestaan is om je eigen rekenmachine te gebruiken.
- Aan het einde van de toets wordt er gemeld dat het toegestaan is om terug te bladeren.

CE wiskunde vmbo kb 2014 alle varianten

- De toets wordt afgenomen in examentester
- De toetsduur bedraagt 120 minuten.

- De toets bestaat uit 20 tot 22 items.
- De maximumscore bedraagt 63 tot 67 punten, dit verschilt per variant. Er zijn 1, 2, 3, 4 of 5 punten per item te behalen en je kan een gedeelte van de punten krijgen.
- De CRM is ten allen tijde beschikbaar.
- De items zijn verdeeld over 6 contexten. Er zijn meerdere items per context.
- De uitleg aan het begin van de toets bestaat uit:
 - o De mededeling dat er kladpapier gebruikt mag worden.
 - o Instructie over de beschikbare knoppen. Dit zijn de knop voor het volume, de knop voor het invoeren van speciale tekens, de knop voor de volgende vraag en de knop voor de rekenmachine.
 - o De mededeling dat het toegestaan is om je eigen rekenmachine te gebruiken.
- Enkele functies van de toolbox in combinatie met de CRM worden toegelicht.
- Aan het einde van de toets wordt er gemeld dat het toegestaan is om terug te bladeren.

Rekentoets vo 2F 2015 tijdvak 1 alle varianten

- De toets wordt afgenomen in examentester en facet. De leerlingen zullen hier geen verschil in merken.
- De toetsduur bedraagt 90 minuten.
- De toets bestaat uit 45 items:
 - o 15 contextloze items.
 - o 30 contextitems, numeriek antwoord of meerkeuze.
- De maximumscore bedraagt 45 punten, 1 punt per item. Je antwoord is goed of fout.
- De eerste 15 items zijn contextloze items, binnen deze items kan teruggebladerd worden.
- De laatste 30 items zijn contextitems, binnen deze items kan teruggebladerd worden.
- Als item 16 is geopend kan er niet meer teruggebladerd worden naar items 1 tot en met 15.
- De Computer Rekenmachine (CRM) mag enkel bij de contextitems gebruikt worden.
- Bij elk item wordt er een nieuwe context gegeven.
- De uitleg aan het begin van de toets bestaat uit:
 - o Instructie over de twee beschikbare knoppen. Dit zijn de knop voor de volgende vraag en de knop voor de rekenmachine.
 - o De mededeling dat er kladpapier gebruikt mag worden.
 - o Instructie over het invoeren van getallen. Hierbij gaat het om negatieve getallen en het feit dat er geen punt in je antwoord mag staan. Bijvoorbeeld: niet 31.342,45 maar 31342,45.
 - o De waarschuwing dat het niet mogelijk is om terug te bladeren.

CE wiskunde vmbo bb 2015 alle varianten

- De toets wordt afgenomen in examentester.
- De toetsduur bedraagt 90 minuten.
- De toets bestaat uit 21 tot 25 items.
- De maximumscore bedraagt 54 punten. Er zijn 1, 2, 3 of 4 punten per item te behalen en je kan een gedeelte van de punten krijgen.
- Er is geen CRM ingebouwd. Kandidaten mogen hun eigen rekenmachine gedurende de hele toets gebruiken.
- De items zijn verdeeld over 6 contexten. Er zijn meerdere items per context.

- De uitleg aan het begin van de toets bestaat uit:
 - o De mededeling dat er kladpapier gebruikt mag worden.
 - o Instructie over de beschikbare knoppen. Dit zijn de knop voor het volume, de knop voor het invoeren van speciale tekens en de knop voor de volgende vraag.
 - o De mededeling dat het toegestaan is om je eigen rekenmachine te gebruiken.
- Aan het einde van de toets wordt er gemeld dat het toegestaan is om terug te bladeren.

CE wiskunde vmbo kb 2015 alle varianten

- De toets wordt afgenomen in examentester.
- De toetsduur bedraagt 120 minuten.
- De toets bestaat uit 21 tot 23 items.
- De maximumscore bedraagt 66 punten. Er zijn 1, 2, 3, 4 of 5 punten per item te behalen en je kan een gedeelte van de punten krijgen.
- De CRM is ten allen tijde beschikbaar.
- De items zijn verdeeld over 6 contexten. Er zijn meerdere items per context.
- De uitleg aan het begin van de toets bestaat uit:
 - o De mededeling dat er kladpapier gebruikt mag worden.
 - o Instructie over de beschikbare knoppen. Dit zijn de knop voor het volume, de knop voor het invoeren van speciale tekens, de knop voor de volgende vraag en de knop voor de rekenmachine.
 - o De mededeling dat het toegestaan is om je eigen rekenmachine te gebruiken.
- Enkele functies van de toolbox in combinatie met de CRM worden toegelicht.
- Aan het einde van de toets wordt er gemeld dat het toegestaan is om terug te bladeren.

Bijlage 8: Scores per koppel items met gelijke inhoudelijke complexiteit

Tabel 15

Uitkomst t-toetsen op itemkoppels van CE wiskunde vmbo bb en de rekentoets vo 2013

Koppelnummer	M(RVO) (SD(RVO))	M(CE) (SD(CE))	t-waarde	p-waarde
1	.80 (.402)	.75 (.436)	t(232) = -1.327	.186
2	.44 (.497)	.92 (.266)	t(222) = 12.522	.000
3	.68 (.468)	.80 (.399)	t(419) = 4.494	.000
4	.83 (.379)	.35 (.479)	t(231) = -12.441	.000
5	.58 (.495)	.47 (.499)	t(482) = -3.282	.001
6	.12 (.330)	.53 (.031)	t(257) = 11.457	.000
7	.11 (.307)	.73 (.444)	t(474) = 25.507	.000
8	.10 (.304)	.11 (.317)	t(185) = .333	.740

Tabel 16

Uitkomst t-toetsen op itemkoppels van CE wiskunde vmbo kb en de rekentoets vo 2013

Koppelnummer	M(RVO) (SD(RVO))	M(CE) (SD(CE))	t-waarde	p-waarde
1	.57 (.496)	.56 (.496)	t(642) = -.171	.865
2	.31 (.463)	.63 (.482)	t(652) = 13.474	.000
3	.90 (.294)	.50 (.501)	t(241) = -11.345	.000
4	.30 (.458)	.89 (.308)	t(141) = 13.719	.000
5	.75 (.431)	.15 (.354)	t(607) = -27.401	.000
6	.77 (.424)	.73 (.443)	t(456) = -1.190	.235
7	.17 (.380)	.78 (.416)	t(554) = 26.395	.000
8	.09 (.282)	.17 (.380)	t(805) = 5.741	.000

Tabel 17

Uitkomst t-toetsen op itemkoppels van CE wiskunde vmbo bb en de rekentoets vo 2014

Koppelnummer	M(RVO) (SD(RVO))	M(CE) (SD(CE))	t-waarde	p-waarde
1	.69 (.465)	.88 (.330)	t(508) = 7.697	.000
2	.42 (.494)	.91 (.282)	t(219) = 13.093	.000
3	.08 (.267)	.27 (.443)	t(246) = 6.053	.000
4	.35 (.477)	.97 (.182)	t(322) = 21.767	.000
5	.38 (.485)	.64 (.479)	t(685) = 11.084	.000
6	.83 (.373)	.56 (.026)	t(365) = -8.613	.000
7	.02 (.152)	.08 (.267)	t(297) = 3.066	.002
8	.13 (.340)	.34 (.029)	t(263) = 6.310	.000

Tabel 18

Uitkomst t-toetsen op itemkoppels van CE wiskunde vmbo kb en de rekentoets vo 2014

Koppelnummer	M(RVO) (SD(RVO))	M(CE) (SD(CE))	t-waarde	p-waarde
1	.46 (.499)	.75 (.435)	t(537) = 11.104	.000
2	.12 (.325)	.48 (.500)	t(973) = 19.421	.000
3	.36 (.480)	.64 (.482)	t(1937) = 18.642	.000
4	.48 (.500)	.57 (.388)	t(1666) = 6.583	.000
5	.20 (.400)	.23 (.423)	t(576) = 1.429	.153
6	.37 (.484)	.56 (.399)	t(976) = 9.942	.000
7	.06 (.230)	.55 (.357)	t(1452) = 45.881	.000
8	.51 (.500)	.46 (.499)	t(573) = -1.977	.048

Tabel 19

Uitkomst t-toetsen op itemkoppels van CE wiskunde vmbo bb en de rekentoets vo 2015

Koppelnummer	M(RVO) (SD(RVO))	M(CE) (SD(CE))	t-waarde	p-waarde
1	.56 (.497)	.98 (.155)	t(810) = 22.774	.000
2	.49 (.500)	.77 (.348)	t(561) = 11.696	.000
3	.15 (.355)	.32 (.338)	t(230) = 6.109	.000
4	.783 (.413)	.753 (.343)	t(514) = -1.263	.207
5	.171 (.377)	.425 (.419)	t(339) = 9.313	.000
6	.084 (.278)	.833 (.373)	t(879) = 49.717	.000
7	.579 (.494)	.492 (.500)	t(776) = -3.643	.000
8	.280 (.449)	.332 (.454)	t(607) = 2.270	.024

Tabel 20

Uitkomst t-toetsen op itemkoppels van CE wiskunde vmbo kb en de rekentoets vo 2015

Koppelnummer	M(RVO) (SD(RVO))	M(CE) (SD(CE))	t-waarde	p-waarde
1	.819 (.385)	.586 (.493)	t(734) = -10.493	.000
2	.778 (.416)	.730 (.343)	t(797) = -2.667	.008
3	.439 (.496)	.846 (.290)	t(1093) = 25.260	.000
4	.375 (.484)	.505 (.500)	t(925) = 6.099	.000
5	.962 (.192)	.931 (.254)	t(706) = -2.603	.009
6	.908 (.291)	.585 (.473)	t(129) = -6.808	.000
7	.853 (.355)	.724 (.308)	t(955) = -8.597	.000
8	.964 (.188)	.792 (.382)	t(656) = -10.220	.000

Bijlage 9: Itemkoppels voor analyse van taligheid

Koppel 1:

Preview Toets
RVO-2F-13

1 punt
Vraag 7 van 61



2500 kg aardappels per uur
24 uur per dag

Deze machine verwerkt aardappels tot frites. De machine werkt dag en nacht door.

Hoeveel ton aardappels verwerkt de machine per dag?

ton



1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

Preview Toets
bb-wi-13

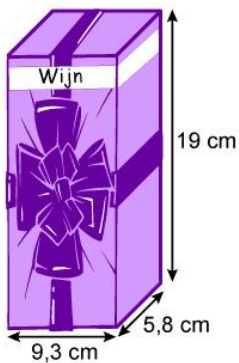
Vraag 21 van 23



Fles of pak

Mario wil zijn wijn ook in kartonnen pakken van 1 liter gaan verkopen.

Hier zie je een tekening van het pak.
De maten zijn in centimeters.



Er geldt: 1 liter = 1000 cm³

(3p) Laat met een berekening zien dat in dit pak 1 liter wijn past.
Typ je berekening in.

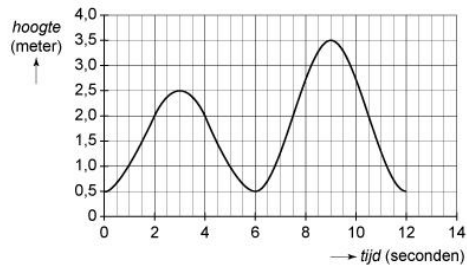
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23

Koppel 2:



Schommelboot

Je ziet de grafiek van de eerste slingerbeweging van de boot. Die eerste slingerbeweging duurt 12 seconden. Deze grafiek geeft het verband aan tussen de *tijd* in seconden en de *hoogte* boven de grond in meter van Hans en Marieke.



(3p) Leg uit hoeveel seconden Hans en Marieke tijdens de eerste 12 seconden minstens 2 meter hoog zijn. Typ je uitleg in.

Navigation bar with question numbers 1 through 23. Question 2 is highlighted.



ENTREETICKETS	KASSA	ONLINE
Kinderen t/m 5 jaar	Gratis	Gratis
Kinderen vanaf 6 t/m 11 jaar	€ 28,50	€ 27,50
Volwassenen en kinderen (12+)	€ 31,00	€ 30,00
Senioren (55+)	€ 28,50	€ 27,50
Mindervalide	€ 28,50	€ 27,50
Parkeerticket	€ 7,50	€ 7,50

Tim is 16 jaar en gaat samen met zijn vader (48), moeder (47), broertje (14) en zusje (10) naar het pretpark Walibi Holland. Ze gaan met de auto en parkeren op het Walibi terrein. Ze betalen alles online.

Hoeveel moeten ze in totaal betalen?

euro



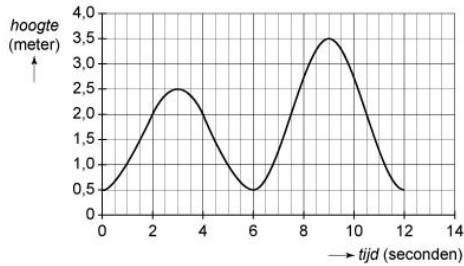
Navigation bar with question numbers 1 through 24. Question 11 is highlighted.

Koppel 3:



Schommelboot

Je ziet de grafiek van de eerste slingerbeweging van de boot. Die eerste slingerbeweging duurt 12 seconden. Deze grafiek geeft het verband aan tussen de *tijd* in seconden en de *hoogte* boven de grond in meter van Hans en Marieke.



(3p) Leg uit hoeveel seconden Hans en Marieke tijdens de eerste 12 seconden minstens 2 meter hoog zijn. Typ je uitleg in.

Navigation bar with question numbers 1 through 23. Question 2 is highlighted.



NS-Jaarabonnement per 1 januari 2012			
2e klas betaling ineens	2e klas betaling per maand *)	1e klas betaling ineens	1e klas betaling per maand *)
3.581,00	373,00	6.088,00	634,10

*) U betaalt 10 maanden per jaar

Sam koopt een NS-Jaarabonnement 2^e klas en betaalt het hele bedrag ineens.
Jolien koopt een NS-Jaarabonnement 2^e klas en betaalt een bedrag per maand.

Hoeveel is Sam goedkoper uit dan Jolien?

euro



Navigation bar with question numbers 1 through 24. Question 14 is highlighted.

Koppel 4:

velux Dakramen (bxh)							
	55x78 cm C02	55x98 cm C04	55x118 cm C06	78x98 cm M04	78x118 cm M06	78x140 cm M08	94x160 cm P10
GGL Grenen	€ 479,-	€ 519,-	€ 585,-	€ 559,-	€ 585,-	€ 605,-	€ 669,-
GGU Kunststof	€ 519,-	€ 559,-	x	€ 599,-	€ 645,-	€ 679,-	€ 749,-
GPL Grenen	x	x	x	x	€ 755,-	€ 769,-	€ 855,-
GHL Grenen	x	€ 615,-	x	€ 665,-	€ 695,-	€ 719,-	x
GPU Kunststof	x	€ 699,-	x	€ 759,-	€ 815,-	€ 859,-	€ 945,-

Henk bestelt een Velux Dakraam van 94x160 cm type GPU kunststof.

Wat is de prijs inclusief 19% BTW?

€



Alle prijzen zijn exclusief 19% BTW.

Triathlon

Stephen doet mee aan de $\frac{3}{8}$ triathlon van Wijdeveen. Stephen moet 1460 m zwemmen.

Bij een $\frac{3}{8}$ triathlon zou hij eigenlijk 1425 m moeten zwemmen. Dat mag volgens de regels ook 5% meer of 5% minder zijn.



(3p) Laat met een berekening zien of het onderdeel zwemmen bij deze $\frac{3}{8}$ triathlon aan de regels voldoet. Typ je berekening in.

Koppel 5:

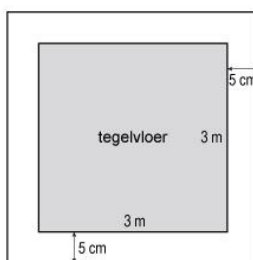


Blokhut



(3p) Bereken hoeveel m^2 plastic Brenda minstens nodig heeft.
Typ je berekening in.

Onder de tegelvloer wil Brenda plastic leggen. Dit plastic moet aan alle kanten 5 cm onder de tegelvloer uitsteken.



VOORJAARSLOTERIJ

**Totaal prijzengeld:
2 miljoen euro!!**

De prijzen:
1 x 1 miljoen euro
5 x 100 000 euro
15 x 10 000 euro
100 x 1000 euro

De overige prijzen zijn elk 500 euro

Hoeveel prijzen van 500 euro zijn er?

prijzen



Bijlage 10: Uitkomst analyse van taligheid

Analyse door Jacqueline Evers-Vermeul.

Koppel 1

Deze twee items ontlopen elkaar niet veel qua talige moeilijkheid. Ik denk dat item 21 net iets complexer is.

Bij **item 7** is een onnodige ambiguïteit gecreëerd door in de zin direct voor de vraag het woord *dag* op een andere manier te gebruiken dan de manier waarop het woord *dag* in de vraag zelf gebruikt wordt. De vraag zelf refereert naar de 24-urige dag (zoals ook onder de foto wordt aangegeven); de zin ervoor zou je kunnen interpreteren als een 12-urige dag.

Leerlingen komen hier verder waarschijnlijk alleen in de problemen als ze de betekenis van het woord *ton* niet kennen, maar ik vermoed dat dit juist bij de kennis hoort die bij dit vak getoetst wordt. Tot slot is *friet* of *patat* misschien bekender dan *frites*.

Bij **item 21** zou met name de vraag/opdracht zelf eenvoudiger geformuleerd kunnen worden.

Bijvoorbeeld: In dit pak past 1 liter wijn. Laat dit met een berekening zien.

(*Typ je berekening in* zou ik achterwege laten -> verdubbeling van informatie.)

Deze vraag kan ook op detailniveau aangescherpt worden:

- de titel *fles* of *pak* is overbodig; leerlingen hoeven daar in deze vraag helemaal geen keuze over te maken.
- vermoedelijk is deze vraag onderdeel van een cluster; anders is het woord *ook* in de eerste zin overbodig.
- *het pak* suggereert dat dit specifieke pak al eerder genoemd is; gebruik daar liever *zo'n pak*
- de opmerking over de maten is overbodig.

Opmerkelijk is overigens dat item 7 de cruciale gegevens op de linkerhelft onder de foto aanbiedt (2500 kg aardappels per uur, 24 uur per dag), terwijl item 21 dit op de rechterhelft van het scherm doet, bovenaan (1 liter = ...). Dit is niet consistent.

Koppel 2

Item 11 is conceptueel waarschijnlijk makkelijker te volgen, omdat de situatieschets daarin meer aansluit bij de alledaagse werkelijkheid dan de situatieschets in item 2. Item 11 lijkt me dan ook een beter voorbeeld van een betekenisvol rekenen dan item 2. Dat er in item 11 personages voorkomen die daadwerkelijk iets doen, draagt hier ook aan bij. Item 2 bevat weliswaar een verwijzing naar Hans en Marieke, maar die spelen geen actieve rol in het geheel.

Detailopmerkingen over **item 2**:

- vermoedelijk is dit item er één uit een cluster; anders komt *de boot* uit de lucht vallen.
- De 3e zin is vrij lang. Ik vraag me bovendien af of deze uitleg nodig is. Is het geen onderdeel van hun vaardigheden dat leerlingen dit soort grafieken kunnen aflezen?
- Ik vind de personages Hans en Marieke hier niets toevoegen, en zou liever gewoon naar de boot verwijzen. Bijvoorbeeld zoals in de volgende herschrijving: *In de grafiek staat de eerste slingerbeweging van de boot. Deze beweging duurt 12 seconden. Voor elk tijdstip zie je hoe hoog de boot boven de grond is.*

- De vraag is behoorlijk informatiedicht. Ik zou 'm vereenvoudigen door een deel van informatie in een zin ervoor te noemen. Bijvoorbeeld:
*Tijdens de eerste 12 seconden is de boot soms 2 meter boven de grond of nog hoger.
Hoeveel seconden is de boot minstens 2 meter hoog? Typ je antwoord met uitleg in.*

Detailopmerkingen over **item 11**:

- Ik vind het logischer als ook de situatieschets van Tim en zijn familie op de linkerhelft van het scherm zou staan.
- de 1e zin zou ik splitsen: *Tim is 16 jaar. Hij gaat samen...*
- de chronologische volgorde in het stuk is niet logisch: eerst stappen ze in de auto en pas daarna wordt genoemd dat ze online betalen... Misschien oplossen door te zeggen: *Ze kopen de tickets vooraf en betalen alles online.*

Koppel 3

Talig gezien is vraag 14 eenvoudiger dan vraag 2.

Je zou **item 14** nog iets kunnen vereenvoudigen door de informatie *per 1 januari 2012* achterwege te laten (deze voegt niets toe). Verder veronderstelt de formulering van de vraag dat Sam goedkoper uit is dan Jolien, iets wat de leerlingen tot dan toe niet hadden kunnen weten.

Koppel 4

Het taalgebruik in item 12 is eenvoudiger dan dat in item 7.

Bij **item 12** zou ik de zin over Henk boven het prijsoverzicht zetten en een expliciete link leggen met dat schema. Bijvoorbeeld: *Henk bestelt... kunststof. Hij kijkt in het schema wat hij hiervoor moet betalen.*

Bij **item 7** kan met name de formulering van de vraag zelf eenvoudiger; deze is nu te informatiedicht. Afhankelijk van de vraag of je wilt weggeven dat deze triatlon aan de regels voldoet, kun je 'm als volgt herschrijven:

Voldoet het onderdeel zwemmen van deze triatlon aan de regels? Typ je antwoord met berekening in.

Overige punten:

- is het noodzakelijk om steeds te noemen dat het een 3/8 triatlon betreft? Of kun je volstaan met een *verkorte triatlon*? Ik weet niet in hoeverre jullie 3/8 in deze opgave als afleidende cijfers zien...
- leerlingen weten mogelijk niet wat een triatlon is. Daar zou je een zinnetje over kunnen toevoegen. De informatie over Wijdeveen kan juist achterwege blijven. Bijvoorbeeld:
*Stephen doet mee aan een verkorte triatlon. Daarbij moet hij een stuk zwemmen, een stuk fietsen en een stuk hardlopen.
Voor deze verkorte triatlon moet Stephen 1460 m zwemmen. De regels voor zo'n verkorte triatlon schrijven voor dat je 1425 m moet zwemmen, maar dat dit ook 5% meer of 5% minder mag zijn.*

Koppel 5

Het taalgebruik in beide items is niet moeilijk. Vermoedelijk is item 10 iets complexer dan item 42.

In **item 10** had de vraag zelf echt als vraag geformuleerd kunnen worden, waarmee ook de verdubbeling van berekenen/berekening verdwijnt: *Hoeveel m2 plastic heeft Brenda minstens nodig? Typ je berekening in.*

Detailopmerkingen over item 10:

- Het plaatje van de blokhut voegt niets toe.
- De volgorde in de eerste zin is onlogisch. Waarom niet: *Brenda wil plastic onder de tegelvloer leggen.* Dat sluit de interpretatie uit dat Brenda onder de tegelvloer zit.
- De toevoeging *onder de tegelvloer* in de zin erna is overbodig: *Dit plastic moet aan alle kanten 5 cm uitsteken.*