

# Toets- en itemanalyses; naar een breder gebruik

■ Inge Lauret en Harry Molkenboer

Drs. Inge Lauret en Ing. Harry Molkenboer zijn studenten van de masteropleiding Toetsdeskundige en hebben in het kader van de opleiding onderzoek gedaan naar soorten programma's voor toets- en itemanalyse, beschikbare programma's in Nederland en hun functionaliteiten

Email: i.lauret@student.fontys.nl, h.molkenboer@student.fontys.nl

Een toets- en itemanalyse geeft handvatten om naar de kwaliteit van toetstaken en een toets te kijken en verbeteringen door te voeren. Voor het uitvoeren van een analyse zijn analyseprogramma's beschikbaar. Waarom zijn dergelijke programma's nuttig, welke soorten programma's bestaan er, welke functionaliteiten bieden ze, waardoor is het gebruik beperkt en wat is daaraan te doen? Op deze vragen wordt in dit artikel een antwoord gegeven. De gegevens zijn verkregen uit een onderzoek dat de auteurs als studenten van de masteropleiding Toetsdeskundige hebben uitgevoerd naar programma's voor toets- en itemanalyses

## Inleiding

Binnen het vakgebied van toetsing wordt voor toets- en itemanalyse vaak de afkorting TIA gebruikt. Voor een TIA worden de antwoorden en de scores van kandidaten op de toetstaken (vragen en opdrachten) gebruikt die na afname van een toets beschikbaar zijn. Deze worden omgezet in informatie over de moeilijkheid van toetstaken en de toets als geheel, het onderscheidend vermogen van een toetstaak, de aantrekkelijkheid van afleiders bij gebruik van gesloten vragen, de betrouwbaarheid en de standaardmeetfout van de toets. Dit zijn de zogenoemde psychometrische gegevens.

Voor het verkrijgen van de psychometrische gegevens zijn verschillende soorten speciaal daarvoor ontwikkelde programma's beschikbaar. Er zijn programma's waarmee direct na een toetsafname geheel geautomatiseerd de psychometrische gegevens beschikbaar komen. Bij andere programma's moeten enkele hande-

lingen verricht worden om de toetsresultaten in te lezen, waarna de gegevens geautomatiseerd beschikbaar komen. Sommige programma's zijn gratis beschikbaar. Sommige programma's zijn onderdeel van een groter systeem zoals een toetsservicesysteem of een learning managementsysteem (LMS).

## Waarom zijn TIA-programma's nuttig?

Goosen (2015) schreef in *Levende Talen Magazine* 2015 | 2 een artikel met als titel "GEBRUIK DE TIA'S". Het artikel was een dringende oproep aan docenten taal in het voortgezet onderwijs (vo) om de analysegegevens uit TIA's te gebruiken. Zichtbaar wordt hoe een klas scoort ten opzichte van het landelijke gemiddelde. Ook kan zichtbaar worden gemaakt welke vragen goed zijn gemaakt en met welke vragen leerlingen moeite hadden. Deze informatie kan in het nieuwe schooljaar worden gebruikt bij individuele begeleiding en de voorbereiding van de leerlingen op komende tentamens en examens.

Om iets te kunnen zeggen over de kwaliteit van een toets is het volgens Eggen (2009) in zijn oratie over de kwaliteit van toetsen “absoluut noodzakelijk” (p. 14) om inzicht te hebben in de psychometrische aspecten van een toets. Als geen TIA wordt gemaakt, wordt niet goed zichtbaar waar kandidaten moeite mee hebben gehad, welke toetstaken kwalitatief voldeden en welke aanpassingen eventueel gewenst zijn. Aanpassingen kunnen een correctie op de cesuur tot gevolg hebben, omdat de kandidaten niet de dupe mogen worden van slechte vragen. Een analyse dient doordacht en met kennis van zaken te gebeuren. Een moeilijke vraag (met een lage p-waarde) kan duiden op een onderwerp dat moeilijk is, niet of slecht is behandeld tijdens de les, door de kandidaten niet is bestudeerd of door alle kandidaten als moeilijk wordt ervaren. Ook kan uit de analyse blijken dat de hoge moeilijkheid wordt veroorzaakt door een sleutelfout (een afleider die door een vergissing de sleutel is geworden), of door een slechte formulering van de stam of de alternatieven. Daarnaast kan uit de analyse blijken dat de vraag een goed onderscheid maakt tussen de betere en de zwakkere kandidaten, dus dat de betere kandidaten de vraag juist wel goed hebben gemaakt en zwakkere kandidaten niet. Het omgekeerde kan ook voorkomen; de betere kandidaten beantwoorden de vraag fout en de zwakkere kandidaten beantwoorden de vraag juist goed. Dan is er mogelijk wat aan de hand met de vraag. Zonder een analyse wordt niet of onvoldoende zichtbaar dat er een probleem met de toetstaak kan zijn en wat het probleem zou kunnen zijn. Kandidaten kunnen daarvan dus onopgemerkt de dupe worden. De kans bestaat dat kandidaten onterecht slagen of zakken. Dit betreft met name de kandidaten die zich rondom de cesuurgrens bevinden, maar dat zijn nu juist de kandidaten waarvoor een beslissing van gezakt of geslaagd belangrijk is. Als snel wordt gezegd dat een toets ‘te moeilijk was’ en de resultaten dan ook vaak ‘te laag’ zijn, dan valt daar zonder een gedegen TIA-analyse eigenlijk nog niets over te zeggen. Vanuit het kwaliteitsdenken hebben toetsaan-



## Inzicht hebben in de psychometrische aspecten van een toets is absoluut noodzakelijk om iets te zeggen over de kwaliteit van de toets

bieders de morele plicht om de kwaliteit van de toetsen zo goed mogelijk te borgen.

TIA's zijn nuttig ter verbetering van het onderwijs, of om een onderbouwde verantwoording af te leggen over de kwaliteit van de toetstaken en de toets. Met een TIA wordt ook voorkomen dat toetstechnisch slechte toetstaken opnieuw terecht komen in een volgende toets. TIA's bieden dus ook handvatten om de kwaliteit van toetstaken en van toetsen structureel te verbeteren.

### Welke soorten TIA-programma's bestaan er?

Als nut en noodzaak ertoe leiden dat psychometrische gegevens gebruikt gaan worden,

moeten ze wel eerst beschikbaar gemaakt worden. De makkelijkste weg is om bij de ontwikkeling, afname en beoordeling van toetsen gebruik te maken van een toetsservicesysteem. Een toetsservicesysteem is een al dan niet online beschikbaar systeem waarmee gesloten en soms ook open vragen en opdrachten ontwikkeld kunnen worden. De afname van de gesloten en open vragen kan digitaal plaatsvinden. Bij gesloten vragen worden de antwoorden van de kandidaten digitaal geregistreerd. Als dat systeem over de TIA-functionaliteiten beschikt, zijn psychometrische gegevens zonder tussenkomst van een persoon beschikbaar. Bij open vragen moet eerst nog een beoordeling van de gegeven antwoorden plaatsvinden. Afhankelijk van de functionaliteiten van een toetsservicesysteem kan dit met verschillende en tweede beoordelaars digitaal plaatsvinden. Zonder tussenkomst van een persoon kunnen, als dat systeem daarin voorziet, psychometrische gegevens na de beoordeling van de open vragen beschikbaar komen. Afhankelijk van hoe opdrachten beoordeeld worden, kunnen op vergelijkbare wijze als bij open vragen, beoordelingen van opdrachten opgenomen worden in het toetsservicesysteem. Vergelijkbaar met toetsservicesystemen beschikken LMS'en soms ook over functionaliteiten voor TIA's. Toetsservicesystemen en LMS'en kunnen de analysegegevens overzichtelijk weergeven, eventueel in fraaie rapportvormen met toelichtingen op de analyseresultaten. Soms zijn deze rapporten voorzien met (automatisch gegenereerde) adviezen op basis van de analysegegevens. Het verdient aanbeveling om doordacht met die automatische adviezen om te gaan en ook nog zelf kritisch de analysegegevens te bestuderen.

Indien er nog geen digitale afnamemogelijkheden zijn, wordt bij gesloten toetsen gebruik gemaakt van zogenaamde schrapkaarten of A4-antwoordformulieren. Deze kaarten of formulieren worden via speciale scanapparatuur gescand waarna de antwoorden door een TIA-programma worden ingelezen en de psychometrische gegevens beschikbaar komen.

Een dergelijk programma kan een toetsservice-systeem zijn, maar ook een speciaal ontwikkeld analyseprogramma, of een programma ontwikkeld op basis van MS-Excel, SPSS- of R. Organisaties die beperkt gebruik maken van digitaal toetsen, kunnen de belangrijkste psychometrische gegevens genereren met Excel of via R. Invoer van toetsresultaten gebeurt dan veelal 'met de hand'. Dergelijke systemen worden vooral gebruikt voor toetsen met een beperkt aantal kandidaten (enkele tientallen). Dergelijke programma's zijn veelal gratis en vrij beschikbaar, maar soms zijn ze alleen bedoeld voor intern gebruik.

### Onderzoek naar functionaliteiten van TIA-programma's

Door de analysemogelijkheden van verschillende TIA-programma's te onderzoeken, rapporten van TIA's te bestuderen en literatuur over toets- en itemanalyses te bestuderen zijn door de auteurs functionaliteiten die relevant zijn binnen TIA's verzameld en geordend. Dit onderzoek is beperkt tot psychometrische analyses op basis van de klassieke testtheorie en is geen aandacht besteed aan de itemresponstheorie. Het onderzoek heeft geleid tot bijna driehonderd te onderscheiden gegevens/functionaliteiten van TIA-programma's. Deze gegevens/functionaliteiten zijn geordend op basis van vijf hoofdgroepen en 87 subgroepen (zie een MS Excel-overzicht dat toegankelijk is via de Toetskaart van Nederland: <https://www.e-xamens.nl/toetskaart-van-nederland>). De hoofdgroepen betreffen: algemene gegevens TIA-programma, toetsgegevens, toetsanalyse, itemanalyse, kandidaatgegevens, rapportage.

Aan 47 mogelijke leveranciers van TIA-programma's of TIA-diensten in Nederland is gevraagd welke functionaliteiten hun programma biedt aan de hand van de lijst met bijna driehonderd functionaliteiten. Van de leveranciers werd gevraagd de functionaliteitenlijst in te vullen voor hun TIA-programma. Daarnaast werd gevraagd om van twee toetsen de antwoorden van kandidaten in te voeren. Een van een toets

met de scores en de antwoorden op gesloten vragen en een van een toets met de scores op open vragen. Hiermee kon nagegaan worden of het TIA-programma correcte analysegegevens opleverde. De toets met open vragen was gebaseerd op twee groepen kandidaten en kende ook twee te onderscheiden onderwerpen waarop getoetst werd. Programma's/leveranciers konden hierdoor ook laten zien of meer gedetailleerde analyses gerelateerd aan de groepen of de onderwerpen mogelijk waren.

Zeven reacties zijn ontvangen: Detmin van Tanis education, TiaPlus en Dexter (R-programma) beide van Cito, een MS Excelprogramma van Fontys LerarenOpleiding Tilburg (FLOT), een Teleformkoppeling van Simac, TestVision Online van Teelen Kennismanagement en de ToetsAnalyseTool (R-programma) van 10voordeleraar. De respons viel de auteurs tegen. Mogelijke redenen hiervoor zijn de onzekerheid over het bestaan van de leverancier in Nederland (niet van alle aangeschreven leveranciers was vooraf duidelijk of zij nog bestonden en wat zij precies konden leveren), het niet geschikt zijn van het programma voor TIA's zoals bedoeld, of gebrek aan tijd van de leverancier om de uitgebreide functionaliteitenlijst en een analysevoorbeeld aan te leveren.

Welke functionaliteiten de zeven programma's bieden is weergegeven in het hiervoor genoemde MS Excel-bestand. De gegevens komen van de leveranciers en zijn niet gecontroleerd. Alle TIA-programma's bieden de basisfunctionaliteiten om te komen tot een goede TIA. De basisfunctionaliteiten betreffen de in de inleiding al genoemde psychometrische gegevens zoals de moeilijkheid van vragen en de toets als geheel, het onderscheidend vermogen van een vraag (maar ook van de afleiders bij gesloten vragen), de aantrekkelijkheid van afleiders bij gebruik van gesloten vragen, de betrouwbaarheid en de standaardmeetfout van de toets.

De ToetsAnalyseTool van 10voordeleraar is ontwikkeld in R en wordt doorontwikkeld op basis van de lijst met mogelijke functionaliteiten.

## Beperkt gebruik van TIA's

Uit verschillende onderzoeken (Draaijer, Hartog & Hofstee, 2007; Draaijer, 2016; Wan & Henly, 2012;) blijkt dat veel docenten in het hoger onderwijs een digitaal toetssysteem niet gebruiken of slechts gebruiken voor het ontwikkelen van multiple-choice vragen. Draaijer et al (2007) wijten dit onder meer aan het ontbreken van voldoende toetsbekwaamheid en tijd bij docenten. Voldoende toetsbekwaamheid en tijd zijn ook nodig om TIA's te gebruiken. Dat het hier bij docenten aan ontbreekt wordt bevestigd uit ervaringen met BKE-trainingen (BasisKwalificatie Examinering). Zelfs al zijn TIA's beschikbaar dan nog worden ze niet of heel oppervlakkig gebruikt. Er wordt dan vooral naar de resultaten van de toets gekeken en soms naar de moeilijkheid van de vragen en de toets. Van de andere psychometrische gegevens wordt aangegeven dat ze lastig te begrijpen zijn of dat de statistiek moeilijk is. Uit verschillende bronnen (Macher,



## Een breder gebruik van TIA's leidt tot een hogere toetskwaliteit

Paechter, Papousek & Ruggeri, 2012; Paechter, Macher, Martskvishvili, Wimmer & Papousek, 2017; Onwuegbuzie & Wilson, 2003) blijkt dat bij veel studenten in het hoger onderwijs sprake is van 'statistics anxiety'. Er zijn geen onderzoeken gevonden waaruit blijkt dat dit ook geldt voor docenten in bijvoorbeeld het hoger onderwijs, maar geheel onwaarschijnlijk is dit niet. Motivatie is een belangrijke reden om TIA's te gebruiken. Dan moeten gebruikers wel overtuigd zijn van nut en noodzaak en zich bekwaam voelen om TIA's te gebruiken. Enig niveau van toetsbekwaamheid is juist dan nodig. Uit BKE-trainingen blijkt dat als docenten wordt geleerd wat de basis psychometrische gegevens inhouden ze enthousiast worden om ze te gebruiken. Als blijkt dat ze die gegevens dan ook nog gaan toepassen, dan is veel gewonnen. Bij trainingen voor examenontwikkelaars in de private sector worden dezelfde ervaringen opgedaan. Natuurlijk mogen deze ervaring niet zomaar worden geëxtrapoleerd. Maar training lijkt hier een goed instrument te zijn om docenten en ontwikkelaars enthousiast te krijgen voor het gebruik van TIA-analyses.

### Naar een breder gebruik van TIA's

Met het onderzoek naar soorten TIA-programma's en hun functionaliteiten en de online publicatie van de functionaliteitenlijst met de gegevens van een aantal programma's, hebben de auteurs een bijdrage willen leveren aan een breder gebruik van TIA's omdat nut en noodzaak alles te maken hebben met het leveren van toetskwaliteit. Kandidaten mogen verwachten dat hen op een kwalitatieve wijze 'de maat' wordt genomen. Het gebruik van TIA's om meer verantwoorde en onderbouwde beslissingen over slagen en zakken te kunnen nemen, hoort daarbij. De leveranciers van de programma's die hebben meegewerkt worden langs deze weg nogmaals bedankt voor hun bijdrage hieraan.

### Literatuur

- Draaijer, S. (2016). Supporting Teachers in Higher Education in Designing Test Items. Academisch proefschrift. Amsterdam: Vrije Universiteit Amsterdam.
- Draaijer, S., & Hartog, R. J. M. (2007). Design patterns for digital item types in higher education. *E-Journal of Instructional Science and Technology (e-JIST)*, 10.
- Eggen, T. J. H. M. (2009). De kwaliteit van toetsen. Rede uitgesproken bij het aanvaarden van het ambt van hoogleraar Psychometrische Aspecten van Examinering aan de Faculteit Gedragwetenschappen van de Universiteit Twente.
- Goosen, H. (2015). Gebruik de Tia's! De Cito-toetsanalyse van de landelijke centrale examens efficiënter benut in de lespraktijk. *Levende Talen Magazine*, 102(2), 18-22.
- Macher, D., Paechter, M., Papousek, I., & Ruggeri, K. (2012). Statistics anxiety, trait anxiety, learning behavior, and academic performance. *European journal of psychology of education*, 27(4), 483-498.
- Onwuegbuzie, A. J., & Wilson, V. A. (2003). Statistics Anxiety: Nature, etiology, antecedents, effects, and treatments--a comprehensive review of the literature. *Teaching in Higher Education*, 8(2), 195-209.
- Paechter, M., Macher, D., Martskvishvili, K., Wimmer, S., & Papousek, I. (2017). Mathematics Anxiety and Statistics Anxiety. Shared but Also Unshared Components and Antagonistic Contributions to Performance in Statistics. *Frontiers in psychology*, 8, 1196.
- Wan, L., & Henly, G. A. (2012). Measurement properties of two innovative item formats in a computer-based test. *Applied Measurement in Education*, 25(1), 58-78.