

SAMENVATTING

De testen uit Muiswerk Testsuite 8 Rekenen 1F-2F-3F zijn beoordeeld met de anker testen van het Ministerie van Onderwijs, Cultuur en Wetenschap. De norm is bepaald op 65%. Dat is de norm waarop kandidaten het betreffende niveau gehaald hebben. Tevens is de betrouwbaarheid bepaald. De 1F test heeft een betrouwbaarheid van 0,81 en de 2F test heeft een betrouwbaarheid van 0,92. Daarmee is aangetoond dat de testen van Muiswerk betrouwbaar en valide zijn.

HET ONDERZOEK

De Commissie voor Toetsen en Examens van het Ministerie van Onderwijs, Cultuur en Wetenschap heeft eind 2013 sets referentietesten Rekenen ter beschikking gesteld voor uitgevers van testmateriaal. Deze referentietests, ook wel anker testen genoemd, zijn bedoeld om bij elk test de norm (cesuur) te bepalen waarop een kandidaat het niveau niet of wel gehaald heeft.

In de periode maart tot juni 2014 zijn de Ankeronderzoeken t.b.v. Muiswerk Testsuite 8 Rekenen 1F-2F-3F uitgevoerd.

De Muiswerk testen bieden voor elk niveau 40 vragen aan de kandidaten aan. Uit de set van 80 Anker vragen per niveau zijn 20 vragen geselecteerd, 5 per domein. Deze vragen zijn in de Muiswerktesten letterlijk overgenomen en in willekeurige volgorde met de andere vragen verwerkt. Er zijn 450 kandidaten getest op het 1F niveau en 550 kandidaten op het 2F niveau.

Alle scores op items zijn gecontroleerd en daaruit zijn geen items naar voren gekomen die te moeilijk of te gemakkelijk waren.

Het eerste doel van dit onderzoek is de norm (cesuur) bepalen bij de Muiswerk testen.

Omdat het hier gaat om het bepalen van een norm en er geen factoren zijn die het testresultaat beïnvloeden, anders dan een gebrek aan kennis en/of vaardigheid van de kandidaat, kan volstaan worden met het onderzoek in hoeverre beide testen als 'parallel' beschouwd kunnen worden.

In Testtheorie van prof. Drenth wordt bewezen dat 2 testen parallel zijn indien hun gemiddelden gelijk zijn, hun varianties gelijk zijn en de correlatie van beide testen met een onafhankelijke parameter gelijk is. Dat hebben we kunnen bewijzen. Omdat de betrouwbaarheid van de anker testen hoog is, is de correlatie tussen anker test en de Muiswerk test voldoende om de betrouwbaarheid van de Muiswerk test te bepalen en hoeven we geen schattingen te doen voor bv een Cronbach's α .

Bij de anker testen zijn gegevens en hulpmiddelen aangeleverd om de cesuur te bepalen. Voor de 1F anker test is de cesuur vastgesteld op 64%. Uit de testresultaten blijkt dat een score van 64% op de anker testen overeenkomt met een score van 65% op de Muiswerk test met een standaard deviatie van 8,7%. Dat betekent dat de cesuur van de Muiswerk 1F test op 65% ligt. De correlatie tussen de beide 1F testen is 0,81. Daarmee is de Muiswerk Testsuite 8 Rekentest 1F ruim voldoende betrouwbaar om ingezet te worden voor het bepalen van het niet/wel halen van het 1F reken niveau.

Bij de itemselectie voor de 2F test blijkt de cesuur bij 47,5% te liggen. Uit de testresultaten blijkt dat een score van 47,5% op de anker test overeenkomt met een score van 65% op de Muiswerk test met een standaard deviatie van 11,1%. Dus ook bij de Muiswerk 2F test ligt de cesuur op 65%. De correlatie tussen de anker test 2F en de Muiswerk test 2F is 0,92. Daarmee is de Muiswerk Testsuite 8

Rekentest 2F zeer ruim betrouw om ingezet te worden voor het bepalen van het niet/wel halen van het 2F reken niveau. Door de hoge betrouwbaarheid en het feit dat de ankertesten valide zijn, is tevens aangetoond dat ook de testen van Muiswerk Testsuite 8 Rekenen 1F-2F-3F valide zijn.

Detail uitwerking

Volgens prof. dr. P.J.D. Drenth en prof. dr. K Sijsma in Testtheorie (2006) zijn twee testen parallel als aan 3 voorwaarden voldaan is: 1. het gemiddelde is gelijk, 2. de varianties zijn gelijk en 3. de correlaties met een willekeurige onafhankelijke variabele is gelijk, zie bladzijde 209 van boven genoemd boek.

Hierbij valt op te merken dat Drenth en Sijsma de derde eis het zwaarste vinden wegen. Wij merken verder op dat het gemiddelde niet gelijk hoeft te zijn, omdat het er juist om ging de cesuur te bepalen. Het zijn dus niet zuiver parallelle test, maar in principe verschoven parallelle tests.

Hieronder noemen we de Muiswerk test de F-test en de Ankertest de A-test.

Voor de 1F test hebben wij de volgende waarden gevonden.

Het gemiddelde van de F-test is 65,29933 en van de A-test 66,66297.

De variantie van F is 251,8 en van A 482,6.

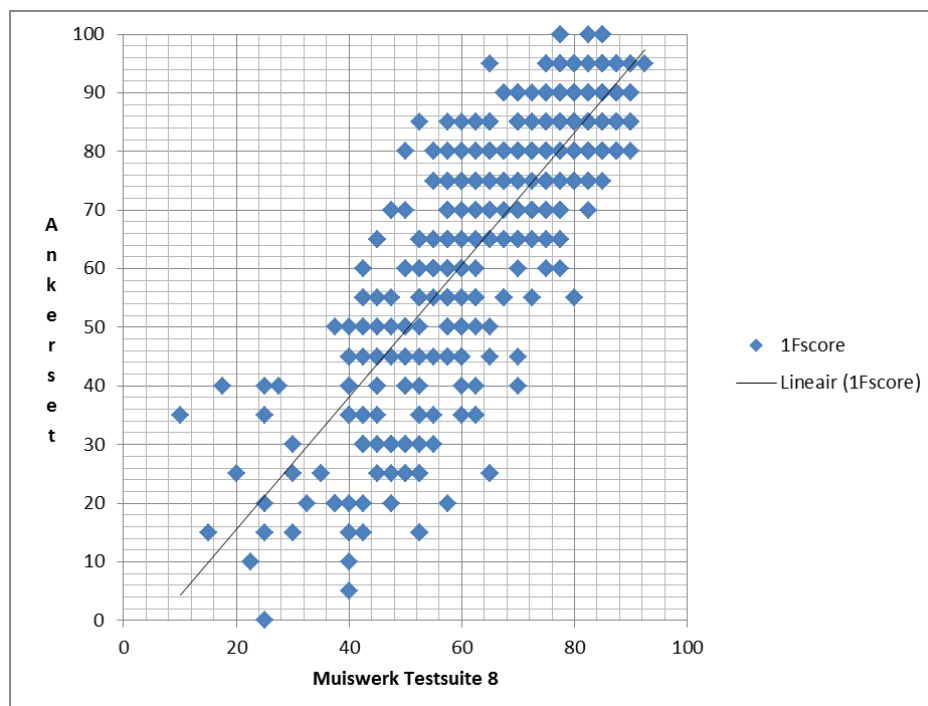
Als onafhankelijke variabele L hebben wij het leerling nummer genomen.

De correlatie tussen F en L is $-2,9965 \cdot 10^{-5}$ en de correlatie tussen A en L is $-3,1567 \cdot 10^{-5}$.

Alleen de variantie wijkt sterk af. Dat is te verklaren uit het geringe aantal vragen van de A-test.

Vervolgens hebben wij de F-scores genomen van de leerlingen bij de cesuur van de A-test, dus een A-score van 65%. Daaruit volgt een gemiddelde F-score van 65% met een standaard deviatie van 8,705.

Hieronder ziet u het scatter diagram van de twee testen met de regressielijn.



Voor de 2F test hebben wij de volgende waarden gevonden.

Het gemiddelde van de F-test is 38,85000 en van de A-test 30,90000.

De variantie van F is 686,90 en van A 577,42.

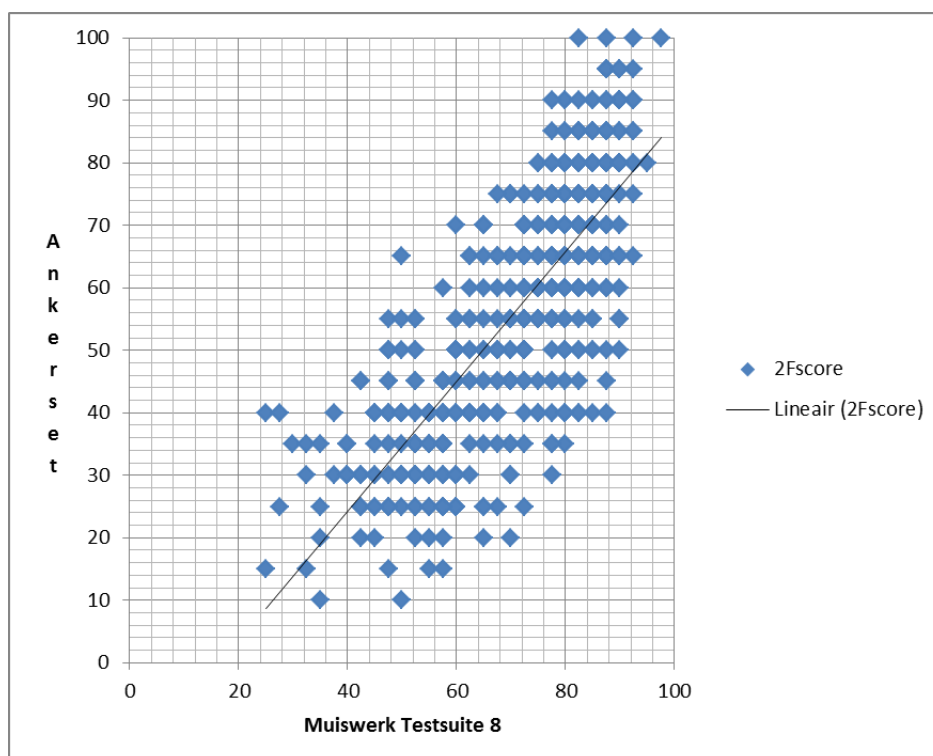
Als onafhankelijke variabele L hebben wij weer het leerling nummer genomen.

De correlatie tussen F en L is $-2,8623 \cdot 10^{-5}$ en de correlatie tussen A en L is $-2,4831 \cdot 10^{-5}$.

Ook hier wijkt de variantie af.

Vervolgens hebben wij de F-scores genomen van de leerlingen bij de cesuur van de A-test, dus een A-score van 47,5%. Daaruit volgt een gemiddelde F-score van 65% met een standaard deviatie van 11,098.

Hieronder ziet u het scatter diagram van de twee testen met de regressielijn.



Formules

Gebruikte formules:

$$\text{Variantie: } S^2(F) = \frac{1}{n} \sum_{i=1}^n (F_i - \bar{F})^2$$

$$\text{Covariantie: } V(F, L) = \frac{1}{n} \sum_{i=1}^n (F_i - \bar{F})(L_i - \bar{L})$$

$$\text{Correlatie: } r(F, L) = \frac{\frac{1}{n} \sum_{i=1}^n (F_i - \bar{F})(L_i - \bar{L})}{S(F)S(L)}$$