rijksuniversiteit groningen

De legitimatie van de invoering van de CITO-toets in Nederland in de jaren zeventig in publieke discussies

Jildou de Vries

S5648025

Bachelorwerkstuk Pedagogische Wetenschappen

Begeleiders en eerste beoordelaar: Dr. J. Exalto & R. Kuijpers

Tweede beoordelaar: Dr. D.D.N.M. Kostons

Rijksuniversiteit Groningen

Faculteit der Gedrags- en Maatschappijwetenschappen

Pedagogische Wetenschappen

Mei 2026

Aantal woorden: 5999

Abstract

This study examines how the introduction of the CITO test in the Netherlands during the 1970s was legitimised in public debates. Newspaper articles from 1970–1979, collected through Delpher, were analysed. Using thematic analysis, the study explored which arguments were used to defend or criticise the test. The central research question was: 'How was the introduction of the CITO test in the Netherlands during the 1970s legitimised in public debates?'

The analysis shows that the test was primarily legitimised through arguments related to objectivity, followed by selection. The CITO test was presented as an instrument contributing to a more systematic assessment of pupils. At the same time, it was emphasised that the test should not be seen as an absolute form of assessment, but as a supplement to teachers' professional judgement, measuring only part of pupils' abilities.

In addition, the test was both defended and criticised as a selection tool. It was seen as a means to reduce differences between schools, but criticised for its influence on educational trajectories and content. The theme of comparability emerged because the CITO test made pupils' and schools' performance more comparable. At the same time, concerns arose regarding relative positions and competition. Concerns about the effects on pupils were less prominent and mainly related to tension surrounding the testing moment.

In summary, this study shows that the introduction of the CITO test in the public debate was not legitimised in a straightforward manner, but was characterised by tensions between different arguments and conditions.

Samenvatting

In dit onderzoek is onderzocht hoe de invoering van de CITO-toets in Nederland in de jaren zeventig werd gelegitimeerd in publieke discussies. Hiervoor zijn krantenartikelen uit de periode 1970–1979 geanalyseerd die via Delpher zijn verzameld. Met behulp van thematische analyse is onderzocht welke argumenten in het publieke debat werden gebruikt ter verdediging of bekritisering van de toets. Centraal stond de vraag: ‘Hoe werd de invoering van de CITO-toets in Nederland in de jaren zeventig gelegitimeerd in publieke discussies?’

Uit de analyse blijkt dat de legitimering van de toets vooral plaatsvond via argumenten rond objectiviteit, gevolgd door selectie. De CITO-toets werd gepresenteerd als een instrument dat kon bijdragen aan een meer systematische beoordeling van leerlingen. Tegelijkertijd werd benadrukt dat de toets niet als absoluut beoordelingsinstrument moest worden gezien, maar als aanvulling op het oordeel van leerkrachten en dat zij slechts een deel van de capaciteiten van leerlingen kon meten.

Daarnaast werd de toets als selectiemiddel enerzijds verdedigd om verschillen tussen scholen te verkleinen, maar anderzijds bekritiseerd vanwege de mogelijke invloed op schoolloopbanen en onderwijsinhoud. Het thema vergelijkbaarheid kwam naar voren doordat de CITO-toets prestaties van leerlingen en scholen beter vergelijkbaar maakte. Tegelijkertijd ontstonden zorgen over relatieve posities en concurrentie. Effecten op leerlingen kwamen minder nadrukkelijk naar voren in het debat en hadden vooral betrekking op spanning rond het toetsmoment.

Samenvattend laat dit onderzoek zien dat de invoering van de CITO-toets in het publieke debat niet eenduidig werd gelegitimeerd, maar werd gekenmerkt door een spanningsveld tussen verschillende argumenten en voorwaarden.

Inleiding

De doorstroomtoets in het Nederlandse basisonderwijs heeft in recente jaren opnieuw geleid tot maatschappelijke discussie. Zo ontstond volgens een bericht van de NOS (2025) debat toen sommige basisscholen aangaven de toets te willen weigeren. Deze scholen stelden dat de toets leerlingen vooral met elkaar vergelijkt en onvoldoende inzicht geeft in de individuele ontwikkeling. Daartegenover wordt in het artikel beschreven dat het idee achter de toets is dat deze een objectieve aanvulling vormt op het schooladvies en moet bijdragen aan het terugdringen van kansenongelijkheid. De toets is wettelijk verplicht, waardoor de discussie niet alleen onder scholen speelde, maar ook de Onderwijsinspectie, de politiek en ouders raakte (NOS, 2025). Dit laat zien dat toetsing onderwerp is van debat op meerdere niveaus, waaronder de maatschappij, de politiek en het onderwijs.

Ook binnen de wetenschap bestaat discussie over selectie binnen het Nederlandse onderwijssysteem (Terwel, 2006) en over vragen rond de validiteit van toetsing en de mogelijke effecten daarvan op gelijke kansen in het onderwijs (Scheider et al., 2023). Tegelijkertijd zijn deze discussies niet nieuw: al gedurende de twintigste eeuw, tijdens de opkomst van gestandaardiseerde toetsing, werd gediscussieerd over de rol van dergelijke toetsen in het onderwijs (Linn, 2001).

De oorsprong van gestandaardiseerde toetsing lag al aan het einde van de negentiende eeuw. In deze periode werden kinderen steeds vaker beschouwd als meetbare en classificeerbare objecten van studie, waarop een wetenschappelijke benadering van onderwijs kon worden gebaseerd. Deze ontwikkeling hing samen met een groeiend vertrouwen in wetenschappelijke methoden en met de uitbreiding van nationale onderwijssystemen. Classificatie van leerlingen werd daarbij gezien als een manier om het onderwijs efficiënter te organiseren (Canales & Polenghi, 2019).

In de twintigste eeuw kregen deze ideeën verdere uitwerking in de Verenigde Staten, waar gestandaardiseerde toetsing uitgroeide tot een belangrijk, maar omstreden onderdeel van het onderwijssysteem. Door de sterke groei van het aantal leerlingen ontstond behoefte aan instrumenten om leerprestaties op een meer efficiënte en ogenschijnlijk objectieve manier te beoordelen. Tegelijkertijd riep deze ontwikkeling kritiek op, onder meer over de meetbaarheid van onderwijsresultaten en de vraag of examens niet een doel op zichzelf werden (Linn, 2001). Ondanks dergelijke kritiek raakten gestandaardiseerde toetsen institutioneel verankerd en werden zij ondersteund door politieke en organisatorische structuren (Hutt & Schneider, 2018).

Ook in Nederland ontwikkelde zich in de twintigste eeuw een professionele testpraktijk, waarin betrouwbaarheid, validiteit en normering steeds belangrijker werden (Van Drunen & Van Strien, 1991). De wetenschappelijke benadering van testen en examineren kreeg in Europa vooral na de Tweede Wereldoorlog merkbare invloed. De kwantitatief-empirische benadering, waarbij meetmethoden werden overgenomen, verspreidde zich vanuit de Verenigde Staten naar West-Europa en vond ook in Nederland ingang. Deze ontwikkeling moest zich verhouden tot een onderwijsstelsel dat werd gekenmerkt door constitutionele onderwijsvrijheid en een politiek pluriforme samenleving (Thio, 1994).

In de jaren zestig begon de potentiële waarde van ‘educational measurement’ in Nederland bredere erkenning te krijgen, niet alleen onder onderwijsonderzoekers en beleidsmakers, maar ook daarbuiten (Thio, 1994). In dezelfde periode bekritiseerde psycholoog A.D. de Groot in zijn boek *Vijven en zessen* de subjectiviteit van leerkrachten bij de doorverwijzing van basisonderwijs naar het voortgezet onderwijs. Hij pleitte voor meer objectieve vormen van beoordeling in het onderwijs (De Groot, 1978). Deze groeiende erkenning leidde in 1968 tot de oprichting van het Centraal Instituut voor Toetsontwikkeling (CITO), dat als onafhankelijke stichting met overheidssteun bestaande initiatieven op het

gebied van toetsontwikkeling overnam (Thio, 1994). CITO groeide in korte tijd uit tot een belangrijke organisatie voor objectieve toetsing en evaluatie in het Nederlandse onderwijs (Broekman, 1981).

Hoewel discussies over toetsing al vanaf de invoering van gestandaardiseerde toetsing aanwezig waren, ontstond in Nederland in het midden van de jaren zeventig meer expliciete weerstand tegen objectieve testvormen. In de vroege jaren zeventig kreeg CITO een steeds meer erkende rol in de ontwikkeling en analyse van deze toetsen en stimuleerde CITO het gebruik ervan vanwege de nadruk op onder andere objectiviteit. Deze ontwikkeling ging gepaard met toenemende kritiek onder leerkrachten en beleidsmakers, onder meer op een mogelijke ongewenste terugslag op het onderwijs en op een verarming van de onderwijsdoelstellingen (Thio, 1994).

Eerder onderzoek heeft aandacht besteed aan de ontwikkeling van testpraktijken (Van Drunen & Van Strien, 1991) en de institutionele geschiedenis van organisaties zoals CITO (Thio, 1994). In bestaand onderzoek blijft echter onderbelicht hoe de invoering van de CITO-toets in het publieke debat werd gelegitimeerd. Vanuit deze spanning tussen institutionalisering en kritiek richt dit onderzoek zich op de volgende onderzoeksvraag: ‘Hoe werd de invoering van de CITO-toets in Nederland in de jaren zeventig gelegitimeerd in publieke discussies?’

De jaren zeventig zijn gekozen omdat deze periode de eerste jaren van de invoering en het gebruik van de CITO-toets omvat en daarmee inzicht biedt in vroege maatschappelijke reacties. In dit onderzoek verwijst legitimatie naar de wijze waarop in publieke discussies argumenten werden ingezet om de invoering van de CITO-toets te rechtvaardigen, te verdedigen of juist ter discussie te stellen. Deze publieke discussies worden onderzocht aan de hand van krantenartikelen uit de jaren zeventig.

Om deze onderzoeksvraag te beantwoorden zijn de volgende deelvragen geformuleerd:

Deelvraag 1: ‘Welke argumenten werden in het publieke debat geuit voor de invoering van de CITO-toets?’

Deelvraag 2: ‘Welke argumenten werden in het publieke debat geuit tegen de invoering van de CITO-toets?’

Door publieke discussies rond de invoering van deze toets te analyseren, beoogt dit onderzoek inzicht te bieden in de wijze waarop toetsing maatschappelijk werd gepresenteerd, verdedigd en bekritiseerd. Hierdoor wordt zichtbaar welke opvattingen over selectie en toetsing een rol speelden. Hiermee draagt dit onderzoek bij aan een beter historisch begrip van de maatschappelijke positionering van gestandaardiseerde toetsing in Nederland.

Om deze onderzoeksvraag te beantwoorden wordt een historische analyse uitgevoerd van krantenartikelen uit de periode 1970–1979 die via Delpher zijn verzameld. Door middel van thematische analyse worden argumenten voor en tegen de invoering van de CITO-toets inductief gecodeerd en samengebracht in overkoepelende thema’s. Hierdoor wordt inzicht verkregen in de wijze waarop de invoering van de CITO-toets in het publieke debat werd gelegitimeerd.

Theoretisch kader

Onderzoek naar de ontwikkeling van gestandaardiseerde toetsing laat zien dat deze deel uitmaakt van bredere internationale ontwikkelingen binnen het onderwijs. Vanaf het einde van de negentiende eeuw ontstond in de Verenigde Staten een toenemende nadruk op het meten en classificeren van leerlingen (Canales & Polenghi, 2019). Deze benadering verspreidde zich vervolgens naar West-Europa, waaronder Nederland, wat erop wijst dat de Nederlandse ontwikkeling van toetsing nauw verweven is met internationale ontwikkelingen (Thio, 1994).

Binnen deze internationale ontwikkelingen nemen de Verenigde Staten een centrale plaats in, niet alleen vanwege de vroege opkomst van ‘educational measurement’ (Thio, 1994), maar ook doordat de invoering van grootschalige toetsing daar gepaard ging met discussies en spanningen tussen voor- en tegenstanders (Hutt & Schneider, 2018; Linn, 2001; Rury et al., 2022). Dit debat wordt gekenmerkt door een blijvend spanningsveld (Haladyna et al., 1998).

Onderwijsonderzoekers Hutt en Schneider (2018) laten zien dat gestandaardiseerde toetsing in de Verenigde Staten al meer dan een eeuw onderwerp is van kritiek, onder andere vanwege negatieve effecten op leerlingen, scholen en het curriculum, terwijl het gebruik van deze toetsen tegelijkertijd is blijven toenemen. Dit verklaren zij onder meer vanuit het meritocratische karakter van het Amerikaanse onderwijssysteem en de rol van onderwijs in sociale mobiliteit. Binnen zo’n meritocratisch systeem kan gestandaardiseerde toetsing worden gezien als een eerlijke manier om leerlingen op basis van prestaties te selecteren en kansen te bieden. Andere studies wijzen erop dat toetsing historisch werd gedreven door factoren als objectiviteit, efficiëntie en verantwoording (Huddleston & Rockwell, 2015).

Waar Linn (2001) terugkerende discussies en verschuivingen rond gestandaardiseerde toetsing bespreekt, laten studies zoals Rury et al. (2022) zien hoe zulke discussies vanuit een

historisch perspectief kunnen worden gereconstrueerd. In het onderzoek van Rury et al. (2022) worden hiervoor onder meer rapporten, wetenschappelijke publicaties en krantenartikelen gebruikt. Uit hun studie blijkt dat gestandaardiseerde toetsing in de Verenigde Staten onder meer werd verdedigd met verwijzingen naar verantwoording, het voorkomen van diploma's zonder voldoende basisvaardigheden en een meer objectieve beoordeling van leerlingen (Rury et al., 2022).

Voor de Nederlandse context ligt de nadruk in bestaand onderzoek vooral op de ontwikkeling en institutionele inbedding van gestandaardiseerde toetsing. Studies richten zich onder andere op de rol van het CITO in de verdere ontwikkeling van gestandaardiseerde toetsing in Nederland (Thio, 1994) en op de projecten en werkwijze van het CITO bij de ontwikkeling en toepassing van toetsen (Broekman, 1981).

Samen laten deze studies zien dat onderzoek bestaat naar maatschappelijke discussies over toetsing, vooral in internationale contexten (Hutt & Schneider, 2018; Linn, 2001; Rury et al., 2022), en naar de opkomst van toetsing en de rol en werkwijze van instituties zoals het CITO (Broekman, 1981; Thio, 1994). In de Nederlandse context blijft een analyse van publieke debatten rondom de invoering van de CITO-toets echter onderbelicht.

Dit onderzoek richt zich daarom op de analyse van publieke discussies rond de invoering van de CITO-toets in Nederland in de jaren zeventig, met als doel inzicht te krijgen in de manier waarop deze toets werd gelegitimeerd. De internationale literatuur helpt om de bevindingen van dit onderzoek te plaatsen binnen bredere discussies over de legitimering van gestandaardiseerde toetsing.

Methode

Aan de hand van een systematisch historisch bronnenonderzoek is onderzocht hoe de invoering van de CITO-toets in Nederland in de jaren zeventig in het publieke debat werd gelegitimeerd. De primaire bronnen bestonden uit krantenartikelen uit de periode 1970 tot en met 1979, geraadpleegd via de digitale databank Delpher. Krantenartikelen waren geschikt om de onderzoeksvraag te beantwoorden, omdat zij inzicht gaven in hoe de invoering van de CITO-toets in het publieke debat werd besproken en gelegitimeerd.

In eerste instantie werden in Delpher de termen ‘CITO’ en ‘landelijke schooltoets’ gecombineerd en gefilterd op de periode 1970–1979. Deze zoekopdracht leverde 7 artikelen op. Vervolgens werd de zoekstrategie verbreed door andere zoektermen te gebruiken, namelijk ‘cito toets lager onderwijs’. Dit leverde 52 krantenartikelen op, waaruit vervolgens, samen met de eerder gevonden artikelen, relevante artikelen werden geselecteerd.

Hoewel de zoekperiode 1970–1979 omvatte, richtte de analyse zich voornamelijk op de jaren 1974–1977, waarin de invoering van de CITO-toets en het publieke debat daarover het meest zichtbaar waren. In de literatuur wordt namelijk gewezen op de groeiende rol van CITO in de loop van de vroege jaren zeventig en op toenemende weerstand tegen objectieve testvormen in het midden van de jaren zeventig (Thio, 1994).

Artikelen werden geselecteerd wanneer zij ingingen op de CITO-toets en argumenten bevatten waarmee de invoering of het gebruik van de toets werd gerechtvaardigd, verdedigd of ter discussie gesteld. De selectie werd voortgezet totdat thematische verzadiging werd bereikt. Dit betekende dat nieuwe artikelen geen wezenlijk nieuwe argumenten meer opleverden. Uiteindelijk zijn 11 artikelen geanalyseerd.

De geselecteerde krantenartikelen werden geanalyseerd met behulp van thematische codering. Argumenten voor en tegen de invoering van de CITO-toets werden eerst inductief gecodeerd, waarna de codes werden samengevoegd tot overkoepelende thema’s. Op deze

manier konden terugkerende patronen in het debat worden gereconstrueerd. De uiteindelijke thema's en subthema's zijn schematisch weergegeven in een codeboom (zie Bijlage A).

Bij het gebruik van historische krantenartikelen werd rekening gehouden met historische bronnenkritiek. Kranten vormen geen neutrale weergave van maatschappelijke opvattingen, maar representeren geselecteerde perspectieven die afhankelijk zijn van onder andere redactionele keuzes, doelgroep en ideologische achtergrond. In de jaren zeventig waren Nederlandse kranten daarnaast nog deels verzuild. Verzuiling verwijst naar de opdeling van de samenleving in verschillende levensbeschouwelijke en maatschappelijke groepen, met eigen organisaties en media. Deze context werd meegenomen in de interpretatie van de bevindingen. Bovendien laten historische bronnen slechts fragmenten van het verleden zien en geven daarmee geen directe weergave van gebeurtenissen (Schrag, 2021). De analyse richtte zich op argumenten in het publieke debat, niet op de feitelijke juistheid van uitspraken. Door deze systematische werkwijze en bronnenkritiek werd gestreefd naar een zo zorgvuldig mogelijke interpretatie van het publieke debat.

Daarnaast kon de digitalisering van kranten via Delpher beperkingen met zich meebrengen, doordat mogelijk niet alle kranten waren gedigitaliseerd en variatie in terminologie kon voorkomen. Door meerdere zoektermen te combineren en artikelen uit verschillende landelijke en regionale kranten te betrekken, werd gestreefd naar een zo systematisch en zorgvuldig mogelijke selectie.

Resultaten

In deze sectie worden de resultaten van het onderzoek gepresenteerd. Op basis van de analyse van 11 geselecteerde krantenartikelen uit de periode 1970–1979 zijn enkele terugkerende thema's geïdentificeerd: objectiviteit, selectie, invloed op onderwijs, effecten op leerlingen en vergelijkbaarheid. Deze thema's worden hieronder besproken en geven inzicht in de wijze waarop de invoering van de CITO-toets in het publieke debat werd gelegitimeerd.

De geanalyseerde artikelen zijn afkomstig uit verschillende Nederlandse kranten. De landelijke kranten omvatten *De Telegraaf* (1971), *NRC Handelsblad* (1972, 1973), *Trouw* (1974, 1975), *De Volkskrant* (1976), *Het Parool* (1977), *De Waarheid* (1977) en *Algemeen Dagblad* (1978). Daarnaast zijn enkele regionale kranten betrokken, waaronder *Tubantia* (1973) en *Limburgsch Dagblad* (1975). Deze kranten vertegenwoordigden in de onderzochte periode verschillende maatschappelijke achtergronden, doelgroepen en journalistieke stijlen. Dit kan van invloed zijn geweest op de wijze waarop argumenten werden gepresenteerd. Zo richtte *NRC Handelsblad* zich relatief sterk op een hoger opgeleid publiek en stond de krant bekend om een meer analytische stijl van berichtgeving. *De Volkskrant* had van oorsprong een katholieke achtergrond en ontwikkelde zich in deze periode in meer progressieve richting. *Trouw* kende een protestants-christelijke signatuur, terwijl *De Waarheid* nauw verbonden was aan de communistische stroming.

Naast verschillen in krantentypen viel op dat bepaalde actoren dominanter aanwezig waren in het debat. In de geanalyseerde artikelen werd het debat over de CITO-toets voornamelijk vormgegeven door journalisten, die daarbij vooral vertegenwoordigers van instituties, zoals CITO-directeur Solberg, onderwijsdeskundigen en schoolhoofden, aan het woord lieten. De perspectieven van bijvoorbeeld leerkrachten, ouders en leerlingen kwamen daarentegen minder direct naar voren. Dit kan van invloed zijn geweest op de wijze waarop de toets in het publieke debat werd gepresenteerd en gelegitimeerd.

Objectiviteit

In de geanalyseerde krantenartikelen kwam objectiviteit regelmatig naar voren als argument voor de invoering van de CITO-toets, waarbij deze werd gepresenteerd als een objectief instrument.

In *NRC Handelsblad* van 29 februari 1972 werd de toets gekoppeld aan het leveren van “objectieve gegevens” die konden bijdragen aan de beoordeling van leerlingen (Van Spaandonk, 1972, p. 9). Een jaar later, op 28 februari 1973, werd in dezelfde krant benadrukt dat de schooltoets werd ingevoerd om de tot dan toe subjectieve en willekeurige beoordeling van leerlingen te vervangen door een meer objectieve en constante beoordeling (Van Spaandonk, 1973).

Ook vanuit het CITO zelf werd dit beeld ondersteund. Solberg, directeur van het CITO, stelde in *NRC Handelsblad* van 28 februari 1973 dat de toets geen “gokspelletje” zou zijn (Van Spaandonk, 1973, p. 3). Daarmee werd aangegeven dat de toets gebaseerd was op systematische meting in plaats van willekeur. In een later artikel van het *Algemeen Dagblad* van 30 september 1978 stelde Solberg dan ook dat de toets had bijgedragen aan een meer transparante en objectieve beoordeling van leerlingen (d’Hamecourt, 1978).

In *De Telegraaf* van 20 februari 1971 werd bovendien gewezen op het feit dat de antwoorden door een computer werden verwerkt en daarmee op gelijke wijze werden nagekeken. Verder maakte de toets het mogelijk om verschillen tussen het oordeel van leerkrachten en toetsresultaten zichtbaar te maken en op termijn te corrigeren. Daarnaast werd de toets ontwikkeld door deskundigen en vooraf getest (Cohen, 1971). De toets werd daarmee gepresenteerd als een zorgvuldig ontworpen en eerlijk beoordelingsinstrument.

Dit beeld werd ook in andere artikelen bevestigd. Zo werd de toets in *De Waarheid* van 26 februari 1977 beschreven als een redelijk objectief hulpmiddel, waarbij werd benadrukt dat de resultaten doorgaans overeenkwamen met de verwachtingen van de

onderwijzer ("Schooltoets niet beslissend", 1977). Hiermee werd de toets gepresenteerd als objectief, maar ook als een instrument dat bestaande oordelen kon bevestigen en ondersteunen.

Objectiviteit werd niet als absoluut gepresenteerd. In meerdere artikelen werd aangegeven dat de toets aanvullend was op andere vormen van beoordeling (Cohen, 1971; "Schooltoets niet beslissend", 1977; “Uitslag kán bepalend zijn voor schoolkeus van zesdeklassers”, 1974; Van Spaandonk, 1973; “Wat in Amsterdam redelijk is, valt in Heerlen tegen”, 1975). Objectiviteit fungeerde daarmee als een belangrijk, maar begrensd argument voor de invoering van de toets.

Zo werd in *De Waarheid* van 26 februari 1977 beschreven dat de toets niet doorslaggevend was, maar slechts als hulpmiddel diende binnen een breder afwegingsproces waarin ook het oordeel van school en ouders een rol speelde. De toets kon met name uitkomst bieden wanneer ouders en school van mening verschilden over het vervolgonderwijs van een leerling, en fungeerde dan als zogenoemd derde middel. Daarbij wogen de opvattingen van ouders en het schoolhoofd ook veel zwaarder ("Schooltoets niet beslissend", 1977).

In het regionale *Tubantia* van 17 maart 1973 werd ook benadrukt dat het advies van het schoolhoofd en de uitslag van de toets samen doorgaans een treffend oordeel over leerlingen opleverden (“Ouders kunnen school niet dwingen hun kind te nemen”, 1973). In *De Volkskrant* van 26 mei 1976 stelden medewerkers van de Amsterdamse Onderwijswinkel bovendien dat het aan onderwijzers zelf moest worden overgelaten hoe en wanneer zij toetsen gebruikten ("Amsterdamse onderwijswinkel: Verbrede toets remt vernieuwing onderwijs", 1976). Dit laat zien dat de objectiviteit van de toets werd erkend, maar doorgaans ondergeschikt bleef aan andere beoordelingsfactoren, zoals het professionele oordeel van leerkrachten.

Deze ondergeschikte rol van de toets hing ook samen met een begrenzing van wat de toets kon meten. In verschillende artikelen werd benadrukt dat de toets slechts een hulpmiddel was en geen volledig beeld gaf van de bredere capaciteiten van een leerling ("Schooltoets niet beslissend", 1977; Van Spaandonk, 1973). Daarbij werd gesteld dat de toets zich voornamelijk richtte op kennis, terwijl andere aspecten zoals karakter en intelligentie (Van Spaandonk, 1973) en inzicht en creativiteit ("Schooltoets niet beslissend", 1977) minder werden meegenomen. In *De Telegraaf* van 20 februari 1971 werd het oordeel van het schoolhoofd daarom minstens even belangrijk geacht, omdat hij kon zeggen hoe het stond met zaken die niet in de toets voorkwamen, zoals studielust, belangstelling en milieu (Cohen, 1971). Ook in *Trouw* van 15 januari 1975 werd aandacht besteed aan wat door de toets werd gemeten. In dit artikel werd beschreven dat ouders in sommige gevallen slechter presteerden dan hun kinderen op onderdelen van de CITO-toets. Daarmee werd gesuggereerd dat de kennis die door de toets gemeten werd niet vanzelfsprekend overeenkwam met alledaagse kennis (Smit, 1975). De objectiviteit van de toets bleef daarmee beperkt tot datgene wat daadwerkelijk werd gemeten.

Hoewel de toets op dit punt werd bekritiseerd, werd ook gewezen op het bredere nut van de gegevens die de toets kon opleveren. Zo werd in *De Telegraaf* van 20 februari 1971 beschreven dat de toets gegevens opleverde die inzicht boden in de kwaliteit van onderwijs op verschillende scholen. Deze gegevens konden daarnaast worden gebruikt voor onderwijsbeleid en onderzoek (Cohen, 1971).

Verder werd in enkele artikelen kritiek geuit op het eenmalige karakter van de meting. De toets werd in *De Waarheid* van 26 februari 1977 beschreven als een eenmalig meetmoment, waardoor leerachterstanden pas laat zichtbaar werden, op een moment waarop bijsturing niet meer mogelijk was ("Schooltoets niet beslissend", 1977). In *Het Parool* van 1 maart 1977 werd het advies van onderwijzers, die een kind doorgaans over een langere tijd

hadden gevolgd, dan ook als “stukken betrouwbaarder” beschouwd dan een eenzijdige momentopname (Dwarshuis, 1977, p. 4). Daarmee werd het eenmalige meetmoment kritisch benaderd, omdat leerachterstanden pas laat zichtbaar werden en het oordeel van leerkrachten over langere tijd door sommigen waardevoller werd geacht.

Objectiviteit vormde in de geanalyseerde artikelen een belangrijk argument voor de invoering van de CITO-toets, maar werd in het publieke debat niet als absoluut gezien en voortdurend begrensd door andere beoordelingsfactoren en kritische kanttekeningen. De vraag in hoeverre toetsresultaten gewicht moesten krijgen binnen bredere beslissingen over leerlingen, raakte direct aan een volgend terugkerend thema: selectie.

Selectie

Naast discussies over objectiviteit werd in de artikelen ook de functie van de CITO-toets als selectiemiddel besproken, met name bij de overgang van het lager naar het voortgezet onderwijs.

In *De Telegraaf* van 20 februari 1971 werd beschreven dat toelating tot hogere vormen van voortgezet onderwijs, zoals havo en vwo, plaatsvond op basis van een combinatie van het oordeel van de school en een gunstige toets- of testuitslag. In hetzelfde artikel werd de toets gepresenteerd als een middel om verschillen in toelating te verminderen. Voorheen konden eisen en beoordelingen per school of plaats verschillen, waardoor leerlingen onder ongelijke voorwaarden werden toegelaten tot het voortgezet onderwijs. De toets zou deze verschillen kunnen verkleinen door een meer vergelijkbare beoordeling mogelijk te maken (Cohen, 1971).

Tegelijkertijd werd het gebruik van de toets als selectiemiddel in meerdere artikelen bekritiseerd. In *NRC Handelsblad* van 28 februari 1973 werd gesteld dat de schooltoets slechts een extra hulpmiddel was bij het bepalen van de verdere schoolloopbaan van leerlingen. Volgens het artikel was het oordeel van onderwijzers daarbij doorslaggevend en

diende dit niet uitsluitend gebaseerd te zijn op behaalde cijfers, maar ook op mogelijkheden van karakter en intelligentie (Van Spaandonk, 1973).

In *De Volkskrant* van 26 mei 1976 werd het vaak inzetten van de CITO-toets als selectiemiddel bij de overgang naar het voortgezet onderwijs zelfs als belangrijkste nadeel van de toets benoemd ("Amsterdamse onderwijswinkel: Verbrede toets remt vernieuwing onderwijs", 1976). In *Trouw* van 1 februari 1974 werd eveneens gesteld dat het belangrijkste bezwaar tegen de toets was dat scholen deze misbruikten om kinderen eenvoudigweg te selecteren (“Uitslag kán bepalend zijn voor schoolkeus van zesdeklassers”, 1974).

Vanuit het CITO zelf leek ook enige terughoudendheid te bestaan over het gebruik van de toets voor selectie. In *NRC Handelsblad* van 28 februari 1973 werd de indruk gewekt dat Solberg, directeur van het CITO, deze functie als weinig zinvol beschouwde (Van Spaandonk, 1973).

In een kritisch artikel uit *Het Parool* van 1 maart 1977, waarin een duidelijk standpunt tegen de toets werd ingenomen, werd daarnaast de noodzaak van voorafgaand onderzoek en toetsing ter discussie gesteld. Daarbij werd betoogd dat de schooltoets, mede in het licht van onderwijshervormingen zoals de Mammoetwet, als overbodig kon worden beschouwd voor de overgang naar het voortgezet onderwijs (Dwarshuis, 1977).

Uit de geanalyseerde bronnen kwam naar voren dat het gebruik van de CITO-toets als selectiemiddel niet vanzelfsprekend werd geaccepteerd. Hoewel de toets in sommige gevallen werd gezien als een hulpmiddel bij toelatingsbeslissingen en het verkleinen van verschillen tussen scholen, werd ook door verschillende bronnen benadrukt dat zij niet doorslaggevend mocht zijn en dat het oordeel van onderwijzers van belang bleef. Daarmee werd de toets enerzijds gepresenteerd als een hulpmiddel en onder voorwaarden geaccepteerd, maar anderzijds ook bekritiseerd of zelfs expliciet afgewezen. In sommige artikelen werd deze

kritiek bovendien nadrukkelijk geformuleerd, waarbij het gebruik van de toets voor selectie werd aangeduid als het belangrijkste nadeel.

Invloed op onderwijs

Naast de functie van de CITO-toets als selectiemiddel werd in de artikelen ook de invloed ervan op het onderwijs besproken. Daarbij werd gewezen op de mogelijkheid dat de toets het onderwijs zelf kon sturen.

Zo werd in *De Waarheid* van 26 februari 1977 benoemd dat leraren hun lesprogramma moesten afstemmen op het afnemen van de toets, waardoor het reguliere onderwijsprogramma kon worden verstoord ("Schooltoets niet beslissend", 1977). Ook in andere artikelen werd gewezen op het risico dat het onderwijs zich zou aanpassen aan de toets. In *De Volkskrant* van 26 mei 1976 werd bijvoorbeeld gesteld dat een uitgebreider toetssysteem de druk op scholen zou vergroten om de inhoud van het onderwijs aan de toets aan te passen, wat volgens het artikel een rem zou kunnen vormen op onderwijsvernieuwing ("Amsterdamse onderwijswinkel: Verbrede toets remt vernieuwing onderwijs", 1976).

In hetzelfde artikel werd deze kritiek verder aangescherpt door C. Wit, het hoofd van een montessorischool in Amsterdam. Hij stelde dat een dergelijk toetssysteem het onderwijs van binnenuit kapot zou maken, doordat onder andere de nadruk steeds sterker zou komen te liggen op intellectuele kennis en er nog minder terecht zou komen van een kindgerichte benadering in het onderwijs ("Amsterdamse onderwijswinkel: Verbrede toets remt vernieuwing onderwijs", 1976).

Het CITO zelf waarschuwde ook voor dit risico. In *NRC Handelsblad* van 28 februari 1973 stelde Solberg, directeur van het CITO, dat de schooltoets niet bepalend mocht zijn voor de inhoud van het onderwijs, omdat het gevaar bestond dat het onderwijs zich zou gaan richten op de toets, terwijl het juist de bedoeling was dat de toets het bestaande onderwijs zou meten (Van Spaandonk, 1973). In een eerder artikel in *NRC Handelsblad* van 29 februari

1972 werd bovendien al benadrukt dat het niet de taak van het CITO was om te bepalen wat leerlingen moesten kennen, aangezien dit tot het domein van het onderwijs zelf behoorde. Daarbij werd de rol van de CITO-toets nadrukkelijk begrensd tot een meetinstrument dat als één van meerdere indicatoren diende binnen het beoordelingsproces (Van Spaandonk, 1972).

Deze zorgen over de invloed van de toets op de inhoud van het onderwijs werden verder versterkt in *Het Parool* van 1 maart 1977, waarin werd gesteld dat de toets het onderwijs in het laatste leerjaar zou kunnen gaan beheersen (Dwarshuis, 1977). De zorgen beperkten zich niet uitsluitend tot de invloed van de toets op lesinhoud en onderwijspraktijk, maar hadden ook betrekking op de bredere positie van het CITO binnen het onderwijs. In het *Algemeen Dagblad* van 30 september 1978 werd gesteld dat de kritiek op de CITO-toets nog steeds aanwezig was, maar dat de emotionele lading daarvan was afgenomen. Tegelijkertijd werd gewezen op zorgen over de invloed van het CITO binnen het onderwijs (d'Hamecourt, 1978).

De legitimering van de CITO-toets werd daarmee niet alleen bepaald door de functie als meetinstrument, maar ook door zorgen over de invloed ervan op het onderwijs. Hoewel de toets werd gepresenteerd als een middel om het onderwijs te meten, werd in meerdere artikelen gewaarschuwd dat zij juist sturend kon werken en daarmee de inhoud en inrichting van het onderwijs kon beïnvloeden. Daarnaast werd gewezen op de groeiende invloed van het CITO binnen het onderwijs.

Effecten op leerlingen

Het thema effecten op leerlingen kwam minder frequent naar voren dan bijvoorbeeld objectiviteit en selectie. In enkele artikelen werd gewezen op de invloed van de CITO-toets op leerlingen zelf.

In *De Waarheid* van 26 februari 1977 werd beschreven dat kinderen nerveus reageerden op de situatie die rond de toets werd opgebouwd. Van Dorp, schoolbegeleider bij

de Stichting Advies en Begeleidingscentrum voor het onderwijs in Amsterdam, gaf aan zich gevoelsmatig in deze reactie te kunnen vinden, maar stelde tegelijkertijd dat de consternatie rond de toets niet terecht was. De moeilijkheidsgraad zou namelijk volgens hem gering zijn ("Schooltoets niet beslissend", 1977).

Ook in *Het Parool* van 1 maart 1977 werd gewezen op de negatieve gevolgen van de toets voor leerlingen. Daar werd gesteld dat de toets in het basisonderwijs zorgde voor spanning en “rompslomp” (Dwarshuis, 1977, p. 4).

Deze artikelen laten zien dat de CITO-toets ook werd bekritiseerd vanwege de spanning die het toetsmoment bij leerlingen kon oproepen. Tegelijkertijd werden deze zorgen in sommige gevallen gerelativeerd. Daarmee kwam dit thema binnen de geanalyseerde publieke discussie relatief beperkt naar voren.

Vergelijkbaarheid

Het thema vergelijkbaarheid kwam in de geanalyseerde artikelen naar voren doordat de CITO-toets nieuwe mogelijkheden bood om prestaties van leerlingen en scholen met elkaar te vergelijken.

In *De Waarheid* van 26 februari 1977 noemde Van Dorp dit expliciet als voordeel. Hij stelde dat de toets het mogelijk maakte om prestaties van leerlingen onderling, zowel binnen een klas als tussen verschillende scholen, eenvoudig te vergelijken. Tegelijkertijd plaatste hij kanttekeningen bij de manier waarop deze vergelijkingen werden geïnterpreteerd. Volgens hem werd te sterk gekeken naar percentielscores ten opzichte van andere leerlingen, terwijl meer aandacht zou moeten uitgaan naar de individuele prestatie van een leerling, zoals het individueel aantal correct gemaakte opgaven ("Schooltoets niet beslissend", 1977).

Een vergelijkbaar punt werd gemaakt in het *Limburgsch Dagblad* van 14 januari 1975, waarin werd benadrukt dat het toetscijfer slechts de positie van een leerling ten opzichte van anderen weergeeft (“Wat in Amsterdam redelijk is, valt in Heerlen tegen”, 1975). In *Trouw*

van 1 februari 1974 werd gewezen op misverstanden rond de uitslag van de toets, waarbij werd aangegeven dat percentielscores wezenlijk verschilden van gewone schoolcijfers (“Uitslag kán bepalend zijn voor schoolkeus van zesdeklassers”, 1974).

In *Het Parool* van 1 maart 1977 werd ook opgemerkt dat het CITO inmiddels landelijke vergelijkingen van toetsresultaten aanbood die mogelijk interessant waren. Volgens het artikel was dit echter niet het oorspronkelijke doel van de toets (Dwarshuis, 1977). In *De Volkskrant* van 26 mei 1976 werd gewezen op mogelijke bredere gevolgen van deze vergelijkbaarheid. Daar werd gesteld dat een uitgebreider toetssysteem de concurrentie tussen leerlingen en scholen zou kunnen versterken ("Amsterdamse onderwijswinkel: Verbrede toets remt vernieuwing onderwijs", 1976).

Deze argumenten laten zien dat vergelijkbaarheid enerzijds werd gezien als een voordeel van de CITO-toets, doordat prestaties op een vergelijkbare manier inzichtelijk werden. Anderzijds riep deze vergelijkbaarheid ook vragen op over de interpretatie van toetsresultaten, omdat de nadruk hierdoor kon verschuiven van individuele prestaties naar de relatieve positie van leerlingen. Daarnaast bestonden zorgen dat dergelijke vergelijkingen de concurrentie tussen leerlingen en scholen zouden kunnen versterken. De legitimering van de toets werd daarmee gekenmerkt door een spanningsveld tussen het nut van vergelijking en zorgen over de gevolgen daarvan.

Overkoepelende bevindingen

Samenvattend laten de resultaten zien dat objectiviteit het meest nadrukkelijk naar voren kwam in de wijze waarop de CITO-toets werd gelegitimeerd, gevolgd door argumenten rond selectie. Vergelijkbaarheid en invloed op het onderwijs keerden eveneens terug, terwijl effecten op leerlingen minder nadrukkelijk aanwezig waren. Gezamenlijk wijzen deze bevindingen erop dat de legitimering van de CITO-toets in het publieke debat een gelaagd karakter had, waarin argumenten voor de toets samengingen met terughoudendheid,

voorwaarden en kritiek. Deze terughoudendheid kwam ook vanuit het CITO zelf. Uit de uitspraken van toenmalig CITO-directeur Solberg blijkt dat de CITO-toets werd verdedigd als objectief meetinstrument, maar dat de rol ervan tegelijkertijd werd begrensd. Zo werd de toets gepresenteerd als hulpmiddel naast andere indicatoren binnen het beoordelingsproces, waarbij werd benadrukt dat zij de onderwijsinhoud niet mocht bepalen.

Conclusie

In dit onderzoek is antwoord gezocht op de vraag: ‘Hoe werd de invoering van de CITO-toets in Nederland in de jaren zeventig gelegitimeerd in publieke discussies?’ Uit de analyse blijkt dat objectiviteit het meest nadrukkelijk naar voren kwam in de wijze waarop de CITO-toets werd gelegitimeerd, gevolgd door argumenten rond selectie. In verschillende artikelen werd de toets gepresenteerd als een instrument dat kon bijdragen aan een meer systematische en minder willekeurige beoordeling van leerlingen. De toets werd verbonden aan objectieve gegevens, gelijke verwerking van antwoorden en deskundige toetsontwikkeling.

Tegelijkertijd werd deze objectiviteit in het publieke debat niet als absoluut beschouwd. De toets werd veelal gezien als hulpmiddel naast andere vormen van beoordeling, zoals het oordeel van leerkrachten, schoolhoofden en ouders. Ook werd benadrukt dat de toets slechts een deel van de capaciteiten van leerlingen kon meten en geen volledig beeld gaf van aspecten als karakter, creativiteit of ontwikkeling over langere tijd. Legitimering vond daarmee voornamelijk plaats onder de voorwaarde dat de toets niet los werd gebruikt van professioneel oordeel.

Ook de selectiefunctie van de toets werd zowel verdedigd als bekritiseerd. Enerzijds werd de toets gezien als middel om de toelating tot het voortgezet onderwijs meer vergelijkbaar te maken. Anderzijds riep juist deze selectiefunctie kritiek op, omdat de toets niet doorslaggevend mocht worden voor de verdere schoolloopbaan van leerlingen. In sommige artikelen werd dit gebruik expliciet bekritiseerd en aangeduid als misbruik of als het belangrijkste nadeel van de toets. Daarnaast werd in een artikel betoogd dat de toets, mede in het licht van onderwijsvernieuwingen, niet noodzakelijk was als selectiemiddel.

Verder werd gewezen op de invloed van de toets op het onderwijs. De toets kon de inhoud en inrichting van het onderwijs sturen, doordat scholen hun lessen gingen afstemmen op de inhoud van de toets.

Vergelijkbaarheid werd ook niet eenduidig gewaardeerd. De CITO-toetsresultaten maakten prestaties inzichtelijk en vergelijkbaar, maar konden ook leiden tot een te sterke nadruk op relatieve posities en concurrentie tussen leerlingen en scholen. Daarbij bestonden ook zorgen over de interpretatie van scores, omdat relatieve posities niet altijd een volledig beeld gaven van individuele capaciteiten.

Opvallend is dat het thema effecten op leerlingen minder prominent naar voren kwam dan de andere thema's. Daarbij werd vooral gewezen op spanning rond het toetsmoment, terwijl deze zorgen in sommige artikelen direct werden gerelativeerd door te benadrukken dat de moeilijkheidsgraad van de toets beperkt was.

De legitimering van de CITO-toets in het publieke debat berustte daarmee niet op één type argument, maar op een spanningsveld tussen argumenten rond objectiviteit, selectie, vergelijkbaarheid, invloed op het onderwijs en effecten op leerlingen.

Discussie

De geanalyseerde publieke discussies over de invoering van de CITO-toets in de jaren zeventig werden gekenmerkt door spanningsvelden rond objectiviteit, selectie, invloed op het onderwijs, effecten op leerlingen en vergelijkbaarheid. De resultaten sluiten daarmee aan bij internationale literatuur waarin gestandaardiseerde toetsing wordt beschreven als een blijvend omstreden fenomeen (Haladyna et al., 1998; Hutt & Schneider, 2018; Linn, 2001).

De nadruk op objectiviteit sluit aan bij onderzoek waarin objectiviteit, efficiëntie en verantwoording worden gezien als belangrijke historische drijfveren achter gestandaardiseerde toetsing (Huddleston & Rockwell, 2015). In het onderzochte debat kwamen vooral argumenten rond objectiviteit naar voren, terwijl efficiëntie en verantwoording minder expliciet aanwezig waren. Mogelijk kreeg objectiviteit veel aandacht omdat dit een belangrijk argument leek te vormen voor de invoering en het gebruik van de toets. Het CITO stimuleerde het gebruik van de toets namelijk mede door de nadruk op objectiviteit (Thio, 1994). Dit kan wijzen op bredere maatschappelijke aandacht voor meetbaarheid in het onderwijs, waarbij systematische meting willekeur in beoordeling kon verminderen. Mogelijk kan dit worden begrepen vanuit meritocratische opvattingen, die een belangrijke rol speelden in het Amerikaanse onderwijssysteem. Binnen meritocratische opvattingen kan gestandaardiseerde toetsing worden gezien als middel om leerlingen eerlijk op basis van prestaties te selecteren en kansen te bieden (Hutt & Schneider, 2018). Mogelijk speelden zulke opvattingen ook een rol in de nadruk op objectiviteit in het Nederlandse debat. Tegelijkertijd laat de terughoudendheid rond de toets zien dat deze nadruk op objectiviteit niet zonder meer betekende dat professioneel oordeel werd vervangen door toetsresultaten.

De resultaten ondersteunen ook Thio's (1994) observatie dat de toenemende erkenning van objectieve toetsing samenging met groeiende kritiek op een mogelijke ongewenste terugslag op het onderwijs en een verarming van onderwijsdoelstellingen. In het publieke

debat werd gewaarschuwd dat scholen hun lesprogramma op de CITO-toets zouden afstemmen. Dit sluit aan bij onderzoek waaruit blijkt dat high-stakes toetsing samenhangt met veranderingen in het curriculum en kan leiden tot een sterkere focus op getoetste onderwerpen (Au, 2007). In ander onderzoek komt een vergelijkbaar patroon naar voren, waarbij gestandaardiseerde toetsing in de Verenigde Staten al langere tijd gepaard gaat met kritiek op effecten op het curriculum en leerlingen (Hutt & Schneider, 2018).

Hoewel het thema effecten op leerlingen minder prominent naar voren kwam, laat ander onderzoek zien dat leerlingen meer spanning en angst kunnen ervaren bij high-stakes toetsing dan bij reguliere toetsen (Segool et al., 2013). Mogelijk weerspiegelt dit de prioriteiten van het toenmalige publieke debat, waarin meer aandacht uitging naar objectiviteit en deskundigheid dan naar de beleving van leerlingen. In de artikelen kwamen vooral onderwijsdeskundigen en toenmalig CITO-directeur Solberg aan het woord, terwijl leerlingen zelf nauwelijks werden gehoord.

Bij dit onderzoek dienen enkele beperkingen te worden genoemd. Hoewel thematische verzadiging werd nagestreefd, hadden aanvullende bronnen mogelijk andere accenten laten zien. Daarnaast zijn landelijke kranten sterker vertegenwoordigd dan regionale kranten, doordat de gebruikte zoektermen vaker resultaten opleverden in landelijke kranten. Uitspraken zijn zoveel mogelijk vanuit de historische context geïnterpreteerd, maar enige mate van interpretatie door de onderzoeker blijft onvermijdelijk. Ook zijn krantenberichten vaak gericht op nieuwswaardige ontwikkelingen, waardoor kritische geluiden relatief sterk vertegenwoordigd kunnen zijn geweest.

Verder blijft dit onderzoek beperkt tot het publieke debat in de geanalyseerde krantenartikelen. Vervolgonderzoek kan hierop voortbouwen door het professionele debat in vaktijdschriften te analyseren. Daarnaast kan vergelijking met recente discussies over de

doorstroomtoets inzicht geven in continuïteit en verandering binnen Nederlandse debatten over toetsing.

Dit onderzoek draagt bij aan een beter historisch begrip van de maatschappelijke legitimering van toetsing in Nederland. De invoering van de CITO-toets was niet alleen een onderwijskundige ontwikkeling, maar ook onderwerp van publiek debat. Door historische discussies over testen in het onderwijs te onderzoeken, kunnen huidige discussies hierover beter begrepen worden. Dit onderzoek maakt zichtbaar welke argumenten rond de CITO-toets in publieke discussies naar voren kwamen en wie daarbij vooral werden gehoord. In de onderzochte krantenartikelen kwamen vooral onderwijsdeskundigen en CITO-directeur Solberg aan het woord, terwijl leerkrachten en leerlingen minder zichtbaar waren. Daarmee laat dit onderzoek niet alleen zien welke argumenten over testen circuleerden, maar ook wie invloed had op de publieke beeldvorming hierover.

Literatuurlijst

Au, W. (2007). High-Stakes Testing and Curricular Control: A Qualitative Metasynthesis. *Educational Researcher*, *36*(5), 258–267. https://doi.org/10.3102/0013189X07306523

Broekman, P. (1981). Projects at the National Institute for Educational Measurement (Centraal Instituut Voor Toetsonwikkeling--CITO), of The Netherlands. *European Journal of Science Education*, *3*(4), 487–489. https://doi.org/10.1080/0140528810304016

Canales, A. Fco., & Polenghi, S. (2019). Classifying children: a historical perspective on testing and measurement*. *Paedagogica Historica*, *55*(3), 343–352. https://doi.org/10.1080/00309230.2019.1611890

De Groot, A. D. (1978). *Vijven en zessen: cijfers en beslissingen: het selectieproces in ons onderwijs* (9e dr.). Wolters-Noordhoff.

Haladyna, T., Haas, N., & Allison, J. (1998). Continuing Tensions in Standardized Testing. *Childhood Education*, *74*(5), 262–273. https://doi.org/10.1080/00094056.1998.10521950

Huddleston, A. P., & Rockwell, E. C. (2015). Assessment for the Masses: A Historical Critique of High-Stakes Testing in Reading. *Texas Journal of Literacy Education, 3*(1), 38–49.

Hutt, E., & Schneider, J. (2018). A History of Achievement Testing in the United States Or: Explaining the Persistence of Inadequacy. *Teachers College Record: The Voice of Scholarship in Education*, *120*(11), 1–34. https://doi.org/10.1177/016146811812001102

Linn, R. L. (2001). A Century of Standardized Testing: Controversies and Pendulum Swings. *Educational Assessment*, *7*(1), 29–38. https://doi.org/10.1207/S15326977EA0701_4

NOS. (2025, 15 januari). *Scholen weigeren doorstroomtoets: 'Laat rechter erover oordelen'.*

https://nos.nl/artikel/2553572-scholen-weigeren-doorstroomtoets-laat-rechter-erover-oordelen

Rury, J. L., Belew, R., & Hurst, J. (2022). The Origins of American Test-Based Educational Accountability and Controversies about Its Impact, 1970-1983. *Teachers College Record: The Voice of Scholarship in Education*, *124*(1), 143–163. https://doi.org/10.1177/01614681221086094

Scheider, S., Rosenfeld, S., Bink, S., & Lecina, N. (2023). Educational inequality due to lack of validity: A methodological critique of the Dutch school system. *International Journal of Educational Research*, *117*. https://doi.org/10.1016/j.ijer.2022.102097

Schrag, Z. M. (2021). Sources: AN INTRODUCTION. In *The Princeton Guide to Historical Research* (pp. 103–119). Princeton University Press. https://doi.org/10.2307/j.ctv1pdrrc9.8

Segool, N. K., Carlson, J. S., Goforth, A. N., von der Embse, N., & Barterian, J. A. (2013). Heightened Test Anxiety among Young Children: Elementary School Students' Anxious Responses to High-Stakes Testing. *Psychology in the Schools*, *50*(5), 489–499. https://doi.org/10.1002/pits.21689

Terwel, J. (2006). *Is de school een sorteermachine? Schoolkeuze en schoolloopbaan van leerlingen van 10–16 jaar*. Vrije Universiteit Amsterdam.

Thio, D. K. (1994). The Educational Testing Movement in the Netherlands: Some Historical and Political Aspects. *Educational Measurement: Issues and Practice*, *13*(2), 21–23. https://doi.org/10.1111/j.1745-3992.1994.tb00792.x

Van Drunen, P., & Van Strien, P. J. (1991). *Op de proef gesteld: Geschiedenis van de psychologische test in Nederland*. Stichting Historische Materialen Psychologie.

Literatuurlijst primaire bronnen

Amsterdamse onderwijswinkel: “Verbrede toets remt vernieuwing onderwijs”. (1976, 26 mei). *De Volkskrant*, p. 6.

Cohen, I. (1971, 20 februari). Eerlijk, maar wel mechanisch. *De Telegraaf*, p. 7.

d’Hamecourt, P. (1978, 30 september). Argwaan tegen ‘meting’ scholier. *Algemeen Dagblad*, p. 55.

Dwarshuis, S. (1977, 1 maart). Schooltoets kan beter weg. *Het Parool*, p. 4.

Ouders kunnen school niet dwingen hun kind te nemen. (1973, 17 maart). *Tubantia*.

Schooltoets niet beslissend: “Over enkele dagen zitten de kinderen weer in de zenuwen”. (1977, 26 februari). *De Waarheid*, p. 5.

Smit, L. (1975, 15 januari). Tabellen lezen is niet niks. *Trouw*.

Uitslag kán bepalend zijn voor schoolkeus van zesdeklassers: Maart: toetstijd voor duizenden kinderen. (1974, 1 februari). *Trouw*.

Van Spaandonk, J. W. M. (1972, 29 februari). CITO op eigen toetsen getoetst. *NRC Handelsblad*, p. 9.

Van Spaandonk, J. W. M. (1973, 28 februari). Zesde klassers van basisonderwijs zijn aan het “toetsen”. *NRC Handelsblad*, p. 3.

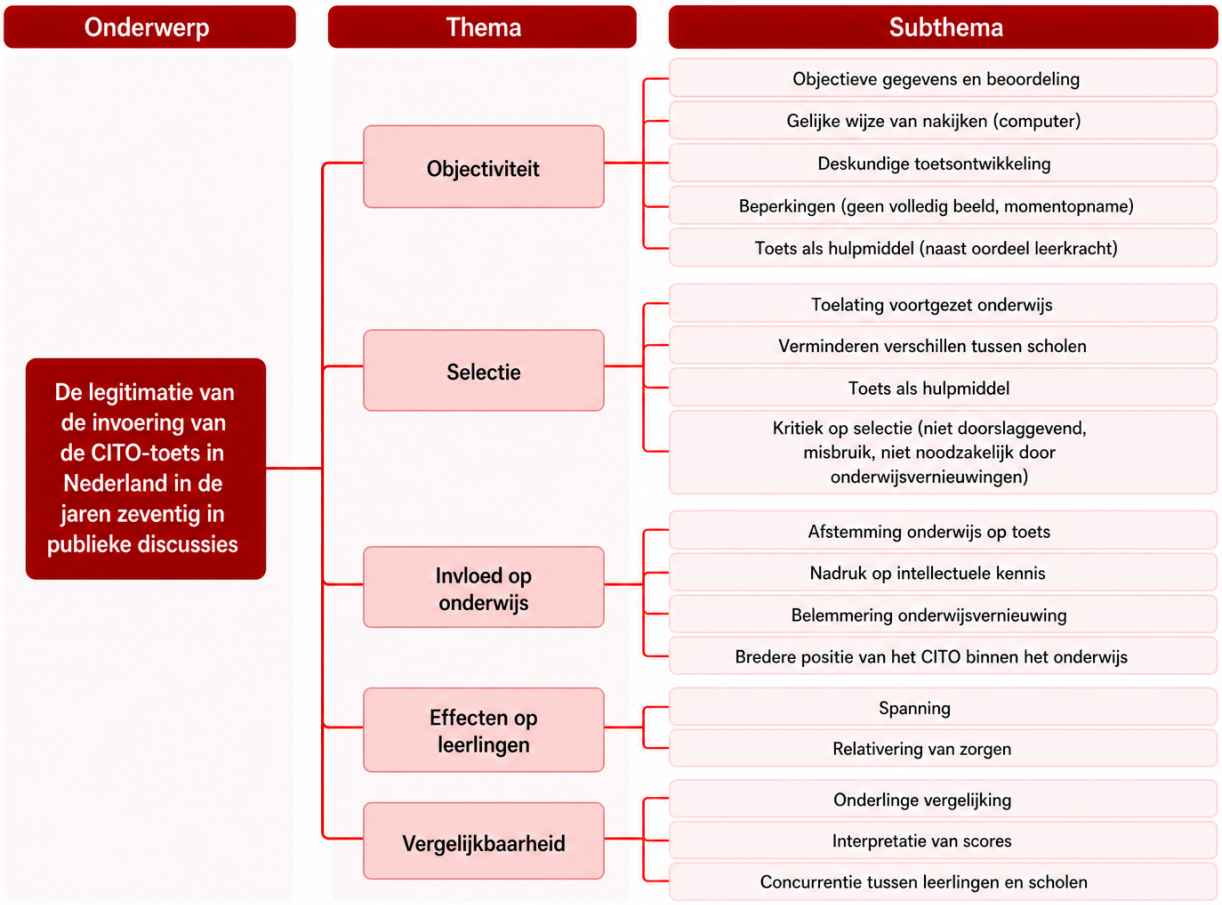
Wat in Amsterdam redelijk is, valt in Heerlen tegen. (1975, 14 januari). *Limburgsch Dagblad*, p. 11.

Bijlage A

Codeboom

Figuur A1

Codeboom van de thematische analyse



Bijlage B

Verklaring gebruik generatieve AI-tools

Inleiding

• Gebruikte AI-systeem/-systemen: ChatGPT via www.chatgpt.com.

• Gebruikte definitieve prompts: “Beoordeel mijn inleiding op basis van de gegeven rubric. Geef per punt aan wat ik nog moet verbeteren, zodat ik dit zelf eventueel kan aanpassen in mijn tekst.”

• Toepassing: Gebruikt voor het verkrijgen van feedback op de structuur, inhoud en formulering van de inleiding.

• Aanpassingen: De feedback is gebruikt als richtlijn en kritisch beoordeeld. De tekst is volledig zelfstandig geschreven en herschreven. AI-gegenereerde tekst is niet direct overgenomen.

Methode

• Gebruikte AI-systeem/-systemen: ChatGPT via www.chatgpt.com.

• Gebruikte definitieve prompts: “Beoordeel mijn methode op basis van de gegeven rubric. Geef per punt aan wat ik nog moet verbeteren, zodat ik dit zelf eventueel kan aanpassen in mijn tekst.” ;

“Geef aan welke woorden nog in de verleden tijd moeten.”

• Toepassing: Gebruikt voor feedback en taalcorrectie (werkwoordstijden).

• Aanpassingen: Suggesties zijn kritisch beoordeeld en eventueel handmatig verwerkt. De tekst is volledig zelfstandig geformuleerd en herschreven. AI-gegenereerde tekst is niet direct overgenomen.

Resultaten

• Gebruikte AI-systeem/-systemen: ChatGPT via www.chatgpt.com.

• Gebruikte definitieve prompts: “Is het verstandig om eerst een begin te maken met de analyse of eerst meer bronnen te zoeken?”;

“Beoordeel mijn resultaten op basis van de gegeven rubric. Geef per punt aan wat ik nog moet verbeteren, zodat ik dit zelf eventueel kan aanpassen in mijn tekst.”

• Toepassing: Gebruikt voor advies over de werkwijze bij de analyse en voor feedback.

• Aanpassingen: Het advies is kritisch beoordeeld en gebruikt om een keuze te maken voor een bepaalde werkwijze. De analyse en interpretatie van de data zijn volledig zelfstandig uitgevoerd. De feedback is gebruikt als richtlijn en kritisch beoordeeld. De tekst is volledig zelfstandig geschreven en herschreven. AI-gegenereerde tekst is niet direct overgenomen.

Conclusie en discussie

• Gebruikte AI-systeem/-systemen: ChatGPT via www.chatgpt.com.

• Gebruikte definitieve prompts: “Beoordeel mijn conclusie en discussie op basis van de gegeven rubric. Geef per punt aan wat ik nog moet verbeteren, zodat ik dit zelf eventueel kan aanpassen in mijn tekst.”

• Toepassing: Gebruikt voor het verkrijgen van feedback op de structuur, inhoud en formulering van de conclusie en discussie.

• Aanpassingen: De feedback is gebruikt als richtlijn en kritisch beoordeeld. De tekst is volledig zelfstandig geschreven en herschreven. AI-gegenereerde tekst is niet direct overgenomen.