Onderzoek naar de kwaliteit van schoolexamens

Technisch rapport

Mei 2026

Inhoud

Managementsamenvatting 4
1. Inleiding 6
1.1 Aanleiding van het onderzoek 6
1.2 Achtergrond 7
1.2.1 Opzet van het Nederlandse eindexamen 7
1.2.2 De kwaliteit van het SE 8
1.2.3 Actoren bij de SE's 11
1.3 Onderzoeksvragen 12
2. Methoden 13
2.1 Methodologische opzet 13
2.2 Vragenlijsten docenten en examencommissie 13
2.2.1 Doel van het onderzoek 13
2.2.2 Ontwikkeling van de vragenlijsten 14
2.2.3 Kenmerken vragenlijsten 15
2.2.4 Werving 16
2.2.5 Analyses 16
2.3 Evaluatie SE's 17
2.3.1 Relevante materialen behorend bij SE's 17
2.3.2 Ontwikkeling van beoordelingskader 19
2.3.3 Beoordelaars 21
2.3.4 Werving 21
2.3.5 Terugkoppeling aan docenten 22
2.3.6 Analyse 22
2.4 Interviews vakverenigingen 22
2.4.1 Methode 22
2.4.2 Werving 23
2.4.3 Verwerking van de gegevens 24
3. Resultaten 25
3.1 Resultaten van het vragenlijstonderzoek 25
3.1.1 Algemeen 25
3.1.2 Constructie 26
3.1.3 Regels en richtlijnen 29
3.1.4 Leerlingen 31
3.1.5 Beoordelen 31
3.1.6 Evaluatie en deskundigheid 32
3.1.7 Positieve punten en verbeterpunten 34
3.1.8 Belangrijkste resultaten en conclusies uit het vragenlijst onderzoek 35

3.2 Resultaten van de evaluatie van SE's 37
3.2.1 Inventarisatie van de gedeelde materialen 37
3.2.2 Inhoudelijke beoordelingen 39
3.2.3 Relatie tussen CE en SE 44
3.2.4 Open vragen: tips en tops voor de docenten 45
3.2.5 Belangrijkste resultaten en conclusies op basis van de evaluatie van SE's 46
3.3 Resultaten van de vakverenigingen 48
3.3.1 Positionering van de vakverenigingen 48
3.3.2 Activiteiten van de vakverenigingen 49
3.3.3 De SE's als thema 50
3.3.4 De kwaliteit van de SE's 50
3.3.5 Vakverenigingen en hun blik op SE's 50
3.3.6 Ondersteuningsbehoefte van vakverenigingen 51
4. Discussie en conclusie 52
4.1 Wat gaat er goed? 52
4.2 Wat kan er beter? 53
4.2.1 Het gebruik van CE's en lesmethodemateriaal bij de SE's 53
4.2.2 Rol examencommissie 53
4.2.3 Beoordeling en normering 54
4.2.4 Transparantie 54
4.2.5 Rol vakverenigingen 54
4.3 Methodologische reflectie 55
4.4 Conclusies en aanbevelingen 56
4.4.1 Aanbevelingen voor vakdocenten 56
4.4.2 Aanbevelingen voor examencommissies 57
4.4.3 Aanbevelingen voor schoolleiding 58
4.4.4 Aanbevelingen voor vakverenigingen 58
Dankwoord en verantwoording 60

Managementsamenvatting

Dit onderzoek richt zich op de vakinhoudelijke en toetstechnische kwaliteit van schoolexamens (SE's) binnen het Nederlandse voortgezet onderwijs. In 2023 voerde de Inspectie van het Onderwijs[1] een grootschalig onderzoek uit naar de kwaliteit van de schoolexaminering, op basis waarvan zij aanbevelingen gaven om de kwaliteit van de schoolexaminering te versterken. Voor scholen betreft dit onder meer het versterken van examencommissies en het verbreden, expliciteren en implementeren van een visie op examinering, gesteund door een verduidelijking van de regelgeving van het ministerie (met de VO-raad). Ook adviseerden zij OCW om landelijk onderzoek te (laten) doen naar de kwaliteit van de SE's zelf, wat de directe aanleiding vormde voor dit rapport. Dit onderzoek beoogt in kaart te brengen hoe de huidige praktijk rond de SE's zich verhoudt tot vakinhoudelijke en toetstechnische kwaliteitscriteria, welke actoren hierbij betrokken zijn, en welke knelpunten en kansen zich voordoen om de kwaliteit van de SE's te versterken.

De methodologische opzet van dit onderzoek combineert kwantitatieve en kwalitatieve onderzoeksmethoden. Zo zijn er vragenlijsten afgenomen onder docenten en leden van examencommissies, waarin zowel de inhoudelijke aspecten van de examens als organisatorische (proces)factoren werden geanalyseerd. Ook is er een beoordelingskader ontwikkeld om de SE's systematisch te evalueren op toetskwaliteit. Daarnaast zijn er interviews gehouden met vertegenwoordigers van vakverenigingen. De deelname aan het onderzoek was vrijwillig om zo de ethische zorgvuldigheid te borgen. Dit heeft wel impact op het aantal respondenten, de representativiteit, en daarmee de diepgang en statistische robuustheid. Hierdoor zijn de resultaten vooral beschrijvend en verkennend van aard.

De resultaten van het onderzoek geven aan dat er stevige fundamenten aanwezig zijn voor het ontwikkelen van kwalitatief goede SE's, maar ook significante aandachtspunten. De vragenlijsten en beoordelingen van de SE's laten ook zien dat er veel goed gaat: de basisopzet, vormgeving, taalkundige en praktische kwaliteit scoren vaak goed, docenten ervaren vaksecties als deskundig en toetsen stemmen doorgaans met het beoogde eindniveau overeen. Echter, dit onderzoek laat ook zien dat veel SE's sterk lijken te leunen op (oude) CE-opgaven en methodemateriaal, waardoor de onderwijsvisie van de school mogelijk niet altijd expliciet en/of bewust terugkomt in de SE's. Daardoor raken twee essentiële doelen uit beeld: het schoolspecifieke karakter van het schoolexamen en de rol ervan als afsluitend onderdeel van het eindexamen. Beide zijn sterk verbonden met de kwaliteit van de examinering (Commissie Kwaliteit Schoolexaminering, 2018)[2]. Daarnaast geven docenten en examencommissieleden aan dat tijdsgebrek een beperkende factor is bij de constructie en evaluatie van de SE's. Vanuit de SE-beoordelingen lijkt ook dat de transparantie, beoordelingsmodellen, het gebruik van toetsmatrijzen en logica van weging regelmatig niet in orde of verdedigbaar zijn. Daarnaast lijkt de samenwerking tussen docenten en examencommissies vaak beperkt. De positie en rol van vakverenigingen worden als waardevol ervaren, maar deze worden nog onvoldoende benut in de uitwisseling van *best practices* en normering. Kortom, de huidige praktijk kent een wisselend kwaliteitsniveau met voldoende aanknopingspunten voor verbetering en verdere professionalisering van de SE's.

Op basis van deze bevindingen komen wij tot aanbevelingen voor docenten, examencommissies, schoolbesturen en vakverenigingen. Daar waar die in het rapport per groep uitgesplitst zijn, vatten we aanbevelingen als volgt samen:

[1] Inspectie van het Onderwijs (2023). Technisch rapport themaonderzoek OP6. Kwaliteit van de schoolexaminering. Utrecht: Inspectie van het onderwijs. Geraadpleegd via: https://www.onderwijsinspectie.nl/documenten/rapporten/2023/10/12/kwaliteit-van-de-schoolexaminering

[2] Commissie Kwaliteit Schoolexaminering (2018). Een volwaardig schoolexamen.

Versterk de rol en zicht van examencommissies

Doelgroep: examencommissies en schoolleiding

Gezien hun positie als kwaliteitsbewaarder verdient het aanbeveling om de rol van examencommissies te versterken in het borgen van de kwaliteit en het afsluitende karakter van SE's. Zij kunnen ook een stimulerende rol hebben in het aandragen van alternatieven voor het gebruik van oude CE opgaven en methodetoetsen. In samenwerking met de schoolleiding kunnen zij ook bespreken hoe de visie van de school op onderwijs en op toetsing nog meer een weerslag kan krijgen in de SE's.

Hiermee kunnen ook heldere afspraken op schoolniveau ontwikkeld worden over cesuren, omzetting van scores naar cijfers en het gebruik van tweede beoordelaars om objectiviteit en betrouwbaarheid te verbeteren. Ook is het van belang om een visie op toetsing te ontwikkelen en communiceren die richting geeft aan het beleid, de toetspraktijk en kwaliteitsbewaking van SE's.

De examencommissie krijgt daarmee niet alleen een controlerende, maar ook een stimulerende rol. Dit vergt aanvullende training en heldere richtlijnen zodat beide rollen goed en eenduidig kunnen worden ingevuld. Bij de uitwerking hiervan kunnen het landelijk platform voor examensecretarissen en examencommissieleden (PLEXS) en/of de VO-Raad mogelijk ondersteuning bieden.

Reserveer tijd voor bijscholing

Doelgroep: schoolleiding, examencommissie en docenten

De docenten die deelnamen aan dit onderzoek voelen zich competent bij het maken van schriftelijk opgaven, al wordt daar nog wel uit efficiëntieoverwegingen regelmatig uit het CE geput om tot opgaven te komen. Het gebruik van een toetsmatrijs is echter vaak een stap te ver. Juist wanneer de variatie in toetsvormen binnen het SE geambieerd wordt, is ondersteuning bij het maken van dergelijke vormen, alsook het beoordelen en normeren ervan van belang. De toets is niet alleen het maken van opgaven en opdrachten, maar ook een borging van de kwaliteit door gebruik van een toetsmatrijs, controle op beoordelingen, en goed onderbouwde doordachte normeringen. Ook wanneer een examencommissie ervaart dat zij op bepaalde onderdelen kennis of kunde missen en om de eerste aanbeveling uit te voeren, is op die punten gerichte bijscholing ook aan te bevelen.

Voor de schoolleiding betekent dit ook een investering in gerichte bijscholing voor toetsconstructie, gebruik van toetsmatrijs en consistente beoordeling/normering, zodat scholen kwalitatief verantwoorde en gevarieerde SE-toetsen kunnen realiseren. Het is van belang dat de inzet van tijd door docenten en examencommissieleden door de schoolleiding gefaciliteerd wordt.

Stimuleer samenwerking tussen docenten

Doelgroep: docenten (binnen een tussen vaksecties); facilitering door schoolleiding

Een stap die mogelijk laagdrempelig is, zijn onderlinge samenwerkingsverbanden binnen een school. Daar waar docenten al wel vaak volgens een vier-ogen-principe werken, kan de interne samenwerking binnen scholen ook versterkt worden. Dit kan gedaan worden door van elkaar te leren in het gebruik van (versimpelde) toetsmatrijzen, of het maken van creatieve opgaven. Zeker als docenten naar bijscholingen zijn geweest kunnen zij het geleerde binnen de school verspreiden.

Samenwerking van docenten over scholen heen met een versterkte rol van vakverenigingen

Doelgroep: vakverenigingen, vakdocenten en schoolleiding

Het is aan te bevelen om vakverenigingen te stimuleren om hun rol in het ondersteunen van scholen bij toetsontwikkeling en kwaliteitsverbetering uit te breiden, bijvoorbeeld door het faciliteren van sectoroverstijgende kennisuitwisseling. Hierbij is het van belang om voldoende ruimte en professionaliseringsmogelijkheden te creëren voor docenten.

1. Inleiding

1.1 Aanleiding van het onderzoek

In Nederland speelt het diploma van het voortgezet onderwijs een cruciale rol: het fungeert als startkwalificatie en opent de deur naar vervolgopleidingen. Dat civiele effect komt echter alleen tot stand als er voldoende vertrouwen bestaat in het diploma. Het maatschappelijke vertrouwen in de waarde van het diploma is cruciaal. De acceptatie van een diploma hangt zowel af van de (ervaren) juistheid van de beslissing om het uit te reiken als van de acceptatie van de inhoud van het eindexamen. Dit eindexamen bestaat uit een centraal examen (CE) en een schoolexamen (SE). De bepalende rol van het eindexamen, inclusief de toegepaste normen, moet waarborgen dat leerlingen die een diploma ontvangen een vastgesteld basisniveau aan kennis en vaardigheden hebben bereikt.

Voor elk vak vormt het SE minimaal de helft van het eindcijfer van het eindexamen, en voor sommige vakken of onderdelen bepaalt het zelfs het gehele eindoordeel. Het SE is daarmee een wezenlijk onderdeel van het eindexamen. Het is zodoende van groot belang dat de kwaliteit van de SE's zelf en het proces van schoolexaminering goed geborgd zijn, zodat er breed vertrouwen bestaat in de kwaliteit, objectiviteit en consistentie van de eindexamens als geheel. Hierbij hebben de twee samenstellende delen van het eindexamen duidelijk een eigen rol.

Het CE levert een uniforme, betrouwbare maatstaf doordat alle leerlingen hetzelfde examen afleggen, ongeacht school of regio. Het SE vult dit idealiter aan door vaardigheden en kennis te toetsen die zich moeilijk laten standaardiseren. Doordat scholen een invulling moeten geven aan het SE, geeft dit hen de ruimte om de toetsing aan te laten sluiten bij hun eigen identiteit en onderwijsvisie. Via het SE kunnen scholen accenten leggen die passen bij hun profiel — bijvoorbeeld technisch vakmanschap, kunstzinnige vorming of maatschappelijke betrokkenheid — en krijgen docenten meer regie over de inrichting van toetsing. Deze autonomie kan het eigenaarschap en de afstemming tussen toetsen en het dagelijkse onderwijs versterken. De combinatie van beide examenvormen maakt het mogelijk zowel validiteit als betrouwbaarheid te waarborgen — een noodzakelijke balans voor een rechtvaardig examenstelsel. De combinatie van CE's en SE's in Nederland is uniek in vergelijking met andere landen.

Gegeven de belangrijke rol die het SE vervult en het gewicht dat het SE heeft in de eindexaminering, is het cruciaal dat de kwaliteit van het SE op orde is. Incidenten met SE's in 2018 (Maastricht) en 2019 (Amsterdam) brachten kwetsbaarheden in het systeem aan het licht. Naar aanleiding hiervan zijn diverse onderzoeken uitgevoerd, onder meer door de Onderwijsraad (2018)[3], de Commissie Kwaliteit Schoolexaminering (2018) [4] in opdracht van de VO-raad, en door het Nationaal Regieorgaan Onderwijsonderzoek gefinancierde onderzoek van Scheerens et al. (2019)[5] . Daarnaast voerde in 2019/2020 de Inspectie van het Onderwijs onderzoek uit naar de zorgvuldigheid van examinering[6]. Dit onderzoek leidde tot een pakket aan aanbevelingen en een wetswijziging in 2021 met bijbehorende regels[7].

[3] Onderwijsraad (2018). Toets wijzer - Naar een eigen(tijdse) wijze van toetsen en examineren. Den Haag: Onderwijsraad.
[4] Commissie Kwaliteit Schoolexaminering (2018). Een volwaardig schoolexamen.
[5] Scheerens, J., Brouwer, A., Sanders, P., Veldkamp, B. en Van der Vegt, A.L (2019). Fundamentele vragen over examens en toetsing. Utrecht: Oberon B.V.
[6] Inspectie van het Onderwijs (2019). Schoolexaminering in het voortgezet onderwijs: Naar een zorgvuldig verloop en een betere Kwaliteitsbewaking. Utrecht: Inspectie van het Onderwijs.
[7] Inspectie van het Onderwijs (2025). Onderzoekskader 2021 voor het toezicht op het voortgezet onderwijs. Versie 2025. Utrecht: Inspectie van het Onderwijs.

In 2023 voerde de Inspectie van het Onderwijs opnieuw een grootschalig onderzoek uit naar de kwaliteit van de schoolexaminering, met aandacht voor onder meer PTA's en het functioneren van examencommissies, om na te gaan of de hervormingen effect hebben gehad. Ook dit rapport bevatte aanbevelingen, onder meer gericht op examencommissies en op het formuleren en implementeren van een toetsvisie. De inspectie adviseerde daarnaast OCW om landelijk onderzoek te laten uitvoeren naar de kwaliteit van de examens zelf, aangezien eerdere rapporten vooral ingingen op het proces van schoolexaminering en de bijbehorende regels. Die aanbeveling is de directe aanleiding voor dit rapport, waarin de vakinhoudelijke en toetstechnische kwaliteit van SE's centraal staat. Evaluatie van de vakinhoudelijke kwaliteit betekent dat onderzocht wordt op welke wijze de inhoud van het vak gerepresenteerd wordt in de toets, bijvoorbeeld door te kijken naar het gebruik van een toetsmatrijs. De toetstechnische kwaliteit betreft de manier waarop het gevraagd wordt correct is, en of technische zaken zoals de normering en de betrouwbaarheid op orde zijn.

In oktober 2024 is in een kamerbrief[8] dit onderzoek naar de kwaliteit van toetsen aangekondigd. Daarin is gesteld dat naast de acties ter verbetering van de schoolexamenorganisatie ook de kwaliteit van de SE's zelf in beeld wordt gebracht. De kamerbrief geeft aan dat dit onderzoek zich richt op de toetstechnische kwaliteit van het SE en de factoren die van invloed zijn hierop. Er is aangegeven dat de focus van het onderzoek ligt op het achterhalen van de sterke en zwakke punten van huidige SE's , om zo vooral te richten op de deskundigheidsbevordering die scholen kan helpen bij het versterken van de kwaliteit. In de kamerbrief is ook de relatie gelegd met de aanbeveling aan het ministerie van OCW uit het inspectierapport.

1.2 Achtergrond

1.2.1 Opzet van het Nederlandse eindexamen

De twee onderdelen van het Nederlandse eindexamen worden in een vaste volgorde afgenomen. Het SE vindt altijd plaats vóór het CE waarbij het CE pas bij een leerling mag worden afgenomen nadat de school het cijfer op het SE van de leerling heeft vastgesteld. Het College voor Toetsen en Examens (CvTE) is verantwoordelijk voor de inhoudelijke kwaliteit, het ontwerp, de normering en de landelijke regels van het CE. Stichting Cito stelt in opdracht van het CvTE de examens samen en adviseert over de normering. DUO verzorgt de aanmeldingen en levert de uitslagen, en de scholen zijn verantwoordelijk voor de afname en het nakijken van het CE. De samenstelling, afname, correctie en cijfermatige beoordeling van de CE's zijn daarmee vooral landelijk geregeld. Het SE wordt echter samengesteld door scholen zelf, en ook normering en administratieve afhandeling zijn de verantwoordelijkheid van de school. Wel vergelijkbaar met het CE is dat scholen ook bij het SE verantwoordelijk zijn voor de afname en het geven van beoordelingen voor de geleverde leerlingprestaties.

Het SE en het CE zijn complementair; op www.examblad.nl staat per onderwijstype en vak aangegeven welk deel van het examenprogramma in het SE en welk in het CE wordt geëxamineerd. Er kan overlap zitten in de onderwijsinhoud van het SE en het CE, maar er wordt gewerkt naar een opdeling waarbij er geen overlap meer is[9].

Het SE biedt veel ruimte aan scholen voor eigen invulling van de leerstof en toetsvormen. In tegenstelling tot de CE's, die in een beperkte periode in het eindexamenjaar worden afgenomen, vinden

[8] Mariëlle Paul (2024). Kamerbrief eindexamenresultaten 2024 en kwaliteit schoolexamen (referentie: 47982768) – 10 oktober 2024. https://open.overheid.nl/documenten/dpc-dd1011931cbb094fb770db6e92bd4567be130136/pdf

[9] Zo wordt in de werkopdracht 2024 (SLO, *Het proces* (actualisatie examenprogramma's pagina 8, punt 8) genoemd dat de inhoudsverdeling qua ontwerpruimte evenwichtig (50%-50%) verdeeld moet worden over de SE- en CE-domeinen, overeenkomstig de weging van het SE en het CE in het eindcijfer. De inhoud zal dan op het niveau van domeinen en subdomeinen (in plaats van losse eindtermen) zoveel mogelijk toegewezen worden aan dan wel het SE dan wel het CE.

de SE's vaak verspreid plaats over een periode van twee à drie jaar. Het zwaartepunt van de afname van het SE ligt meestal in het examenjaar, maar niet binnen een landelijk vastgestelde periode. Een uitzondering vormen sommige onderdelen, zoals de kijk-en-luistertoets voor moderne vreemde talen, die vaak wordt ingekocht en wel aan een specifieke afnameperiode is gebonden.

De ruimte die scholen hebben om het SE in te richten maakt ook dat zij zich moeten verantwoorden, onder meer richting de Inspectie van het Onderwijs. De inspectie ziet landelijk toe op de procedures rond de afname van SE's en hanteert een onderzoekskader waarin staat aan welke eisen scholen moeten voldoen bij de uitvoering van SE's. Jaarlijks vóór 1 oktober moeten scholen het zogenaamde examenreglement en het Programma van Toetsing en Afsluiting (PTA) aan de inspectie toezenden en aan leerlingen verstrekken. Het PTA bevat gedetailleerde informatie over het SE-programma. Voor elk examenvak staat hierin informatie over welke onderdelen worden getoetst, de leerstof en inhoud per toets, herkansingsmogelijkheden, het gewicht van de toets voor het SE-cijfer, et cetera. Het examenreglement bevat regels rondom de organisatie van het eindexamen, zoals maatregelen die scholen nemen bij ziekte of onregelmatigheden. Voor vakken waarbij een leerling een CE maakt, is het cijfer van het eindexamen het gemiddelde van het SE-cijfer en het CE-cijfer, waarbij beide even zwaar meetellen. Bij enkele vakken, zoals maatschappijleer, wordt het vak alleen met een SE afgesloten. In dat geval is het eindcijfer geheel bepaald door het SE. De slaag-zakregeling bepaalt op basis van de eindcijfers of een leerling slaagt voor het eindexamen als geheel[10]. Opgenomen in de slaag-zakregeling is onder meer dat het ongewogen gemiddelde van alle centrale examencijfers (zonder afronding) minstens 5,5 moet zijn. Voor de kernvakken geldt aanvullende regelgeving: havo- en vwo-leerlingen mogen bijvoorbeeld maximaal één vijf hebben voor Nederlands, Engels en wiskunde; voor vmbo geldt een vergelijkbare eis ten aanzien van Nederlands.

1.2.2 De kwaliteit van het SE

Van der Ploeg en Weijers (2018[11], zie ook checklist 3, VO-raad[12]) onderscheiden vier perspectieven bij het onderzoeken van de kwaliteit van SE's en het proces eromheen; de schoolexaminering. Dat zijn (1) de verdeling en invulling van de verschillende verantwoordelijkheden (zoals die van de schoolleider, de examensecretaris, vakdocenten en examencommissie), en (2) de conceptuele, (3) procesmatige en (4) vakinhoudelijke en toetstechnische kwaliteit van het SE. De eerste drie punten betreffen vooral de inrichting en de borging van het examineringsproces, en vallen vooral onder de schoolexaminering. Dit is uitvoerig door de Inspectie van het Onderwijs onderzocht in de eerder benoemde inspectierapporten (Inspectie van het Onderwijs, 2023[13]). Het laatste punt, de vakinhoudelijke en toetstechnische kwaliteit van het SE, had daarbij minder aandacht en staat in dit onderzoek centraal. Hiervoor duiken we in de werkwijze waarop scholen de inhoud en vorm van de examens bepalen zodat zij met de SE's daadwerkelijk bij de leerlingen beoordelen wat ze willen toetsen. Het gaat er daarbij ook om of en hoe scholen ervoor zorgen dat het SE de onderwijsvisie van de school weerspiegelt en voldoende onderscheidend is van het CE.

Voor onderzoek naar vakinhoudelijke en toetstechnische kwaliteit van tests en toetsen zijn diverse kaders beschikbaar. De oudste daarvan is het beoordelingssysteem van de COTAN – de Commissie Testaangelegenheden Nederland. De COTAN is als commissie van het Nederlands Instituut van Psychologen (NIP) opgericht in 1959 met als missie om de kwaliteit van tests en testgebruik in Nederland

[10] Zie Uitvoeringsbesluit WVO 2020: https://wetten.overheid.nl/BWBR0045787/2025-10-01
[11] Van der Ploeg, S., & Weijers, S. (2018). Kwaliteitsborging schoolexaminering voortgezet onderwijs, Oberon.
[12] VO Raad (2023). Checklist 'De kwaliteit van schoolexaminering' (versie 2025-2026).
[13] Inspectie van het Onderwijs (2023). Technisch rapport themaonderzoek OP6. Kwaliteit van de schoolexaminering. Utrecht: Inspectie van het Onderwijs.

te bevorderen. Sinds 1969 beoordelen zij tests met hun beoordelingssysteem, waarbij de meest recente versie in 2009 gepubliceerd is (Evers, et al, 2009)[14].

Andere kaders waarmee in Nederland kwaliteit van tests en toetsen geëvalueerd worden, hebben vaak inspiratie geput uit het COTAN-beoordelingssysteem. Dit geldt voor wettelijke kaders zoals die voor de evaluatie van doorstroomtoetsen (Staatcourant, 2025)[15] en leerlingvolgsystemen (2026)[16], alsook voor het beoordelingskader van het Research Center voor Examinering en Certificering (RCEC). Ook Stichting Cito heeft kaders ontwikkeld voor het beoordelen van toetsen, waarbij er ook specifieke aandacht was voor SE's. Dit betreft onder andere een kader dat samen met de Inspectie van het Onderwijs is ontwikkeld voor eerder onderzoek om de vakinhoudelijke en toetstechnische kwaliteit van SE's in kaart te brengen (Kuhlemeier en Van Til, 2016; pp 64-69)[17]. Daarnaast zijn meer beoordelingen geweest van SE's door bijvoorbeeld Stichting Cito[18] en de Inspectie van het Onderwijs[19]. Ten slotte faciliteert de VO-raad scholen in de organisatie en ontwikkeling van de SE's. Zo zijn op de website van de VO-raad verschillende checklists, handreikingen en een zelfevaluatie-instrument te vinden, waaronder de checklist 'De kwaliteit van schoolexaminering'[20] die gedeeltelijk gericht is op de vakinhoudelijke en toetstechnische kwaliteit van SE's.

Wat in al deze bronnen terugkomt, is dat voor toetsen moet gelden dat deze ten minste valide en betrouwbaar zijn. Met andere woorden: het is noodzakelijk dat de toets meet wat het beoogt te meten en dat de beoordeling zoveel mogelijk vrij van meetfout is. Dit zijn zodoende ook criteria die van belang zijn voor het SE.

1.2.2.1 Validiteit

Er zijn veel verschillende opvattingen over hoe de validiteit van een toets in kaart zou moeten worden gebracht. Traditioneel wordt er onderscheid gemaakt tussen verschillende soorten validiteit. Twee bekende vormen van validiteit zijn begripsvaliditeit en inhoudsvaliditeit (Cronbach & Meehl, 1955)[21]. Begripsvaliditeit wil zeggen dat daadwerkelijk het begrip-zoals-bedoeld wordt gemeten en inhoudsvaliditeit wil zeggen dat de inhoud van opgaven representatief is voor de inhoud van het construct.

Hoewel deze vormen nog wel onderscheiden worden, is het tegenwoordig meer de opvatting dat de verschillende soorten validiteit niet al te strikt moeten worden geïnterpreteerd. In plaats van verschillende vormen van validiteit, ziet men het vaker als verschillende manieren om informatie over de validiteit te verzamelen (waarbij validiteit gezien wordt als een 'ondeelbaar begrip'; zie p. 38 van het COTAN beoordelingssysteem).

Deze visie op validiteit is nog verder doorgevoerd bij de nog recentere aanpak van de *argument-based approach to validation* (Kane, 2013[22]; in het Nederlands aangeduid en toegepast als bewijsgericht

---

[14] Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2009). COTAN Beoordelingssysteem voor de Kwaliteit van Tests (geheel herziene versie). Amsterdam: NIP.

[15] Staatcourant (2025). Regeling van het College voor Toetsen en Examens van 3 februari 2025, CvTE-25.00472, houdende wijziging van de Regeling beoordelingskader doorstroomtoets PO Staatscourant 2025, 7373 | Overheid.nl > Officiële bekendmakingen

[16] Staatcourant (2026). Regeling van het College voor toetsen en examens van 1 december 2025, nummer CvTE-25.012.12, houdende wijziging van de Regeling beoordelingskader doorstroomtoets PO in verband met de wijziging van de Regeling toetswijzer doorstroomtoets PO en enkele technische verbeteringen. Staatscourant 2026, 100 | Overheid.nl > Officiële bekendmakingen

[17] Kuhlemeier, H.,& Van Til, A. (2016). *Toetsen op School - Voortgezet Onderwijs*. Arnhem: Cito.

[18] Cito (2017). Screeningslijst toetskwaliteit.

[19] Inspectie van het Onderwijs (2023). Technisch rapport themaonderzoek OP6. Kwaliteit van de schoolexaminering. Utrecht: Inspectie van het Onderwijs.

[20] VO Raad (2023). Checklist 'De kwaliteit van schoolexaminering' (versie 2025-2026).

[21] Cronbach, L. J., & Meehl, P. E. (1955). Construct Validity in Psychological Tests. Psychological Bulletin, 52, 281-302.

[22] Kane, M. T. (2013). The argument-based approach to validation. *School Psychology Review*, *42*(4), 448-457. https://doi.org/10.1080/02796015.2013.12087465

valideren; Wools, 2015[23]). Deze aanpak onderscheidt ook geen verschillende vormen van validiteit. Het splitst het hele toetsproces op, van toetsconstructie via toetsafname tot beslissing, in allerlei onderliggende aannames. Door de onderbouwing van de aannames tezamen te evalueren, kan bepaald worden of het aannemelijk is dat de geconstrueerde toets valide is voor het beoogde doel bij de doelgroep. Validering van de toets houdt dan vervolgens in dat al deze onderliggende aannames worden onderbouwd en getoetst. Voorbeelden van dergelijke aannames zijn dat een gelijke prestatie op de toets leidt tot een gelijke score en dat gewenste vaardigheden worden gemeten (en ongewenste vaardigheden niet). De aanpak van het bewijsgericht validering maakt inzichtelijk welke zaken bij een specifieke toets nog een belemmering kunnen zijn voor validiteit. Het gedachtegoed van het bewijsgericht valideren helpt daarmee om zogenaamde 'validiteitsbedreigers' te achterhalen.

1.2.2.2 Betrouwbaarheid

De betrouwbaarheid van een toets gaat over de mate waarin het gelukt is om meetfouten te voorkomen. De invloed van toevallige factoren op de toetsresultaten moet zo veel mogelijk beperkt zijn. Je kan hierbij denken aan verschillende toevalsfactoren: toevallige keuzes bij selectie van opgaven, de invloed van een beoordelaar van de toets, maar ook toevallige factoren die invloed hebben op de geleverde prestatie. Bij dat laatste kan je denken aan dat iemand iets wel kan of weet, maar bijvoorbeeld minder presteert doordat deze slecht geslapen heeft. Dat kan je in je meting niet altijd voorkomen, maar het helpt dan mensen vaker te meten.

De invloed van de toevallige opgaven wordt vaak onderzocht met betrouwbaarheidsmaten gebaseerd op inter-itemrelaties, zoals Cronbach's alfa. Een manier om de invloed van een toevallige opgave te verminderen is door de toets uit meer opgaven te laten bestaan. Je moet er dan natuurlijk wel voor zorgen dat de test niet te lang wordt, omdat anders wellicht concentratie- dan wel uithoudingsvermogen te veel een rol kan spelen, terwijl je dat vaak niet wil meten.

Een belangrijke toevallige invloed die vermeden moet worden is die van de beoordelaar: het testresultaat moet uitsluitend door kenmerken van de geteste persoon worden bepaald, niet door wie beoordeelt. Wanneer twee verschillende beoordelaars onafhankelijk bij dezelfde prestatie (schriftelijk of praktisch) dezelfde scores geven, is het beoordelingsproces objectief, wat weerspiegeld wordt door een hoge interbeoordelaarsbetrouwbaarheid. Je kunt ook intra-beoordelaarsbetrouwbaarheid meten: een enkele beoordelaar moet bij dezelfde prestatie consequent dezelfde scores geven.

Beide vormen verbeteren aanzienlijk door een helder beoordelingsmodel en door training van beoordelaars/docenten (Sanders, 2013; pp. 62-64)[24]. Training vergroot ook het bewustzijn van vaak voorkomende, onbedoelde beoordelaarsfouten — zoals het halo-effect, waarbij een voor de beoordeling irrelevant positief aspect doorklinkt in het oordeel — en helpt deze zo veel mogelijk te vermijden. Duidelijke, praktische beoordelingscriteria maken consistent beoordelen makkelijker. Het model moet echter hanteerbaar blijven; een te complex maar volledig model kan juist onbetrouwbare resultaten opleveren (zie ook Kuhlemeier, 2017)[25].

1.2.2.3 Aanvullende kwaliteitseisen en specifieke kenmerken van SE's

Naast validiteit en betrouwbaarheid zijn er verschillende aanvullende kwaliteitseisen waaraan schoolexamens idealiter zouden moeten voldoen om tot een goed SE te komen[26]. Ten eerste moeten SE's voldoen aan de inhoudelijke specificaties uit het examenprogramma en aansluiten bij de eindtermen. Het

[23] Wools, S. (2015). *All about validity: An evaluation system for the quality of educational assessment [Doctoral Thesis]*.

[24] Sanders, P.F.(2013). Toetsen op School. Cito: Arnhem. Online: https://cito.nl/kennisbank-stichting-cito/toetsen-op-school/

[25] Kuhlemeier, H. (2017). Praktijktoetsen en praktische opdrachten: hoe ontwikkel je ze? Toetswijzer (Praktijktoetsen en praktische opdrachten: hoe ontwikkel je ze? Deel 2 van Serie over Praktijktoetsing).

[26] Dit zijn toetstechnische eisen waar de scholen zelf voor zorgen. Hoewel het wel zaak is dat scholen goede SE's maken, zijn er geen wettelijke vereiste omtrent de betrouwbaarheid, validiteit en normering van de SE's.

is daarbij van belang dat het SE niet op zichzelf staat, maar gezien moet worden in relatie tot het CE, zodat het eindexamen een voldoende breed beeld geeft van wat de leerling kent en kan bij het vak. Op de website www.examenblad.nl van de rijksoverheid is aangegeven wat vakinhoudelijk in het CE komt en wat in het SE opgenomen mag, dan wel moet, worden. Daarbij is dus een gedeeltelijke vakinhoudelijke overlap mogelijk tussen CE en SE. Er zijn ook specifieke vaardigheden (denk aan praktische vaardigheden) die echt in het SE horen, omdat deze lastiger centraal getoetst kunnen worden.

Ten tweede dient de weging van de verschillende toetsen passend te zijn. Deze weging moet gecommuniceerd zijn aan leerlingen, wat meestal via het PTA gebeurt.

Ten derde moet de normering van de SE's op orde zijn. Het is ongewenst als de leerlingen op scholen met makkelijkere toetsen en/of mildere beoordelingen bevoordeeld worden, ten opzichte van leerlingen die mogelijk moeilijkere toetsen krijgen en/of strenger beoordeeld worden op een andere school. Eerlijke kansen betekent in dit kader dat een eindexamenvak op de ene school behaald kan worden op basis van hetzelfde niveau van prestaties als bij een andere school.

Naast kwaliteitseisen kent het SE ook verschillende kenmerken. Een specifiek kenmerk van het SE is dat deze doorgaans bestaat uit meerdere toetsen, verspreid over meerdere jaren en toetsmomenten. Een laatste kenmerk van het SE is dat scholen de mogelijkheid hebben het schooleigen karakter (identiteit, schoolvisie, specialisatie) terug te laten komen in de examinering.

1.2.3 Actoren bij de SE's

Bij de SE's spelen verschillende actoren een rol. In de eerste plaats zijn dat de vakdocenten; naast het voorbereiden en begeleiden van hun leerlingen en het nakijken van het examenwerk zijn zij met hun vaksecties verantwoordelijk voor de samenstelling van de SE's. Het examenmateriaal (zoals toetsen, werkstukken, en praktische opdrachten) voor de SE's kan de docent geheel zelfstandig ontwikkelen, maar kan dat ook doen in samenwerking met anderen, vaak mede-docenten in de vaksectie. Het is ook mogelijk gebruik te maken van extern ontwikkeld examenmateriaal door educatieve uitgevers. Wat ook kan gebeuren, is dat scholen gebruik maken van toetsen of toetsopgaven uit methoden en opgaven uit oude CE's[27].

Een andere actor is de schooldirectie die het PTA opstelt in samenwerking met de vaksecties. Hiervoor wordt gebruik gemaakt van de kaders van de examencommissie; een andere belangrijke actor. Naast het bepalen van de kaders en richtlijnen voor het PTA is de examencommissie verantwoordelijk voor de kwaliteitsborging van de SE's binnen een school. Daarnaast spelen examensecretarissen binnen scholen een cruciale rol in de organisatie van zowel het SE als het CE. Dit is een uitvoerende rol waarbij zij ook voor het CvTE fungeren als de centrale contactpersoon rond de CE's. Op een deel van de scholen maken zij ook deel uit van de examencommissie.

Vakdocenten van de verschillende vakken en leergebieden zijn verenigd in een vakverenigingen waar vakdocenten lid van kunnen zijn. Vakverenigingen helpen onder andere bij het delen van kennis (door bijvoorbeeld artikelen en conferenties) en ondersteunen vakdocenten bij hun onderwijspraktijk. Op dit moment spelen deze vakverenigingen (nog) geen actieve rol in de vakinhoudelijke vormgeving van de SE's en het proces rondom deze SE's. De voor dit rapport geïnterviewde docenten gaven echter meermaals aan dat deze vakverenigingen deze rol mogelijk in de toekomst wel zouden kunnen vervullen.

[27] Inspectie van het Onderwijs (2019) Schoolexaminering in het voortgezet onderwijs: naar een zorgvuldig verloop en een betere kwaliteitsbewaking. Utrecht: Inspectie van het Onderwijs.

1.3 Onderzoeksvragen

De hoofdonderzoeksvraag van dit rapport luidt: *Wat is de vakinhoudelijke en toetstechnische kwaliteit van SE's in het voortgezet onderwijs?* Om deze vraag te beantwoorden onderzoeken we de kwaliteit van het SE, de factoren die hierop van invloed zijn en de wijze waarop deze kwaliteit kan worden verbeterd. De kwaliteit onderzoeken we door SE's te beoordelen: dit zijn vaak schriftelijke toetsen, maar kunnen ook praktische opdrachten zijn. De factoren die hierop van invloed zijn brengen we in kaart door docenten en examencommissieleden te bevragen over wat scholen doen om zo goed mogelijke SE's te realiseren. Het ontdekken van het "hoe" en "waarom" achter het ontwerp van een SE is belangrijk om tot aanwijzingen voor verbetering te komen.

Met deze opzet wordt uitvoering gegeven aan één van de aanbevelingen aan het ministerie van OCW uit het inspectierapport[28] en is het vergelijkbaar met en aanvullend op eerder onderzoek door Stichting Cito naar de kwaliteit van SE's (Erkens, et al., 2007[29]; Cito, 2008[30]; Kuhlemeier & Van Til, 2016[31]).

Aangezien het doel van het onderzoek ook is om handvatten te geven voor het versterken van de kwaliteit van de SE's, hebben we ook gebruik gemaakt van signalen van docenten tijdens het onderzoek. Daarbij kwam naar voren dat vakverenigingen mogelijk een rol zouden kunnen gaan spelen in het versterken van de kwaliteit (zie sectie 1.2.3). Om die reden zijn ook interviews gehouden met vakverenigingen. Met deze interviews wordt aanvullend antwoord gegeven op een extra, tweede onderzoeksvraag: *Welke potentiële, toekomstige rol zien vakverenigingen voor zich in de ondersteuning van de kwaliteit van SE's?* Deze onderzoeksvraag heeft een meer exploratief karakter en is beantwoord aan de hand van semigestructureerde interviews met bestuursleden van de verschillende vakverenigingen.

In het volgende hoofdstuk gaan we dieper in op de methoden die toegepast zijn om de onderzoeksvragen te beantwoorden. Daarin wordt ingegaan op de stappen die genomen zijn om tot de vragenlijsten en het beoordelingskader te komen en hoe de respondenten benaderd zijn. In hoofdstuk 3 volgen de resultaten van het onderzoek, en in het afsluitende hoofdstuk 4 volgen de discussie, conclusies en aanbevelingen.

[28] Inspectie van het Onderwijs (2023). Technisch rapport themaonderzoek OP6. Kwaliteit van de schoolexaminering. Utrecht: Inspectie van het Onderwijs.
[29] Erkens, T., Kuhlemeier, H., & Weerden, J. van (2007). Op weg naar betere schoolexamens: de kwaliteitsmonitor schoolexamens van Cito. Examens, 3, 5-9.
[30] Cito (2008). Het SE in het voortgezet onderwijs. Verslag van een onderzoek naar de kwaliteit van het SE bij de vakken Engels, Nederlands, biologie en wiskunde. Arnhem: Cito.
[31] Kuhlemeier, H. , & Van Til, A. (2016). *Toetsen op School: Voortgezet onderwijs*. Arnhem: Cito.

2. Methoden

2.1 Methodologische opzet

Voor het beantwoorden van de onderzoeksvragen hebben we drie deelonderzoeken uitgevoerd. De eerste twee deelonderzoeken geven antwoord op de hoofdvraag. Het eerste deelonderzoek naar hoe scholen het proces van SE's inrichten, bestaat uit vragenlijstonderzoek onder vakdocenten en leden van examencommissies. Voor het tweede deelonderzoek is de toetstechnische kwaliteit van ingestuurde SE's beoordeeld door toetsdeskundigen aan de hand van een ontwikkeld beoordelingskader. Voor het derde deelonderzoek zijn interviews afgenomen met bestuursleden van vakverenigingen.

Voorafgaand aan het uitvoeren van de eerste twee deelonderzoeken is toestemming verkregen van de ethische commissie van Stichting Cito. Zoals beschreven is het derde deelonderzoek in de loop van het onderzoek ontstaan. In dit onderzoek heeft elke respondent mondeling toestemming gegeven voor deelname aan het onderzoek.

2.2 Vragenlijsten docenten en examencommissie

2.2.1 Doel van het onderzoek

Met dit deelonderzoek wordt inzicht verkregen in het proces van het inrichten van SE's, om inzicht te krijgen op de kwaliteit hiervan. Dit onderzoek is ten eerste uitgevoerd bij vakdocenten; zij zijn verantwoordelijk voor het samenstellen van de SE's. Ten tweede is een vragenlijst uitgezet bij de examencommissieleden; zij zijn verantwoordelijk voor de kwaliteitsborging van de SE's (zie sectie 1.2.3).

Op basis van deze vragenlijsten beogen we antwoord te krijgen op vragen zoals:

- Hoe worden toetsen en examenopdrachten inhoudelijk vormgegeven en geëvalueerd door scholen?
- Welke keuzes maken scholen in de samenstelling en inhoud van hun SE's?
- In hoeverre is er sprake van inhoudelijke en toetstechnische kwaliteitsborging bij de ontwikkeling van SE-toetsen?

Vragenlijsten zijn gekozen als onderzoeksmethode omdat hiermee zichtbaar wordt gemaakt welke stappen en motieven achter keuzes zitten, welke acties docenten ondernemen bij het opstellen van SE's en in welke mate zij zich daarbij bekwaam voelen. Bovendien zijn vergelijkbare vragen ook aan leden van examencommissies voorgelegd, zodat de bevindingen vanuit beide perspectieven (docenten en examencommissies) geanalyseerd kunnen worden. Deze informatie is essentieel om te bepalen waar ondersteuning en professionalisering kan bijdragen aan betere SE's en op welke punten docenten versterking nodig hebben. Dit is aanvullend op het beoordelen van de uiteindelijke SE's, zoals is beschreven in *sectie 2.3 Evaluatie SE's*.

2.2.2 Ontwikkeling van de vragenlijsten

De volgende stappen zijn gevolgd bij de ontwikkeling van de vragenlijsten:

2.2.2.1 Bepalen van relevantie en thema's

Allereerst is vastgesteld welke onderwerpen opgenomen moesten worden. Daarbij is onderscheid gemaakt tussen thema's die direct betrekking hebben op vakinhoudelijke en toetstechnische kwaliteit (bijvoorbeeld validiteit, betrouwbaarheid, afstemming op kerndoelen) en onderwerpen die helpen om deze kwaliteit van SE's in een wat breder perspectief te plaatsen.

Voor de thema's rond de kwaliteit is onder meer inspiratie gehaald uit de materialen voor het beoordelen van tests, zoals het eerder genoemde COTAN beoordelingssysteem (Evers et al., 2010)[32], praktijkrichtlijnen zoals die van de VO-raad (Kwaliteitsborging schoolexaminering)[33], en het onderzoek door de Inspectie van het Onderwijs (2023)[34]. Op basis hiervan is een lijst van onderwerpen gemaakt die aan bod dienden te komen in de vragenlijst. Hierbij is gekeken naar kenmerken van goede SE's, en welke stappen nodig zijn om die samen te stellen, welke hulpmiddelen daarbij gebruikt worden, naar welke mate van samenwerking en ervaren kennis van toetsconstructie er is bij mensen op school die de SE's produceren en evalueren.

2.2.2.2 Afstemming op doelgroepen

Vervolgens is bekeken hoe de geselecteerde onderwerpen relevant zijn voor de twee beoogde doelgroepen: (a) docenten die SE's opstellen en uitvoeren, en (b) leden van examencommissies. Per doelgroep is bepaald welke invalshoeken en woordkeuzes het meest geschikt zijn om relevante en bruikbare antwoorden te verkrijgen. Er is gekeken welke thema's beter bij de ene groep dan wel de andere groep pasten, en welke nuanceverschillen er zijn om de thema's te benaderen voor de verschillende groepen.

2.2.2.3 Eerste opzet van de vragen

Na de uiteenzetting van de thema's zijn de vragen voor de vragenlijst opgesteld, waarbij gelet is op hoe passend deze zijn voor de ondervraagde personen. Zijn het vragen die zij kunnen beantwoorden, en welk taalgebruik past daarbij? Bij de formulering van vragen is inspiratie gehaald uit de teksten van de VO-Raad die ook bij het opstellen van de relevante thema's gebruikt zijn. De vragen zijn geschreven vanuit de blik van de beoogde ondervraagde persoon, omdat het bij diens ervaringen moest aansluiten: welke vragen zouden in die context relevant zijn om vervolgens rechtstreeks aan de docent te stellen?

2.2.2.4 Opstellen en thematische ordening van vragen

Op basis van bovenstaande bronnen is een eerste lijst met vragen opgesteld en thematisch geordend. Dit maakte het mogelijk om samenhang tussen vragen te waarborgen en dubbele of overlappende items vroegtijdig te signaleren. Hier vond een eerste selectie plaats, door de dubbele vragen te verwijderen.

Om de beoogde respondenten niet te ontmoedigen, diende de lijst ook niet te lang te worden. Voor een extra selectie is een beoordeling gemaakt van de relevantie van de vragen. Vragen zijn verwijderd, waarvan het oordeel was dat het mogelijk interessant kon zijn, maar niet cruciaal voor het beantwoorden

[32] Evers, A., Lucassen, W., Meijer, R. R., & Sijtsma, K. (2009). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. Nederlands Instituut van Psychologen/Commissie Testaangelegenheden Nederland.
[33] VO-raad (2023). Checklist 3 'De kwaliteit van schoolexaminering' (editie 2025-2026). Geraadpleegd van https://www.vo-raad.nl/onderwerpen/examinering/praktijk-ondersteuning
[34] Inspectie van het Onderwijs (2023). Technisch rapport themaonderzoek OP6. Kwaliteit van de schoolexaminering. Utrecht: Inspectie van het Onderwijs. Geraadpleegd van https://www.onderwijsinspectie.nl/documenten/rapporten/2023/10/12/kwaliteit-van-de-schoolexaminering

van de onderzoeksvraag. Vragen waar nog enige onzekerheid over de relevantie over was, waren in deze fase nog behouden.

2.2.2.5 Toetsing door praktijkcollega's en externen

De conceptlijst is voorgelegd aan een reeks collega's met ervarenheid als docent en/of lid van examencommissies in het voortgezet onderwijs. Zij beoordeelden per vraag de urgentie en relevantie, de herkenbaarheid van het thema, en de bewoording: waren de antwoordalternatieven logisch en helder, of moesten formuleringen worden aangepast? Bij de selectie van de praktijkcollega's is er zorg voor gedragen dat er variatie was in ervaring met de verschillende leerwegen en vakken.

De toetsing door praktijkcollega's is in twee ronden gebeurd. In iedere ronde kregen zij de vragen, gaven hun commentaar, en bespraken die met de onderzoekers. Na iedere ronde werden de vragen verbeterd. De interne evaluatie richtte zich op bewoording en volgorde, mogelijke dubbelingen, de mate waarin de vragen de onderzoeksvragen kunnen beantwoorden en de vraag of de toetstechnische kwaliteit voldoende aan bod komt.

Tot slot zijn ook externe partijen betrokken bij de laatste toetsing, onder meer OCW, VO-raad, het landelijk platform voor examensecretarissen en examencommissieleden (PLEXS) en SLO. Hun feedback heeft bijgedragen aan de afronding en legitimatie van de vragenlijsten.

2.2.2.6 Technische uitwerking in Microsoft Forms

Op basis van de terugkoppeling in de tweede ronde zijn aanpassingen doorgevoerd en is de vragenlijst technisch vormgegeven in Microsoft Forms. Hierbij is ook aandacht besteed aan de volgorde van vragen en de presentatie, zodat de responsbeleving zo optimaal mogelijk is. Deze vorm is andermaal kritisch doorgenomen, om de laatste onvolkomenheden weg te werken.

2.2.2.7 Definitieve vaststelling

De twee vragenlijsten zijn op 22 januari 2025 definitief vastgesteld. Deze versies zijn gedeeld met de respondenten.

Met deze systematische ontwikkelstappen is beoogd vragenlijsten te ontwerpen die enerzijds robuuste informatie opleveren over de praktijk van SE's en anderzijds concrete aanknopingspunten bieden voor versterking van de deskundigheid van docenten en de kwaliteitsbewaking door examencommissies.

2.2.3 Kenmerken vragenlijsten

2.2.3.1 Vragenlijst docenten

De vragenlijst voor docenten bevat in totaal 36 vragen. De eerste acht vragen betroffen goedkeuring voor deelname, en achtergrond informatie. De overige 27 vragen daaropvolgend betroffen de kwaliteit van de SE's, en zijn opgedeeld onder de thema's constructie, regels en richtlijnen, leerlingen, beoordelen, evaluatie en deskundigheid en algemene vragen. De meeste vragen hadden de vorm dat de respondent een enkel antwoord, dan wel meer antwoorden kon aanklikken. In vraag 32 is gevraagd naar ervaren deskundigheid op zeven onderwerpen, waardoor deze vraag uit zeven subvragen bestond. Vraag 35 was een open vraag. In de laatste vraag werd gevraagd of de respondenten hun SE's met ons wilde delen. De gehele vragenlijst is te vinden in Bijlage 2.

2.2.3.2 Vragenlijst examencommissies

De vragenlijst voor leden van de examencommissie bevat 32 vragen. Ook deze is opgebouwd uit de volgende onderdelen: achtergrondinformatie, constructie, regels en richtlijnen, beoordelen,

deskundigheid en algemene vragen. Hoewel de kopjes van de onderdelen grotendeels overeenkomen, verschilt deze vragenlijst inhoudelijk (deels) van de vragenlijst voor docenten. De vragenlijst voor leden van de examencommissie is te vinden in Bijlage 3.

2.2.3.3 Toestemming

Docenten en examen-commissieleden die de vragenlijst invulden kregen de vraag (1) of zij kennisgenomen hadden van het doel van het onderzoek, (2) of zij begrepen op welke wijze de gegevens werden verzameld, opgeslagen, en gebruikt voor onderzoeksdoeleinden en (3) of zij toestemming gaven om hun gegevens te gebruiken voor dit onderzoek. Bij deze vraag was een link gegeven naar de projectpagina op de website[35], naar de Toestemmingsverklaring en de Informatiebrief[36]. Eveneens waren er contactgegevens beschikbaar in het geval er vragen zouden zijn. Indien een respondent niet akkoord ging, werden de vragen niet voorgelegd en werden er geen gegevens verzameld.

2.2.4 Werving

Voor de werving van respondenten voor de vragenlijsten onder docenten en examencommissies is een aantal kanalen ingezet. Allereerst is een oproep geplaatst in de nieuwsbrieven van OCW, de VO-raad, PLEXS en het CvTE. Ook zijn oproepen geplaatst op CitoLab's LinkedIn pagina, en zijn oproepen gedeeld door medewerkers, ook via OCW.

Naast deze werving, heeft er ook een werving plaatsgevonden door congressen te bezoeken waar personen aanwezig waren die tot de onderzoekspopulatie behoren: docenten die SE's maken en examencommissieleden. Dit betrof het SPV Examenfestival[37], de algemene ledenvergadering ALV en het PLEXS- en het VO-congres[38]. Deze persoonlijke vorm van werking is ook toegepast door persoonlijke contacten van Cito-medewerkers die tot de onderzoekspopulatie horen te vragen om deel te nemen aan de vragenlijsten.

Een andere route die is toegepast, is het bereiken van respondenten via de mailinglijst van Cito's Bureau Externe Medewerkers. Tot slot is via de vakverenigingen gevraagd of zij docenten konden aanschrijven.

Er is bij deze werving gekozen voor een wervingsmethode die inzet op een maximaal aantal vrijwillige deelnamen. Voor het invullen van de vragenlijst was er geen mogelijkheid voor een beloning voor de deelname. Er was zo ook geen gebruik gemaakt van een steekproefkader: in principe viel de gehele populatie in de steekproef. Hoewel een groot aantal responses werd nagestreefd, kon ondanks de inzet van diverse wervingsmethoden een hoge opkomst niet worden gegarandeerd, omdat vrijwillige deelname als randvoorwaarde gold. Dat heeft ook tot gevolg dat er sprake zal zijn van zelfselectie, waarmee ook de representativiteit niet geborgd kon worden.

2.2.5 Analyses

Bij de analyses van de vragenlijsten zijn opvallende percentages tussen antwoorden en categorieën vergeleken. Per categorie (zoals schoolexamenvak, of schooltype) waren er niet genoeg respondenten om

[35] CitoLab, *Draag bij aan betere schoolexamens*.
[36] CitoLab, Toestemmingsverklaring – Onderzoek Kwaliteit Schoolexamens.
[37] SPV Examenfestival (10 & 11 maart 2025): congres georganiseerd door Stichting Platforms vmbo in Nieuwegein, waarbij er specifiek aandacht wordt gegeven aan de aankomende cspe's en er workshops zijn voor docenten over schoolexaminering.
[38] VO-Congres (27 maart 2025): congres georganiseerd door de VO-Raad in Nieuwegein. Het thema is dit jaar digitalisering, en de doelgroep is vooral schoolbestuurders en schoolleiders. Hoewel dit niet direct de doelgroep was kunnen zij zorgen voor grote aantallen responses bij de doelgroep.

uitgebreide statistische analyses uit te voeren. De resultaten zijn daarom beschreven, maar niet statistisch getoetst. De aanvullende antwoorden van respondenten zijn handmatig gecodeerd en vervolgens per code inhoudelijk beschreven. Vragen met zeer veel eigen antwoorden zijn automatisch samengevat en de belangrijkste punten zijn eveneens in tekst beschreven.

2.3 Evaluatie SE's

2.3.1 Relevante materialen behorend bij SE's

Een SE bestaat niet uit een enkele toets, maar vaak uit een verzameling aan toetsen en (praktische) opdrachten. Ook een enkele toets of opdracht staat niet op zichzelf. Naast de SE-toetsen zelf (zie o.a. Kuhlemeier en Van Til, 2016, pp.62-70)[39], wordt er ook gekeken naar eventuele correctievoorschriften, beoordelingsmodellen, toetsmatrijzen en de visie van de school.

In een *correctievoorschrift* staan onder meer algemene en vakspecifieke regels voor de beoordeling. Oftewel: het bevat de richtlijnen die de docent toepast bij het nakijken van het examenwerk. Een voorbeeld: op basis van het correctievoorschrift is duidelijk op welke aspecten de docent/beoordelaar wél en niet moet focussen gedurende de gehele beoordeling. Het correctievoorschrift bevat ook het *beoordelingsmodel* of de beoordelingsschaal per opgave of vraag. Een dergelijk model of een dergelijke schaal laat zien hoe punten (scores) toegekend worden aan zwakke, middelmatige en goede prestaties op elk van de afzonderlijke opgaven of vragen. Scholen zouden ongeveer even streng moeten zijn bij deze puntentoekenning om te voorkomen dat leerlingen met een vergelijkbaar prestatieniveau verschillende cijfers krijgen afhankelijk van de school. Hoe uitgebreider en preciezer het beoordelingsmodel of de beoordelingsschaal is, des te meer houvast heeft de docent bij het beoordelen. Dat voorkomt verschillen in soepelheid bij de beoordeling en voorkomt ook zoveel mogelijk verschillende eigen interpretaties. Een voorbeeld van een beoordelingsschaal is te zien in Figuur 2.3.1.

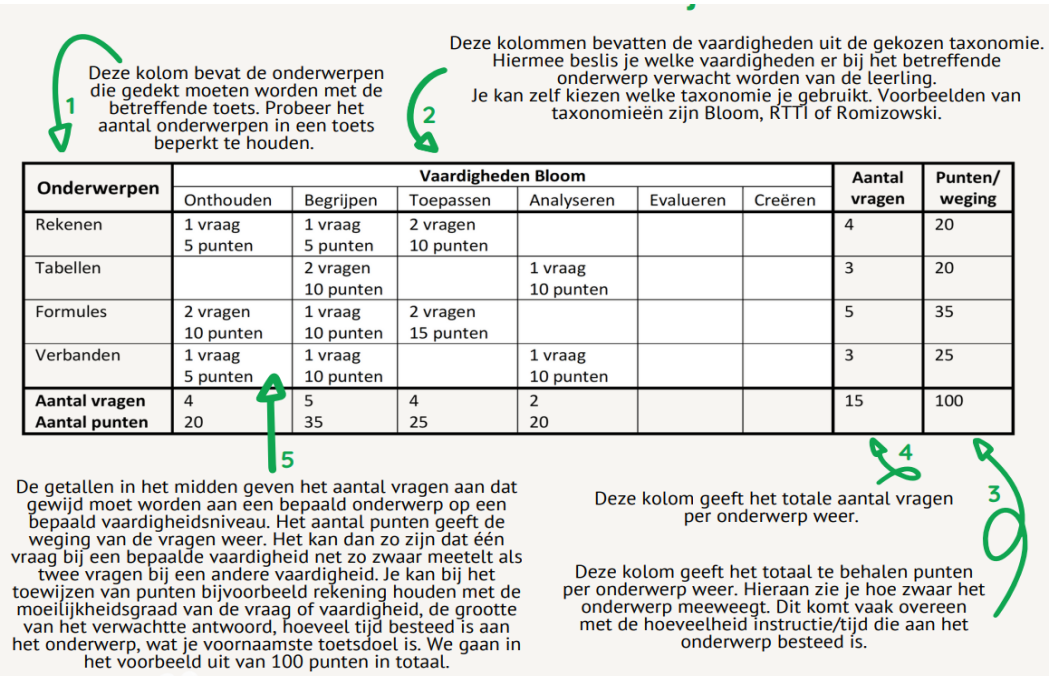
Figuur 2.3.1 *Voorbeeld van een generieke beoordelingsschaal. Uit: Toetsen op school, Cito, p.134*

Schriftelijk presenteren	**Score**
Communicatieve strekking	
De tekst als geheel is volledig begrijpelijk, de boodschap komt helemaal over	**2**
Boodschap komt niet helemaal over	**1**
Boodschap komt niet over	**0**
Briefconventies: aanhef en afsluiting	
Zowel aanhef als afsluiting zijn passend	**2**
Of aanhef of afsluiting is niet passend	**1**
Noch aanhef noch afsluiting is passend	**0**
Indeling tekst	
Tekstindeling bestaat duidelijk uit een inleiding, middendeel en slot	**2**
Tekstindeling is niet geheel duidelijk	**1**
Tekstindeling is niet duidelijk	**0**
Interpunctie, spelling en grammatica	
Geen enkele fout	**2**
Eén of twee fouten	**1**
Drie of meer fouten	**0**

[39] Kuhlemeier, H. & Van Til, A. (2016). Toetsen op school Voortgezet Onderwijs. Arnhem: Cito (online: https://cito.nl/media/qx0knak2/toetsen-op-school-voortgezet-onderwijs.pdf)

Een *toetsmatrijs* biedt een schematisch overzicht van de onderwerpen die gedekt worden in een toets samen met welke type vaardigheden worden getoetst, vaak geordend op basis van een taxonomie zoals die van Bloom[40]. Idealiter wordt een dergelijke toetsmatrijs opgesteld voordat de toets wordt samengesteld, om ervoor te zorgen dat alle onderwerpen en vaardigheden voldoende aan bod komen. Het biedt ook de mogelijkheid van te voren na te denken over de weging van verschillende opgaven. Figuur 2.3.1 toont een voorbeeld van een toetsmatrijs.

Figuur 2.3.2 Voorbeeld van een toetsmatrijs. Uit: Toetsmatrijs in 5 stappen, CitoLab



Onderwerpen	Vaardigheden Bloom						Aantal vragen	Punten/ weging
	Onthouden	Begrijpen	Toepassen	Analyseren	Evalueren	Creëren		
Rekenen	1 vraag 5 punten	1 vraag 5 punten	2 vragen 10 punten				4	20
Tabellen		2 vragen 10 punten		1 vraag 10 punten			3	20
Formules	2 vragen 10 punten	1 vraag 10 punten	2 vragen 15 punten				5	35
Verbanden	1 vraag 5 punten	1 vraag 10 punten		1 vraag 10 punten			3	25
Aantal vragen **Aantal punten**	4 20	5 35	4 25	2 20			15	100

Ten slotte is gevraagd naar de *visie van de school*. Het wel of niet hebben van een visie maakt de toets op zichzelf niet beter. Het maakt wel het SE als geheel beter omdat het een betere afspiegeling is van het kennen en kunnen van een leerling. Een eigen visie op toetsen kan er ook beter toe leiden dat het SE ook veel meer is dan een voorbereiding op het CE. Het afsluitend karakter van het SE is daarmee eerder gegarandeerd. Idealiter is het SE niet simpelweg een verzameling van enkele losse toetsen, maar is deze samengesteld op basis van een onderliggende, schoolbrede visie op toetsing. Deze visie zou samen moeten hangen met de onderwijskundige visie van de school. Op basis van hun visie kiezen scholen vervolgens hoe vaak ze toetsen, wat ze toetsen, met welke toetsvormen, met welk doel en hoe de kwaliteit wordt geborgd. De visie is opgenomen in het schoolplan of het strategisch beleidsplan[41]. De VO-raad[42] heeft in 2019 een adviesrapport opgesteld met als titel 'De ontwikkeling van een visie op toetsing en schoolexaminering in het VO' die scholen kan helpen een dergelijke schoolbrede visie op toetsing en examinering op te stellen. Hier staat o.a. een stappenplan in en een format voor het voeren van interne gesprekken over de schoolvisie.

[40] Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
[41] SLO (2020). *Visie op toetsing en examinering*.
[42] VO-raad (2019). De ontwikkeling van een visie op toetsing en schoolexaminering in het VO. Utrecht: VO-raad.

2.3.2 Ontwikkeling van beoordelingskader

De beoordelingen van de afzonderlijke SE's zijn uitgevoerd met behulp van een beoordelingskader. Dit beoordelingskader bestaat uit vier delen, met sub-thema's indien van toepassing (zie Tabel 2.3.2). Het beoordelingskader is opgesteld door toetsexperts van stichting Cito aan de hand van de bestaande kaders die toegelicht zijn in sectie 1.2.2. Dit zijn beoordelingssystemen die veel gebruikt worden (bijvoorbeeld van COTAN[43]), of specifiek eerder gebruikt zijn voor onderzoek naar SE's (Cito, 2008[44]; Cito, 2024[45]). Conform de conceptuele benadering van bewijsgericht valideren (Bijlage 1), is daarnaast gezocht naar een set van potentiële validiteitsbedreigers die een rol kunnen spelen in alle SE's. Hierbij is vooral gefocust op de SE's zelf, maar is soms, om deze in perspectief te plaatsen, ook gekeken naar de docenten die de toetsen maken en controlemechanismes in het proces van examinering (PTA's, bijvoorbeeld). Enkele voorbeelden van vragen die op deze manier aan het beoordelingskader zijn toegevoegd zijn bijvoorbeeld: 'Is de weging tussen de opgaven passend en niet vertekenend voor het resultaat?' en 'Is de puntentoekenning transparant en logisch verdeeld?' Het beoordelingskader is terug te vinden in Bijlage 6.[46]

Tabel 2.3.2 Beoordelingskader voor het beoordelen van de aangeleverde SE's

Onderdeel	Sub-thema
1. Inventarisatie van de gedeelde informatie	a. Aanwezigheid van materiaal b. Aanvullende aangeleverde informatie per SE
2. Inhoudelijke beoordeling	a. Vorm, inhoud, dekking en niveau b. Transparantie en standaardisatie c. Vraagformulering & taalgebruik d. Beoordeling: antwoord- of beoordelingsmodel en beoordelingsprocedure e. Toetsmatrijs f. Validiteit & gelijke kansen g. Scoren en normeren h. Praktijkopdrachten
3. Relatie tussen CE en SE	n.v.t.
4. Open vragen	n.v.t.

Het eerste deel van het beoordelingskader betreft een inventarisatie van de gedeelde informatie. Dit omvat allereerst een inventarisatie van **1a. de aangeleverde materialen**, zoals de aanwezigheid van een beoordelingsmodel, en ten tweede **1b. aanvullende aangeleverde informatie** per SE, zoals of de tijdsduur van de toets duidelijk vermeld staat.

Na deze inventarisatie volgt de inhoudelijke beoordeling. De beoordeling is ingedeeld op een achttal thema's. Per thema wordt een aantal dichotome (ja-nee) vragen gesteld om te evalueren in welke mate het beoordeelde SE aan de eisen voldoet. De thema's en vragen zijn gebaseerd op kenmerken die

[43] Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2009). COTAN Beoordelingssysteem voor de Kwaliteit van Tests (geheel herziene versie). Amsterdam: NIP.
[44] Cito (2008). Het SE in het voortgezet onderwijs. Verslag van een onderzoek naar de kwaliteit van het SE bij de vakken Engels, Nederlands, biologie en wiskunde. Arnhem: Cito.
[45] Cito (2024). "Screeningslijst toetskwaliteit VO". Intern document.
[46] De indeling van het beoordelingsmodel in Bijlage 6 correspondeert niet volledig met de gebruikte indeling in het Resultaten hoofdstuk, waarbij een verwerkingsslag is gedaan om tot 4 hoofdcategorieën te komen. Ter verduidelijking: Onderdeel 1 (Inventarisatie van de gedeelde informatie) in het rapport correspondeert met vraag 1 en 2 in de bijlage. Onderdeel 2 (Inhoudelijke beoordeling) zijn vragen 3-10. Onderdeel 3 (Relatie tussen CE en SE) is vraag 11, en onderdeel 4 (Open vragen) is vraag 12. Hierbij heeft elke vraag weer een aantal subvragen (steeds aangegeven in de Bijlage met a-s), variërend van minimaal 3 tot maximaal 17.

relevant geacht worden voor goede SE's en genoemd worden door COTAN (2010)[47], Kuhlemeier & Van Til (2016)[48], de VO-raad (2023)[49] en eerdere beoordelingen van SE's (Cito 2017[50]; Inspectie van het Onderwijs, 2023[51]).

Het thema "**2a. Vorm, inhoud, dekking en niveau**" heeft te maken met de ingezette opgaven in het SE. Het is van belang dat deze wat betreft de vorm duidelijk zijn, en dat de inhoud passend is, zowel wat betreft de beoogde dekking van onderwerpen als wat betreft het niveau van de bevraging. Het bevraagde niveau moet namelijk passend zijn bij de doelen van het onderwijstype. Bij het thema "**2b. Transparantie en standaardisatie**" is de essentie dat leerlingen in de voorbereiding en tijdens de afname onder gelijke omstandigheden de SE's maken. Dit betekent onder andere dat aan alle leerlingen in de voorbereiding duidelijk gemaakt wordt wat van hen verwacht wordt bij het SE. De standaardisatie van de afname moet ervoor zorgen dat toevallige factoren, zoals willekeur in geboden tijd om het examen te maken, geen invloed hebben op het resultaat. Bij de evaluatie van de "**2c. Vraagformulering & taalgebruik**" wordt geëvalueerd in welke mate er onduidelijkheden in de vraagstelling gevonden worden, en het taalgebruik van de vragen passend is bij het schooltype waar het SE voor bedoeld is. Bij het evalueren van de "**2d. Beoordelingsmodel en -procedure"** hebben de inhoudsdeskundigen gekeken naar het antwoord- of beoordelingsmodel: is het duidelijk welk antwoord welke punten oplevert? En is in het geval van een praktische opdracht duidelijk welk type respons welk puntenaantal oplevert? Dit moet ervoor zorgen dat alle leerlingen op dezelfde eerlijke manier beoordeeld worden. Bij de beoordeling wordt ook gekeken of de beoordelingsprocedure zo is ingericht dat er een kwaliteitscontrole is op het toegepaste beoordelingsmodel. Wanneer er een "**2e. Toetsmatrijs"** gedeeld is, wordt gekeken in hoeverre de opgaven en wegingen in de toets zijn terug te voeren op de matrijs en of deze aansluiten op het PTA en examenprogramma. Bij **"2f. Validiteit & gelijke kansen**" draait het erom of voldoende verschillende vaardigheden of competenties aan bod komen en deze inclusief genoeg gemeten worden. Dat betekent dat de toetsvragen vrij van bias (vooroordeel/vertekening) moeten zijn, en mogelijke persoonlijke belemmeringen die niet met de gemeten vaardigheid te maken hebben een optimale prestatie niet in de weg staan. Wanneer er voldoende informatie over beschikbaar was, beoordeelden de experts bij het thema "**2g. Scoren en normeren**" wat de kwaliteit was van de stap van toetsscore naar een cijfer, of ander oordeel over de geleverde prestatie, op het SE. In het geval van "**2h. Praktijkopdrachten"** is ook gekeken naar specifieke kwaliteitskenmerken die passend zijn bij praktische opdrachten. Dit zijn vaak activiteiten die een leerlingen ter plekke op school uitvoeren en dan beoordeeld worden, zoals spreekbeurten of opdrachten representatief zijn voor een realistische beroeps- of leersituatie. Het kunnen echter ook zaken waar een leerling over een langere periode (thuis of op school) aan werkt, zoals een werkstuk.

Het aantal subvragen verschilde per thema van 3 tot en met 17 (zie Bijlage 6 voor deze details). Niet alle thema's kunnen bij alle SE's beoordeeld worden. Zo is het laatste thema specifiek gericht op praktijkopdrachten, en wordt deze zodoende niet bij schriftelijke toetsen beoordeeld. De beoordeling van een deel van de andere thema's is afhankelijk van de geleverde materialen. Zo kan het thema "Toetsmatrijs" alleen beoordeeld worden als er ook een toetsmatrijs gedeeld is.

Het derde onderdeel uit het beoordelingskader betrof de relatie tussen het SE en het CE van hetzelfde examenvak. Alhoewel het huidige rapport zich louter richt op de kwaliteit van het SE (en niet van het CE), vormen het SE en CE samen het eindexamen waardoor we de relatie SE-CE niet konden

[47] COTAN (2010). *Beoordelingssysteem voor de kwaliteit van tests.* Nederlands Instituut van Psychologen. Geraadpleegd van https://nip.nl/wp-content/uploads/2021/11/COTAN-Beoordelingssysteem-2010.pdf

[48] Kuhlemeier & Van Til (2016). *Toetsen op School. Voortgezet onderwijs.* Arnhem: Cito. Geraadpleegd van https://cito.nl/media/qx0knak2/toetsen-op-school-voortgezet-onderwijs.pdf

[49] VO-raad (2023). *Checklist 3 'De kwaliteit van schoolexaminering' (editie 2025-2026).* Geraadpleegd van https://www.vo-raad.nl/onderwerpen/examinering/praktijk-ondersteuning

[50] Cito (2017). Screeningslijst toetskwaliteit. Geraadpleegd van screeningslijst.cito.nl/question/form/1

[51] Inspectie van het Onderwijs (2023). Technisch rapport themaonderzoek OP6. Kwaliteit van de schoolexaminering. Utrecht: Inspectie van het Onderwijs. Geraadpleegd van https://www.onderwijsinspectie.nl/documenten/rapporten/2023/10/12/kwaliteit-van-de-schoolexaminering

negeren. Hierbij is onder andere geteld hoeveel opgaven op welke manier aan het CE te relateren zijn, bijvoorbeeld door oude CE-opgaven te gebruiken in het SE. Vanuit een toetstechnisch perspectief is het nuttig om zo een beeld te krijgen van de herkomst van SE-opgaven. Vanuit vakinhoudelijk perspectief is het interessant zo te weten te komen welke kennis en vaardigheden worden gemeten die niet in het CE zitten, en op een manier die niet in het CE gebeurt. Na het invullen van de acht thema's en het bekijken van de relatie tussen het SE en het CE voor alle ingediende SE's afzonderlijk, zijn er ook een vijftal open vragen ingevuld voor het gehele pakket aan SE's per school, vak en schooltype, door de inhoudelijke toetsexperts. Wanneer de docent enkel een afzonderlijk SE heeft ingediend, zijn dezelfde vragen voor deze enkele SE ingevuld. De vijftal vragen stelden de inhoudsexperts in de gelegenheid meer algemene observaties over het pakket als geheel te documenteren. De vragen luiden als volgt:

1. Geeft het SE de indruk dat er een eigen onderwijsvisie achter ligt? Waarop baseer je deze indruk?
2. Kan je een toelichting geven op je antwoorden op herkomst, opbouw en kwaliteit van de vragen in het SE?
3. Wat is er goed gegaan in dit SE of deze verzameling van SE's?
4. Wat zijn de belangrijkste verbeterpunten?
5. Welke tip of tips zou je de samensteller(s) van het SE willen meegeven?

2.3.3 Beoordelaars

De beoordelingen zijn uitgevoerd door zowel het onderzoeksteam als door inhoudsexperts. De focus van het onderzoeksteam lag op de inventarisatievragen. De betrokken onderzoekers vulden de inventarisatievragen in het beoordelingskader in, alvorens deze te delen met de inhoudsexperts. De inhoudsexperts richtten zich vervolgens op de inhoudelijke vragen aangaande de afzonderlijke SE's en de pakketten als geheel. Bij iedere dichotome beoordeling was de beoordelaar ook in de gelegenheid om een kwalitatieve toelichting te geven.

De inhoudsexperts zijn allen medewerkers van de afdeling Centrale Toetsen en Examens (CTE) van Stichting Cito. De experts zijn geselecteerd op basis van hun kennis en ervaring met het te beoordelen vak en het schooltype. De experts zijn in de meeste gevallen verantwoordelijk, of eerder verantwoordelijk geweest, voor het CE van dat vak voor dat schooltype. Zij kennen hierdoor de eisen van het vak goed, en weten zodoende wat in de SE's zou kunnen dan wel moeten zitten. Velen van hen hebben eerder ook als docent gewerkt en in het verleden zelf SE's gemaakt.

2.3.4 Werving

De SE's zijn verkregen via de vragenlijsten. Aan docenten is gevraagd of zij hun SE's met de onderzoekers wilden delen. Er is gevraagd naar de gehele verzameling van alle SE's die op de school in het PTA genoemd werden voor een vak en schooltype, met bijbehorende beoordelingsmodellen en correctievoorschriften, toetsmatrijzen indien deze gebruikt worden en een eventuele visie op toetsen van de school. Deze verschillende elementen worden toegelicht in sectie 2.3.1. Zowel de afzonderlijke SE's als het ingestuurde pakket als geheel (de verzameling SE's per school, per vak, per schooltype) zijn beoordeeld door zowel onderzoekers als inhoudsexperts. Net als bij de vragenlijsten was deelname geheel vrijwillig. Met het insturen van een of meer toetsen en opdrachten die deel uitmaakten van een SE, werd toestemming verleend de resultaten te gebruiken. Aangegeven was dat in de rapportage de resultaten van de beoordeling van de SE's, dit niet met verwijzing naar de school of de docent(en) zou zijn, net zoals bij de vragenlijsten.

2.3.5 Terugkoppeling aan docenten

Als tegenprestatie voor het aanleveren van hun SE(s) kregen de docenten een terugkoppeling over de gedeelde SE's. Deze rapportages zijn geschreven met het doel dat deze voor het maken van toekomstige SE's van de docenten behulpzaam zouden zijn. De terugkoppelingen zijn in eerste instantie gegenereerd door CitoChat (de interne, beschermde GPT van Stichting Cito, waarbij de invoer niet gedeeld wordt met OpenAI). Het onderzoek is toegelicht aan CitoChat en per pakket zijn het beoordelingskader en de antwoorden op de open vragen ingevoerd. Vervolgens is per pakket aan CitoChat gevraagd om op basis van het beoordelingskader en de antwoorden op de open vragen een terugkoppeling te schrijven die gericht is aan de docent, waarin de belangrijkste bevindingen terugkomen, en waarin de volgende onderdelen terugkomen: samenvattend oordeel, sterke punten, aandachtspunten en aanbevelingen. De gegenereerde terugkoppelingen zijn naderhand gecontroleerd door de betrokken onderzoekers en indien nodig aangepast. In Bijlage 7 staat een voorbeeld van een terugkoppeling.

2.3.6 Analyse

Er is per SE een verschillend aantal materialen aangeleverd. Aangezien een SE per vak en schooltype vaak uit een verzameling aan toetsen, werkstukken en andere opdrachten bestaat, hebben de meeste docenten meer toetsen, werkstukken en dergelijke ingestuurd. Een verzameling van dergelijke materialen dat betrekking heeft op een SE van een vak bij een schooltype wordt een pakket genoemd. Per toets of opdracht kunnen ook meer materialen meegestuurd zijn. Denk hierbij naast de toets of de opdracht ook aan het beoordelingsmodel, een toetsmatrijs, of de gebruikte normering. De eerste stap van de analyses betreft de inventarisatie van de gedeelde materialen.

De stap hier opvolgend is de analyse van de ontbrekende waarden. In tegenstelling tot de vragenlijsten die alleen ingeleverd konden worden als alle vragen beantwoord waren, was het wel mogelijk ontbrekende waarden te hebben bij de beoordelingen. Dat was zo ingesteld omdat wanneer specifieke onderdelen van een SE ontbreken, zoals een toetsmatrijs, de beoordelingen ervan dienen te ontbreken. Ook wanneer een specifieke vraag in het kader bij een aangeboden SE niet te beoordelen was, kon dat een ontbrekende waarde opleveren. In de resultaten is per criterium weergegeven over welk aantal SE's de beoordeling is ingevuld.

In de beschrijving van de resultaten ligt de focus dan ook op de losse SE's, maar we geven de resultaten eveneens op het niveau van de pakketten, zodat we ook corrigeren voor de verschillen in de grootte van de pakketten (aantal SE's per pakket). Per onderdeel uit het beoordelingskader is berekend hoeveel procent van de antwoorden 'ja' of 'nee' is. Om meer duiding te geven aan deze kwantitatieve gegevens, was aan de inhoudsdeskundigen ruimte gegeven om nuancerende of verklarende opmerkingen te plaatsen bij deze kwantitatieve gegevens. Andere kwalitatieve gegevens die gebruikt zijn betreffen de antwoorden op de vier open vragen aan het einde van het beoordelingskader. Daarmee geeft de beoordelend expert een overkoepelend, inhoudelijk oordeel over de sterke punten van het beoordeelde pakket SE's, en wat punten zijn die mogelijk versterkt kunnen worden.

2.4 Interviews vakverenigingen

2.4.1 Methode

Aangezien er slechts beperkte inzichten waren in de rol die vakverenigingen binnen de context van het SE spelen, is er gewerkt met een aantal open vragen. De voornaamste vragen waren:

1. Welke rol speelt de vakvereniging op dit moment bij SE's?
2. Welke activiteiten onderneemt u in dit verband?
3. Is er binnen de vakvereniging behoefte aan een andere rol (bijv. groter/kleiner)?
4. Welke ondersteuning heeft uw vereniging nodig om deze rol te vervullen?

Om antwoord op deze vragen te krijgen, is er gekozen voor de interviewvorm, waarin met een vertegenwoordiger van de vakvereniging kort ondervraagd wordt (beoogde tijdsduur: 15 minuten). Deze interviews zijn digitaal uitgevoerd via Microsoft Teams. De antwoorden zijn automatisch getranscribeerd met behulp van de geïntegreerde functionaliteit van Teams. Van te voren waren de geïnterviewden ingelicht via de e-mail over de hoofdvragen.

Na de eerste twee gesprekken bleek dat vakverenigingen mogelijk een rol voor Stichting Cito zagen. Daarom is de vraag 'Welke rol zou Cito eventueel kunnen spelen in de ondersteuning van uw vakvereniging in de context van de SE's?' in de latere gesprekken toegevoegd.

2.4.2 Werving

In september 2025 zijn per e-mail de eerste uitnodigingen verstuurd aan alle 22 vakverenigingen die bij ons bekend waren. In oktober 2025 zijn herinneringen verzonden naar de vakverenigingen van wie tot dat moment geen reactie was ontvangen. Het overzicht van de verenigingen en respons op het verzoek is te vinden in Tabel 2.4.1. De gesprekken hebben plaatsgevonden tussen 22 september en 21 november 2025.

Tabel 2.4.1 Overzicht van benaderde en gesproken vakverenigingen

Vakvereniging	Vak	Interview
VONKC*	Beeldende Vakken	Ja
VLS*	Muziek	Ja
NVON*	Biologie, Natuurkunde, Scheikunde/Nask2	Ja
NNV	Natuurkunde	Via NVON
NVvW*	Wiskunde	Ja
LVT – Arabisch*	Arabisch	Geen Reactie
LVT – Duits*	Duits	Ja
LVT – Engels*	Engels	Ja
LVT – Frans*	Frans	Geen Reactie
LVT – Fries*	Fries	Nee
VCN	Grieks en Latijn	Ja
LVT – Nederlands*	Nederlands	Geen Reactie
LVT – Russisch*	Russisch	Geen Reactie
LVT – Spaans*	Spaans	Geen Reactie
VDSN	Spaans	Ja
LVT – Turks*	Turks	Geen Reactie
KNAG	Aardrijkskunde	Ja
Vecon	Bedrijfseconomie en Economie	Geen Reactie
VFVO	Filosofie	Ja
VGN	Geschiedenis	Ja
KNHG	Geschiedenis	Via VGN
NVLM	Maatschappijkunde	Ja

* zijn georganiseerd als vakbond

Van de 22 uitnodigingen zijn er twaalf geaccepteerd. Bij twee vakken gaf een vakvereniging aan dat de antwoorden gegeven zouden worden via de andere vakvereniging die ook hetzelfde vak (mede) dekt. Bij de overige acht verenigingen is geen interview gehouden omdat ze dit afwezen (één vereniging) of omdat van die vereniging ook na herinnering geen reactie ontvangen was (zeven verenigingen). Van de acht verenigingen die niet geïnterviewd waren, betreffen er zeven talenvakken. Van deze zeven betrof vijf talen waar relatief weinig kandidaten eindexamen in doen, maar ook het verplichte vak Nederlands, waar iedereen een eindexamenresultaat op moet hebben om een diploma te krijgen.

2.4.3 Verwerking van de gegevens

In de praktijk duurden de interviews tussen de 20 en 45 minuten. In het gesprek met de Engelse sectie van de Levende Talenvereniging schoof één vertegenwoordiger van het algemene bestuur van Levende Talen aan, zodat het perspectief van de overige Levende Talen aan tafel werd vertegenwoordigd. Bovendien zijn er twee vakverenigingen voor Spaans, waarvan de vereniging die niet onder Levende Talen valt, is geïnterviewd.

De transcripten van de interviews zijn vervolgens geanalyseerd en de relevante uitspraken zijn per vraag geëxtraheerd. Vervolgens zijn deze resultaten uit de gesprekken geclusterd tot een aantal breed gedragen of veel voorkomende onderwerpen, uitdagingen, of vragen die binnen de vakverenigingen leven. Deze onderwerpen zijn verder beschreven in de resultaten sectie van dit rapport.

3. Resultaten

3.1 Resultaten van het vragenlijstonderzoek

3.1.1 Algemeen

3.1.1.1 Respondenten docentenvragenlijst

In totaal hebben 241 respondenten de vragenlijst ingevuld, verdeeld over 181 scholen. Per provincie zijn vier tot veertien van de deelnemende scholen gehuisvest zijn, waarbij de provincies Gelderland, Noord-Brabant en Noord-Holland enigszins oververtegenwoordigd zijn vergeleken met de andere provincies. Van de respondenten maken er 153 SE's voor zowel het examenjaar als een voor-examenjaar, maken er 77 alleen SE's voor het examenjaar en maken 11 respondenten alleen SE's voor een voor-examenjaar. Binnen de groep respondenten is er een grote variëteit aan vakken waarvoor ze de SE's maken en voor welke onderwijsniveaus ze dit doen (zie Tabel 3.1.1.1). Als onderwijsniveau is vwo wellicht oververtegenwoordigd, gezien de helft van de respondenten aangeeft SE's te maken voor het vwo.

Tabel 3.1.1.1 Combinaties van schooltype en vaksoort

	Schooltype				
Vaksoort	vmbo bb/kb	vmbo gl/tl	havo	vwo	**Totaal (per vak)**
Zaakvakken					**60**
AK/Ges/ML	2	6	13	17	**38**
Economie/ondernemen	7	1	4	10	**22**
Talen					**77**
Nederlands	1	6	9	11	**27**
Moderne vreemde talen	2	9	5	23	**39**
Klassieke talen	0	0	0	11	**11**
Science-vakken					**86**
Biologie	3	6	1	10	**20**
Natuurkunde/scheikunde	2	4	8	12	**26**
Wiskunde/informatica	4	6	7	23	**40**
Overig					**18**
Profielvakken vmbo	7	2	0	0	**9**
Kunst/drama/muziek	0	2	0	3	**5**
Overig	1	1	1	1	**4**
Totaal (per schooltype)	**29**	**43**	**48**	**121**	**241**

3.1.1.2 Respondenten vragenlijst voor examencommissies

In totaal hebben 85 respondenten de vragenlijst ingevuld. De examencommissies bestonden uit 1 tot 8 leden. De meeste commissies hadden 3, 5 of 7 leden. Van de deelnemende examencommissieleden is bijna de helft (42 respondenten) examensecretaris. De geïnterviewde examencommissieleden werken vaak voor een combinatie van schooltypes.

3.1.2 Constructie

Aan zowel docenten als leden van de examencommissie is gevraagd naar de manier(en) waarop de SE's samengesteld worden op de school waar zij werkzaam zijn. Hiermee is inzicht verkregen in wie er doorgaans betrokken is, welke toetsvormen toegepast worden, welke bronnen gebruikt worden, en welke (hulp)middelen er gebruikt worden bij het samenstellen van de SE's.

De meerderheid van de leden van de examencommissie geeft aan dat docenten op de school waar zij werken de ruimte krijgen om individuele keuzes te maken in de constructie, afname, en beoordeling van de SE's (77,6%, zie Bijlage 5 Tabel 5.1). Maar een klein deel van de leden van de examencommissie ervaart deze ruimte niet zijn of haar school (9,4%).

3.1.2.1 Toetsvormen

Voor het SE worden voornamelijk schriftelijke toetsvormen gebruikt. Leden van de examencommissie geven aan dat docenten op hun school voornamelijk schriftelijke toetsen gebruiken (70,6%, zie Bijlage 5, Tabel 5.2). Maar een klein deel geeft aan dat er in min of meer gelijke mate gebruik gemaakt wordt van schriftelijke en andere toetsvormen (16,5%) of dat er veel variatie is tussen vakken als het gaat om toetsvormen (12,9%). Docenten geven hetzelfde beeld: voor 30,7% van de vakken is de SE-toets (bijna) altijd schriftelijk, en voor 51% van de vakken meestal schriftelijk (Zie Bijlage 4, Tabel 4.9). Maar 0,8% van de docenten geeft aan dat er voor zijn of haar vak nooit of bijna nooit schriftelijk getoetst wordt. Er zijn geen opvallendheden tussen schooltypen in het aandeel schriftelijke toetsen voor het SE. Opvallend is dat bij de taalvakken zeer weinig docenten (10,4%) aangeven (bijna) altijd schriftelijk te toetsen, ten opzichte van 35% bij de zaakvakken en 48,8% bij de science-vakken. Dit is voornamelijk zo bij Nederlands (3,7%) en moderne vreemde talen (0%), omdat bij klassieke talen juist 63,6% van de docenten aangeeft vrijwel altijd schriftelijk te toetsen.

Schriftelijke toetsen worden dan ook door bijna alle docenten gebruikt als toetsvorm voor het SE (door 98,8% van de deelnemende docenten). Docenten konden meerdere toetsvormen aangeven die zij gebruiken voor de SE's. Ook praktijkopdrachten die thuis kunnen worden gemaakt (39,4%), praktijkopdrachten die ter plekke worden uitgevoerd (34,4%) en mondelinge toetsen (32,8%) worden geregeld ingezet als vorm voor SE's (zie bijlage 4 Tabel 4.2). Reproductieve praktijktoetsen zoals luistertoetsen komen minder vaak voor (17,4%), net als groepswerk (26,6%). Schriftelijke toetsen worden in alle schooltypen en vakken gebruikt, terwijl andere vormen meer vakgebonden worden gebruikt. Zo maakt het vmbo bb/kb relatief veel gebruik van praktijkopdrachten ter plekke (51,7% van de docenten). Docenten van moderne vreemde talen gebruiken naast schriftelijke toetsen ook veel niet-schriftelijke toetsen, zoals reproductieve praktijktoetsen (74,4%) en mondelinge toetsen (100%). Bij de zaakvakken aardrijkskunde, geschiedenis en maatschappijleer (gezamenlijk geanalyseerd) gebruiken relatief veel docenten praktijkopdrachten die leerlingen thuis kunnen maken (68,8%), wat aansluit bij de schrijf- en projectgerichte aard van deze vakken. Profielvakken vmbo en cultuurvakken kunst, drama en muziek (gezamenlijk geanalyseerd) zijn beroeps- en vaardigheidsgericht en kennen weinig thuisopdrachten, maar juist relatief veel praktijkopdrachten ter plekke, wat aansluit bij hun praktijkgerichte karakter.

Als aanvulling op de in de enquête genoemde toetsvormen noemen docenten voornamelijk specifiekere varianten van de hiervoor genoemde toetsvormen. Bij praktijkopdrachten zijn dit praktijkopdrachten die op school uitgevoerd moeten worden, waarvan leerlingen vervolgens thuis een verslag of vragen maken die pas na afloop, niet ter plekke door de docent, beoordeeld worden (genoemd door 5 respondenten). Daarnaast worden ook presentaties of debat genoemd als toetsvorm (4 respondenten). Digitale toetsen en computeropdrachten worden ook als alternatieve toetsvorm genoemd (5 respondenten). Minder vaak genoemde toetsvormen zijn portfolio's en excursies (beide genoemd door 1 respondent). Twee respondenten benoemen dat andere toetsvormen voornamelijk ingezet worden bij

leerlingen met bijzondere omstandigheden[52]. In dat geval wordt een schriftelijke toets bijvoorbeeld mondeling afgenomen.

Docenten geven aan dat zij de vorm van een SE-toets of opdracht voornamelijk bepalen door de toetsvorm af te stemmen op de inhoud (kennis, vaardigheden, inzicht, anders) die getoetst moet worden (zie bijlage 4, Tabel 4.6). 78,8% van de docenten geeft dit als (een van de) manieren om de toetsvorm te bepalen. Daarna worden de gebruikelijke wijze voor het specifieke examen wat de docent af gaat nemen en praktische overwegingen, zoals beschikbare tijd en ruimte, door een aanzienlijk deel van de docenten genoemd (respectievelijk 42,7% en 42,3% van de docenten). Consequent dezelfde toetsvorm hanteren (bijvoorbeeld altijd schriftelijk) wordt door een aanzienlijk minder groot deel van de respondenten gedaan (24,9%). Er zijn hierin wel verschillen tussen vaksoorten: 43,3% van de docenten die zaakvakken geven, geeft aan dat er wel consequent dezelfde toetsvorm wordt afgenomen. Dat is opvallend meer dan bij de talen (19,5%) en de science-vakken (8,2%). Verder geven alle docenten van de klassieke talen aan dat zij praktische afwegingen nooit meenemen bij het bepalen van de toetsvorm, en hanteren zij juist vaker consequent dezelfde toetsvorm (45,5% van de 11 docenten klassieke talen geeft dit als manier om de toetsvorm te bepalen). Voor het bepalen van de toetsvorm zijn er geen opvallende verschillen in schooltypen.

3.1.2.2 Bronmaterialen

Met betrekking tot de gebruikte bronnen voor SE vragen valt op dat veel docenten zelfgemaakte nieuwe opgaven (80,1%) en materiaal uit oude CE's (75,9%) gebruiken (zie bijlage 4, Tabel 4.3). Daarnaast wordt ook door veel docenten materiaal uit oude SE's (66%) en lesmethodes (61%) ingezet. Gebruik van digitale itembanken (16,6%) en materiaal van beroepsverenigingen (6,6%) komt veel minder voor. Dit patroon is vergelijkbaar over alle schooltypen. Moderne vreemde talen onderscheidt zich door een hoog aandeel eigen invulling en relatief frequent gebruik van itembanken. Materiaal van vakverenigingen wordt vooral, maar wel beperkt, gebruikt binnen aardrijkskunde, maatschappijleer en geschiedenis, en nauwelijks in andere vakken. Als aanvulling op de in de vragenlijst genoemde bronmaterialen, noemen docenten voornamelijk dat ze gebruik maken van materiaal van collega's, waaronder ook collega's buiten de school of vanuit een professionele leergemeenschap (PLG[53]; 6 respondenten). Twee respondenten geven aan om voor productieve vaardigheden (zoals spreken) gebruik te maken van materiaal van het Goethe instituut. Ook maken docenten gebruik van bestaand materiaal dat gebruikt is in olympiades (2 respondenten). Eén respondent geeft aan gebruik te maken van AI voor het maken van SE-toetsen en opdrachten.

Een duidelijke meerderheid van de docenten (zie bijlage 4, Tabel 4.4) betrekt in enige mate centraal-examenstof in hun SE, ook als dit volgens de syllabus niet verplicht is. Slechts een minderheid (14,1%) doet dit nooit. Dit beeld wordt bevestigd door leden van de examencommissie. Volgens leden van de examencommissie worden er voor de constructie van oude SE's vaak oude CE-opgaven gebruikt (82,4%, zie bijlage 5, Tabel 5.4), materiaal uit oude SE's (72,9%) en materiaal uit de lesmethode (71,8%). Ook bedenken docenten relatief vaak zelf nieuwe opgaven volgens de leden van de examencommissie (64,7%). Andere bronnen zoals materiaal van vakverenigingen (24,7%), materiaal in digitale itembanken (34,1%) worden ook volgens examencommissieleden minder vaak gebruikt.

Meer dan de helft van de docenten geeft aan dat ze centraal-examenstof opnemen voor meer dan de helft van de schriftelijke opgaven. Moderne vreemde talen wijkt hiervan af, doordat bij dit vak bijna de helft van de deelnemende docenten aangeeft niet of slechts sporadisch oude CE-vragen te gebruiken. De voornaamste genoemde reden dat docenten opgaven uit de oude CE's gebruiken, is dat de docent zo makkelijker en/of sneller toetsen van hoge kwaliteit kan maken (67%, zie bijlage 4, Tabel 4.5). Iets minder

[52] Dit antwoord was gegeven op een open vraag waarbij de docent aangaf dat ook andere vormen gebruikt zijn. Toelichtingen hierop waren verder niet heel exact welke omstandigheden dit betrof.
[53] https://plgonderwijs.nl/

dan een derde van de docenten geeft aan dat het beperken van verschillen met de cijfers van het CE een reden is voor het gebruik van oude CE-opgaven.

Hoewel er uiteraard overlap is tussen CE- en SE-stof, kan het relatief hoge gebruik van oude CE-opgaven mogelijk problematisch zijn: het SE is immers ook bedoeld om een ander (aanvullend) deel van het curriculum te toetsen dan het CE. Bij het gebruik van CE-opgaven is het daarom belangrijk om goed zicht te houden op de verhouding CE- en SE-leerstof. Hierin valt nog winst te behalen bij examencommissies: 18,8% van de leden van de examencommissie geeft aan dat er als examencommissie geen zicht is op welke bronnen gebruikt zijn bij de constructie van schriftelijke SE-toetsen (zie bijlage 5, Tabel 5.4). Een ander potentieel nadeel van een groot deel CE-opgaven is dat er mogelijk weinig ruimte overblijft om het schooleigen karakter terug te laten komen in het SE. Toch geeft de meerderheid van de leden van de examencommissie aan dat de SE's op school passen bij de visie van de school op toetsing en examinering en het daarop gebaseerde toetsbeleid (63,5%, zie bijlage 5, Tabel 5.3). Volgens 23,5% van de leden van de examencommissie sluiten de toetsen maar soms (23,5%) of zelden (1,2%) aan bij de visie van de school. Opvallend is dat een deel van de leden aangeeft de visie op toetsing niet te kennen (11.8%). Het gaat hierbij om de *ervaren* visie door vakdocenten en de examencommissie. Er is in de vraag niet geëxpliciteerd of het gaat om een onderwijskundige visie (inclusief schoolidentiteit en specialisatie), om een visie op toetsing (beleid, normering, etc.) of een combinatie van beide soorten visies. Idealiter zijn vakdocenten op de hoogte van zowel de onderwijsvisie als de visie op toetsing, om zowel de toetstechnische kwaliteit en het schooleigen karakter binnen het SE te kunnen waarborgen. Er is veel variatie tussen docenten in hoeverre zij aangeven dat het SE wat zij maken recht doet aan het schooleigen karakter (zie Bijlage 4, Tabel 4.12). Een aanzienlijk deel (38,6%) van de docenten is daar niet bewust mee bezig. Een derde van de docenten geeft aan dat de helft (14,9%) of minder (17,8%) van het SE past bij het schooleigen karakter. Voor 24,1% van de docenten voldoet (bijna) alle SE-onderdelen aan het schooleigen karakter, en bij 4,6% helemaal niet. Met name bij vmbo bb/kb geeft een groot deel (48,3%) van de docenten aan niet bewust bezig te zijn met het schooleigen karakter in vergelijking met de andere schooltypen (32,6 – 38%). Bij Nederlands en kunst/drama/muziek geeft een opvallend hoog percentage van de docenten (44,1% voor Nederlands en 40% voor kunst/drama/muziek) aan dat alle onderdelen van het SE bij het schooleigen karakter passen. Tussen de overkoepelende vaksoorten (science, zaakvakken, talen) zijn geen opvallende verschillen gevonden.

3.1.2.3 Samenwerking en (hulp)middelen bij constructie

In het algemeen hebben docenten bij de constructie van SE-vragen en -opdrachten een sterke voorkeur voor samenwerking in duo's binnen vaksecties (61%, zie bijlage 4, Tabel 4.1). Een klein deel van de docenten (14,5%) geeft aan de SE's samen te stellen met alle docenten binnen de vaksectie of juist individueel (19,9%). Brede samenwerking, zowel sectie-overstijgend (0,8%) als met de examencommissie (2,1%), komt daarentegen nauwelijks voor. Bij sommige vakken, zoals klassieke talen, de profielvakken vmbo, en cultuurvakken als kunst, drama en muziek, komt het relatief vaker voor dat docenten alleen verantwoordelijk zijn voor de toetsconstructie dan bij andere vakken.

De meerderheid van de docenten gebruikt de volgende drie middelen om in het constructieproces de kwaliteit van een schriftelijke SE toets te borgen: zeker weten dat de toets op eindniveau meet (74,7%), de toets door een collega laten becommentariëren (78,4%) en het hebben van een duidelijk beoordelingsmodel waarmee de toets nagekeken kan worden (91,3%, zie Bijlage 4, Tabel 4.7). Een aanzienlijk kleiner deel van de docenten geeft aan de toets samen te stellen door middel van een toetsmatrijs (29,5%) of stelt de hele toets samen met een collega op (30,3%). Er zijn voor de manieren van samenstellen geen opvallende verschillen tussen schooltypes of vaksoorten. Bij het samenstellen van praktische opdrachten geeft de meerderheid van de docenten (71%) aan dat zij hierbij een duidelijk beoordelingsformulier heeft waarmee de toets nagekeken kan worden en de toets ook door een collega laten becommentariëren (53,9%, zie bijlage 4, Tabel 4.7). Dit komt overeen met de ingezette

hulpmiddelen voor schriftelijke toetsen voor het SE. Het meten op eindniveau, gebruik van een toetsmatrijs en het samenstellen van de toets met een collega wordt door de minderheid van docenten gedaan voor praktische opdrachten (respectievelijk 42,3%, 20,7%, en 40,7% van de docenten). Er zijn geen opvallende verschillen tussen de schooltypes bij het samenstellen van een praktische opdracht. Bij talen wordt er door de meerderheid van de docenten aangegeven dat ze zeker weten dat ze op eindniveau meten (55,9%) tegenover een aanzienlijk minder groot deel bij de zaakvakken (28,3%) en science-vakken (37,2%). Als aanvulling op de in de vragenlijst genoemde middelen noemen drie respondenten dat zij bij praktische opdrachten gebruikmaken van een beoordelingsrubric om transparantie te waarborgen. Eveneens drie respondenten benadrukken het vooraf duidelijk formuleren van de opdracht en beoordelingscriteria. Externe opdrachten van vakverenigingen of universiteiten worden door drie anderen toegepast, terwijl twee respondenten expliciet inzetten op het toetsen van brede vaardigheden of het werken op het juiste niveau. De overige antwoorden zijn uniek en variëren van het bieden van keuzevrijheid aan leerlingen tot het maken van unieke opdrachten, of het geheel weglaten van een praktische opdracht.

Zoals hierboven genoemd, wordt er zowel bij schriftelijke als praktische opdrachten weinig gebruik gemaakt van toetsmatrijzen. Op het gebruik van toetsmatrijzen wordt ook niet heel streng gelet door examencommissies (zie Bijlage 5, Tabel 5.5). 36,5% van de examencommissieleden geeft aan dat het gebruik van toetsmatrijzen wel gestimuleerd wordt, maar dat ze niet vaak gebruikt worden. 42,5% van de examencommissieleden geeft aan dat toetsmatrijzen op hun school wel gebruikt kunnen worden, maar dat de examencommissie niet let op het gebruik ervan. Bij maar 5,9% van de examencommissieleden wordt het gebruik van een toetsmatrijs gestimuleerd en ook daadwerkelijk door docenten gebruikt. 9,4% van de examencommissieleden geeft aan helemaal geen toetsmatrijzen op school te gebruiken. Er valt in het gebruik van toetsmatrijzen mogelijk nog winst te behalen in de constructie van SE's. Een toetsmatrijs zorgt ervoor dat het examen systematisch alle relevante leerdoelen en vaardigheden dekt en dat de cognitieve niveaus (bijv. onthouden, toepassen, analyseren) in de juiste verhoudingen terugkomen, wat de inhoudsvaliditeit verhoogt. Daarnaast bevordert het consistentie en transparantie bij samenstelling en beoordeling, waardoor toetsing eerlijker en betrouwbaarder wordt..

3.1.3 Regels en richtlijnen

Aan zowel docenten als leden van de examencommissie is gevraagd naar de regels en richtlijnen omtrent de SE's op de school waar zij werkzaam zijn. Hiermee is inzicht verkregen in hoe SE's in het PTA worden gewogen, welke criteria daarbij een rol spelen en hoe consistent toetsen, inhaal- en herkansingstoetsen zijn binnen scholen. Ook komt aan bod welke verschillen er tussen schooltypes en vakken bestaan en in hoeverre examencommissies toetsen vooraf op kwaliteitseisen beoordelen.

De meeste docenten (78,4%) geven aan dat afzonderlijke SE-toetsen en opdrachten bij het vak dat zij geven in het PTA geen gelijk gewicht heeft om het SE cijfer te bepalen (Zie bijlage 4, Tabel 4.10). Bij vmbo bb/kb is de verdeling gelijker dan bij de andere schooltypes: bij vmbo bb/kb geeft ongeveer de helft van de docenten (51,7%) aan dat de verdeling gelijk is, en de helft dat dit verschilt tussen toetsen (48,3%). Bij vmbo gl/tl geeft het merendeel van de docenten aan dat toetsen niet allemaal hetzelfde gewicht krijgen (67,4%) en bij havo (89,6%) en vwo (85,1%) geven bijna alle docenten aan dat toetsen verschillende weging hebben. Bij economie/ondernemen geeft 40,9% aan dat toetsen wel een gelijke weging krijgen, wat hoger is dan bij andere vakken, en bij docenten van profielvakken vmbo geeft de meerderheid (66,7%) aan dat alle toetsen dezelfde weging krijgen. Tussen de overkoepelende vaksoorten (talen, zaakvakken, science) zijn geen opvallende verschillen te vinden in de weging van toetsen in het PTA.

Voor het bepalen van de weging van de afzonderlijke SE's houden ze op school volgens examencommissieleden vooral rekening met de hoeveelheid lesstof (60%), de inhoud van de lesstof

(48,2%) en de moeilijkheidsgraad (47,1). Door minder leden werd ook het leerjaar (43,5%), de toetsvorm (41,2%) en de toetsduur (36,5%) als reden genoemd. Een deel van de examencommissieleden (23,5%) geeft aan dat de examencommissie geen zicht heeft op hoe de weging van het SE bepaald wordt.

Docenten bevestigen dit beeld: voor het bepalen van de weging van afzonderlijke toetsen en opdrachten van het SE spelen volgens docenten voornamelijk verschillen in de hoeveelheid lesstof (55,2%), inhoud van de lesstof (51,9%), leerjaar (45,6%) en de toetsvorm (48,5%) (Zie bijlage 4, Tabel 4.11). Moeilijkheid (38,2%) en toetsduur (26,6%) spelen in mindere mate ook een rol in de weging. Leerjaar is vooral bij vwo en havo een reden om de weging aan te passen (respectievelijk 64,6% en 52,9%) en minder bij vmbo bb/kb (24,1%) en vmbo gl/tl (18,6%). Bij vwo wordt moeilijkheid door een minder groot deel van de docenten (25,6%) als reden voor de weging aangegeven dan bij de andere schooltypes (44,8% tot 58,1%). Bij zaakvakken en science-vakken wordt de hoeveelheid van de lesstof door de meeste docenten als reden voor weging opgegeven (70% en 63,9%), bij talen de inhoud van de lesstof (67,3%).

Bijna alle docenten (95%) geven aan dat leerlingen in hetzelfde examenjaar precies dezelfde SE's krijgen (herkansingen en inhaaltoetsen daargelaten) (zie Bijlage 4 Tabel 4.13). Ook volgens bijna alle examencommissieleden (96,4%) krijgen leerlingen in dezelfde SE-periode op hun school voor hetzelfde vak meestal exact dezelfde SE toets of opdracht (zie Bijlage 5, Tabel 5.8). De meerderheid van de docenten (71,4%) geeft aan dat inhaaltoetsen dezelfde inhoud dekken als de originele toets, maar andere opdrachten bevatten (zie Bijlage 4, Tabel 4.14). Bij een kleiner deel (27%) is de toets vrijwel identiek aan de originele toets, en zeer weinig docenten (1,7%) laten zowel de toetsvorm als de inhoud van een inhaaltoets verschillen. Er zijn duidelijke verschillen tussen vmbo en havo/vwo. Bij havo en vwo geeft maar een klein gedeelte van de docenten (16,7 en 12,4%) aan dat de inhaaltoets gelijk is aan de originele toets, en bevat de toets veel vaker dezelfde inhoud maar andere opgaven (81,3% en 87,6%). Bij vmbo is de toets vaker identiek aan de inhaaltoets (51,7 en 62,8%) dan bij havo/vwo. Op basis van de antwoorden is niet te achterhalen of dit alleen praktische opdrachten of ook schriftelijke toetsen betreft. Gezien de hoge percentages is het aannemelijk dat hier ook schriftelijke toetsen bij zitten. Er zijn geen opvallende verschillen tussen de overkoepelende vaksoorten. In de specifieke vaksoorten valt op dat bij kunst/drama/muziek geen enkele docent aangeeft dat de toets vrijwel identiek is aan de originele toets.

Dit beeld is vrijwel hetzelfde voor herkansingstoetsen. De meerderheid van de docenten (87,1%) geeft aan dat herkansingstoetsen dezelfde inhoud dekken als de originele toets, maar andere opdrachten bevatten (zie Bijlage 4, Tabel 4.15). Bij een kleiner deel (10%) is de toets vrijwel identiek aan de originele toets, en zeer weinig docenten (0,8%) laten zowel de toetsvorm als de inhoud van een herkansingstoets verschillen. Ook hierbij zijn er duidelijke verschillen tussen vmbo en havo/vwo. Bij havo en vwo geeft maar een zeer klein gedeelte van de docenten (2,1 en 6,6%) aan dat de herkansing gelijk is aan de originele toets, en bevat de toets veel vaker dezelfde inhoud maar andere opgaven (95,8% en 91,7%).Bij vmbo is dit ook zo, maar geven meer docenten (24,1% en 18,6%) aan dat de herkansing ook weleens gelijk is qua inhoud aan de originele toets. Bij talen is de herkansing vaker vrijwel identiek aan de originele toets (22,1%) dan bij zaakvakken en science-vakken. Verder is opvallend dat bij biologie, natuurkunde/scheikunde en kunst/drama/muziek geen van de docenten aangeeft dat de toets vrijwel identiek kan zijn aan de originele toets.

Bijna de helft van de examencommissieleden (49,4%) geeft aan niet van tevoren de SE's te toetsen op kwaliteitseisen (zie Bijlage 5, Tabel 5.7). De kwaliteitseisen waar vooraf wordt beoordeeld of een toets voldoet zijn vrij gelijk verdeeld: of de toetsvorm goed aansluit bij de lesstof (40%), of de instructie duidelijk is (38,8%), of de opdrachten passen bij de leerdoelen (38,8%), of het taalniveau past bij de doelgroep (35,3%) en of de opdrachten niet te makkelijk of te moeilijk zijn (30,6%). Aanvullende manieren om de kwaliteit te beoordelen zijn steekproefsgewijs examens beoordelen (2 respondenten) en het 4-ogenprincipe toepassen, waarbij er altijd 2 docenten naar een examenbeoordeling kijken (2 respondenten).

3.1.4 Leerlingen

Een belangrijk aspect om de kwaliteit van de SE's te borgen is de voorbereiding en ervaring van leerlingen. Hierbij gaat het om in hoeverre leerlingen vooraf weten hoe ze worden beoordeeld, en de communicatie over toegestane hulpmiddelen rond SE-afnames. Ook is er gevraagd of leerlingen tijdens de SE's tijdgebrek ervaren.

De meerderheid van de docenten (68%) geeft aan dat leerlingen altijd bij alle toetsen en toetsvormen van tevoren weten hoe ze beoordeeld worden (door bijvoorbeeld het aantal punten per opgave te weten, of de beoordelingscriteria) (zie Bijlage 4, Tabel 4.17). Sommige docenten (17%) geven aan dat dit wel altijd zo is bij schriftelijke toetsen, maar niet altijd bij andere toetsvormen. Bij een klein deel van de docenten weten leerlingen dit niet altijd bij schriftelijke toetsen, maar wel bij andere toetsvormen (5,4%) of verschilt het bij alle toetsvormen hoe goed leerlingen op de hoogte zijn van de beoordelingscriteria (9,5%). Er zijn geen opvallende verschillen tussen schooltypes en vaksoorten.

Daarnaast vinden bijna alle examencommissieleden dat er op school tijdig wordt gecommuniceerd over de toegestane hulpmiddelen tijdens de afname van SE's (94,1%) (zie Bijlage 5, Tabel 5.8).

Docenten geven aan dat leerlingen soms (45,2%) of zelden (42,3%) te maken hebben met tijdgebrek bij de afname van SE-toetsen en opdrachten (zie Bijlage 4, Tabel 4.18). Maar een paar docenten zeggen dat dit nooit (4,1%), vaak (6,6%) of altijd (0,8%) voorkomt. Een aantal docenten (0,8%) weet niet of leerlingen tijdgebrek ervaren. Er zijn geen opvallende verschillen tussen schooltypes en vaksoorten.

3.1.5 Beoordelen

Om zicht te krijgen op het proces van beoordeling van de SE's, is docenten en leden van de examencommissie gevraagd welke methoden zij gebruiken om objectiviteit en normering te waarborgen, welke afspraken scholen hebben over score-naar-cijferbepaling, en welke praktijken rond vaststellen en bijstellen van de cesuur gebruikt worden.

De meerderheid van de docenten zorgt ervoor dat alleen leerlingen die de leerstof goed beheersen een voldoende kunnen halen door de toetsvorm goed aan te laten sluiten bij de leerstof (83%) en door toepassings- en inzichtopdrachten te gebruiken (56,8%) (zie Bijlage 4 Tabel 4.16). In mindere mate geven docenten aan het niveau van de opdrachten hoog te leggen (43,6%), de toetsen lang genoeg te maken (34,9%) en geen algemene opdrachten te gebruiken (11,2%). Er zijn hierbij geen opvallende verschillen tussen schooltypes en vaksoorten. Aanvullend op de in de vragenlijst gegeven antwoorden zijn er 5 respondenten die ervoor zorgen dat alleen leerlingen die de leerstof beheersen een voldoende halen door variatie in benodigde vaardigheden te toetsen, zoals reproductie, toepassing, inzicht en eindtermgerichte vragen. 3 respondenten letten vooral op het juiste niveau, bijvoorbeeld door het afstemmen op het centrale examen of de plaats in het schoolprogramma. 2 respondenten richten zich op het gebruik van examenvorm of stijl (stijl CE) om het niveau en herkenbaarheid te borgen. 1 respondent noemt expliciet het hanteren van de N=1.0-normering als borging van voldoende eis. 1 respondent kiest voor een variatie in moeilijkheid met een mix van gemakkelijke en moeilijke vragen. 1 respondent legt de lat bewust hoog, met strenge cesuur (hoog niveau). Andere unieke factoren zijn: inzet (1 respondent), het meenemen van het proces (1 respondent), herkenbare opzet (1 respondent), intrinsieke interesse (1 respondent) en constateren dat het momenteel niet het geval is (1 respondent).

Voor de grens tussen voldoende en onvoldoende oordeel op een toets of opdracht gebruikt 43,6% een van tevoren vastgesteld aantal scorepunten, maar dat kan variëren per toets afhankelijk van de ingeschatte moeilijkheid (zie bijlage 4, Tabel 4.20). 39,8% van de docenten stelt het van tevoren vastgestelde aantal scorepunten bij als het percentage onvoldoendes veel hoger of lager is dan ze verwachten en bij 38,6% staat er een percentage van de maximumscore vast en wordt hier niet van afgeweken. Daarnaast geeft 3,3% van de docenten aan dat de grens tussen voldoende en onvoldoende

wordt meegeleverd met de toets of opdracht en zij bepalen dat dus niet zelf en slechts 2,5% stelt van tevoren geen aantal scorepunten vast en laat dit afhangen van een vastgesteld percentage onvoldoendes.

Docenten zorgen op meerdere manieren voor een objectieve beoordeling bij SE's (zie Bijlage 4, Tabel 4.19). De meest genoemde manieren zijn door duidelijke beoordelingscriteria op te stellen waar ze zich aan houden (97,9%) en door te overleggen en af te stemmen met collega's (81,3%). Ongeveer de helft van de respondenten beoordeelt zonder te kijken naar van wie het werk is ten behoeve van de objectiviteit (52,7%) en laat hun eigen gevoel of mening zoveel mogelijk buiten beschouwing bij het nakijken (50,2%). Docenten van de talen kijken minder vaak na zonder te weten van wie het werk is (36,3%) dan docenten van zaakvakken (56,7%) en science-vakken (66,3%). Daarnaast zorgt 36,5% er bewust voor dat ze uitgerust aan de slag gaan en niet te lang achter elkaar aan het werk zijn. Slechts 5% laat een tweede beoordelaar een praktisch SE nakijken en 2,5% laat een tweede beoordelaar een schriftelijke SE nakijken.

Leden van de examencommissie geven aan dat er op hun school vaak (56,5%) of altijd (24,7%) gebruik wordt gemaakt van antwoord- en beoordelingsmodellen bij de beoordeling van niet-schriftelijke opdrachten van het SE (Zie Bijlage 5, Tabel 5.8). Een kleiner deel van de examencommissieleden (4.7%) denkt dat dit soms wel, maar soms ook niet gebeurt. 14.1% van de examencommissieleden geeft aan niet te weten of antwoord- en beoordelingsmodellen gebruikt worden bij de beoordeling van het SE. Bij het SE wordt er voor schriftelijke opdrachten in de meeste gevallen nooit (43,5%) of zelden (36,5) gebruikgemaakt van een tweede beoordelaar. Voor praktische opdrachten gebeurt dit soms (40%), maar ook vaak niet of zelden (42,4%). Bij praktische opdrachten wordt er volgens examencommissieleden dus vaker gebruik gemaakt van een tweede beoordelaar dan bij schriftelijke opdrachten.

Op slechts 20,3% van de scholen zijn er regels voor het bepalen welke score of prestatie welk cijfer oplevert, hoewel de docenten aangeven uiteindelijk zelf te bepalen hoe ze het berekenen (zie Bijlage 4, Tabel 4.21). Op de meerderheid van de scholen zijn geen regels over het bepalen welke score of prestatie welk cijfer oplevert en wordt dit per SE moment of opdracht bepaald (45,6%). Sommige docenten geven aan dat zij samen met collega's regels hiervoor hebben opgesteld (25,7%) of dat er wel regels zijn maar ze het uiteindelijk zelf bepalen (20,3%). De meerderheid van de examencommissieleden (71,8%) geeft aan dat er op school geen regels en richtlijnen zijn voor het bepalen welke score of prestatie op een toets welk cijfer oplevert (Zie bijlage 5, Tabel 5.9). 17,6% heeft die wel en vindt ze ook bruikbaar voor docenten, en 10,6% heeft ze wel maar vindt ze niet altijd of vrijwel niet bruikbaar voor docenten.

Docenten wijken ook wel eens bewust af van de vooraf opgestelde beoordelingscriteria, voornamelijk als er antwoorden zijn die ook goed zijn maar niet voorzien in de beoordelingscriteria (78,4%) of als de vraag onduidelijk bleek te zijn (68,9%) (Zie bijlage 4, Tabel 4.22). In sommige gevallen wordt ook afgeweken als een vraag moeilijker of makkelijker bleek dan verwacht (25,7%).

3.1.6 Evaluatie en deskundigheid

Dit thema beschrijft hoe vaksecties en examencommissies SE's evalueren, welke aspecten daarbij centraal staan en wie bij deze evaluaties betrokken zijn. Daarnaast wordt ingegaan op de deskundigheid van docenten en hoe deze wordt ingeschat door henzelf en door examencommissies, evenals de manieren waarop scholen deze deskundigheid bevorderen.

3.1.6.1 Evaluatie

Vaksecties evalueren SE's het meest op of het SE de eindtermen toetst zoals beschreven in het PTA (70,5%), op de kwaliteit van individuele opgaven en opdrachten (63,9%), of het SE het juiste kennis- en vaardigheidsniveau heeft (60,2%) en op het proces van beoordeling en normering (54,4%) (zie Bijlage 4, Tabel 4.23). Regelmatig wordt ook geëvalueerd op of het aantal onvoldoendes passend is (42,7%) en op de manier waarop toetsen worden gemaakt en het bronmateriaal dat daarvoor is gebruikt (40,2%). In mindere mate wordt ook de moeilijkheid van de opgaven geëvalueerd (26,6%), wordt geëvalueerd of het

SE de juiste vorm, opbouw en lay-out heeft (24,5%) en of het SE de juiste weging van de onderdelen hanteert (18,3%).

Examencommissieleden geven aan dat vooral de vaksectie betrokken is bij de systematische evaluatie van SE's (77,6%), gevolgd door de examencommissie (64,7%) en de schoolleiding (37,6%) (zie Bijlage 5, tabel 5.12). Een kleinere groep examencommissieleden geeft aan geen systematische evaluatie uit te voeren (17,6%). Leerlingen worden minder vaak betrokken (11,8%), terwijl de medezeggenschapsraad (3,5%) en het bestuur (2,4%) slechts zelden een rol spelen. De examencommissie evalueert het vaakst intern de dekking van de eindtermen/domeinen door het gehele SE (60%). Daarnaast wordt relatief vaak gekeken naar de vorm, opbouw en lay-out van iedere toets/opdracht in het SE (45,9%) en naar de weging van de verschillende onderdelen (44,7%). Ook het proces van beoordeling en normering komt bij een aanzienlijk deel aan de orde (34,1%), net als het kennis- en vaardigheidsniveau van de SE's (31,8%). Minder vaak wordt aandacht besteed aan de manier waarop SE's worden gemaakt en het bronmateriaal dat daarvoor is gebruikt (22,4%). Bij 24,7% van de examencommissieleden speelt geen van de genoemde punten.

Examencommissieleden geven aan dat per toets of opdracht analysegegevens na de afname wisselend worden gebruikt: 7,1% geeft aan dat dit nooit gebeurt, 21,2% zegt zelden, 32,9% soms en 24,7% vaak; niemand antwoordt altijd (0%), en 14,1% weet het niet (zie Bijlage 5, Tabel 5.14).

3.1.6.2. Deskundigheid

Docenten konden aangeven hoe zij hun deskundigheid zouden inschatten voor een aantal onderdelen rondom SE's. Dit konden ze inschatten op 'helemaal niet deskundig', 'beperkt deskundig', 'voldoende deskundig' of 'zeer deskundig' (zie Bijlage 4, Tabel 4.24 t/m 4.30). Voor het maken van schriftelijke opgaven schat de meerderheid zich in als zeer deskundig (56,8%) en een groot deel als voldoende deskundig (41,5%). Slechts vier docenten schatten dit in als beperkt deskundig. Bij de zaakvakken schatten de docenten hun deskundigheid hierin iets lager in; hier noemen docenten zich vaker voldoende deskundig (55%) dan zeer deskundig (43,3%). Bij het maken van praktische opdrachten zijn de antwoorden wat meer verspreid. De meerderheid geeft aan hierin voldoende deskundig te zijn (56,4%), gevolgd door de groep die zich inschat als zeer deskundig (30,3%). Negentien docenten noemen zichzelf hierin beperkt deskundig en slechts één docent helemaal niet deskundig. Ook bij het gebruik van toetsmatrijzen zien we dat de meerderheid zichzelf als voldoende deskundig inschat (44,8%). Hier is de groep docenten dat zich beperkt deskundig inschat echter groter (22,0%). 21,2% van de docenten noemen zichzelf zeer deskundig hierin en 5,4% juist helemaal niet deskundig. De deskundigheid op het beoordelen van lange antwoorden in een toets wordt daarentegen vrij hoog ingeschat door de docenten. De meerderheid noemt zichzelf zeer deskundig op dit gebied (53,5%), nauw gevolgd door de groep die voldoende deskundig is (41,9%). Slechts 6 docenten (2,5%) noemen zichzelf hierin beperkt deskundig. Bij de zaakvakken schatten de docenten hun deskundigheid hierin iets lager in; 55% noemt zich maar voldoende deskundig en 43,4% zeer deskundig. Deskundigheid in het beoordelen van praktijkopdrachten in werkstukvorm schatten de docenten gemiddeld gezien iets lager in. De meerderheid noemt zichzelf voldoende deskundig (57,7%). 27,4% van de docenten zouden zichzelf omschrijven als zeer deskundig en 7,1% docenten als beperkt deskundig. Ook voor het beoordelen van praktijkopdrachten in een live setting schat de meerderheid zichzelf als voldoende deskundig in (45,6%). 24,1% van de docenten noemt zichzelf zeer deskundig op dit gebied en 11,6% noemt zichzelf beperkt deskundig. Slechts 2,9% noemt zichzelf hierin helemaal niet deskundig. Docenten schatten hun deskundigheid op het bepalen van de cijfers in als voldoende deskundig (45,2%) of zeer deskundig (54,4%). Slechts één docent (0,4%) schatte dit in op beperkt deskundig en geen docent noemt zichzelf helemaal niet deskundig. Voor al deze vragen rondom deskundigheid zijn geen bijzonderheden te zien voor schooltype of vaksoort.

Het merendeel van de examencommissieleden vindt docenten over het algemeen voldoende of zeer deskundig op het gebied van het maken van opdrachten (75,3%), het beoordelen van open antwoorden

en essayvragen (78,9%), het beoordelen van praktijkopdrachten in een live setting (75,3%) en in het beoordelen van praktijkopdrachten in werkstukvorm (80%) en in het bepalen van de cijfers (82,4%) (Zie bijlage 5, Tabel 5.10). Meestal wordt de deskundigheid ingeschat als voldoende deskundig (tussen de 54 en 67%) en soms als zeer deskundig (11 tot 22%). Alleen in het gebruik van toetsmatrijzen zouden examencommissieleden docenten helemaal niet (17,6%) of beperkt deskundig (43,5%) inschatten.

Volgens examencommissieleden zorgen de meeste scholen ervoor dat collega's deskundig zijn in het opstellen van SE's door kennis en ervaringen uit te wisselen binnen de school (69,4%) (Zie bijlage 5, tabel 5.11). Daarnaast wordt ook vaak geleerd van de kennis van de examensecretaris (37,6%). Ook geeft een aanzienlijk deel van de respondenten aan dat collega's deskundig blijven door het volgen van trainingen over examinering (30,6%). Het uitwisselen van kennis en ervaringen met examencommissies in de regio of binnen het eigen netwerk wordt door 24,7% genoemd. Slechts een klein aantal scholen raadpleegt actuele informatie over examens en wetgeving (2,4%).

3.1.7 Positieve punten en verbeterpunten

3.1.7.1 Positieve punten

Er werd aan de docenten gevraagd wat zij als positieve aspecten van hun SE's zouden beschrijven en wat verbeterpunten zijn (Zie Bijlage 4, Tabel 4.31 en 4.32). De positieve aspecten die het meest gekozen zijn door de docenten zijn dat er veel ruimte is voor eigen invulling van het PTA en SE (78,4%) en dat ze een deskundige vaksectie hebben (78%). Daarnaast werden goede afspraken binnen de school (35,7%) en goede afstemming met de examencommissie (35,5%) regelmatig genoemd. In mindere mate kwamen mogelijkheden voor bijscholing (25,6%) en goede sturing vanuit de schoolleiding (14,1%) naar voren.

De meerderheid van de examencommissieleden vindt de SE's op school van voldoende kwaliteit (62,4%) (zie bijlage 5, Tabel 5.15). Een klein deel (7,1%) vindt dat dit niet zo is, en 15,3% weet niet of de kwaliteit voldoende is. Door examencommissies wordt ruimte voor eigen invulling van het PTA en SE door vaksecties (78,8%) het vaakst als positief aspect gezien (zie Bijlage 5, Tabel 5.16). Daarnaast wordt deskundigheid binnen vaksecties vaak genoemd (62,4%) en goede afstemming binnen de vaksecties door 57,6% van de respondenten. Goede afspraken binnen de school worden door 42,4% benoemd, terwijl 34,1% mogelijkheden voor bijscholing aangeeft. Slechts 11,8% ervaart goede sturing vanuit de schoolleiding, en 32,9% meldt een goede afstemming tussen vaksecties en de examencommissie.

3.1.7.2 Verbeterpunten

Een verbeterpunt dat er bij docenten duidelijk bovenuit sprong was het gebrek aan tijd (56,4%). Regelmatig werd ook aangegeven dat er geen duidelijke afspraken binnen de school zijn (22,4%). Andere verbeterpunten kwamen in minder dan 11% van de respons voor. Er zijn 14 respondenten die expliciet aangeven geen verbeterpunt te hebben, of tevreden te zijn met het huidige SE. Daarnaast hebben enkele respondenten eigen verbeterpunten geformuleerd: 3 respondenten noemen meer overleg binnen de sectie of vakgroep als verbeterpunt, en 2 respondenten willen bredere samenwerking tussen vakken. 2 respondenten benoemen werkbaarheid en de invloed van AI als probleem bij toetsen en beoordeling. 2 respondenten ervaren weerstand van collega's tegen verandering of wettelijke eisen. Unieke, door één respondent genoemde verbeterpunten zijn onder andere: behoefte aan bijscholing, steun van leidinggevenden, beoordeling van groepswerk, modernisering, ontbreken van een toetsvisie, kennisontwikkeling, inmenging door schoolleiding, actualiteit in toetsvragen, nieuw op school of nieuw vak, kwaliteitsprobleem vorige docent, kleine klassen/betrouwbaarheid, in ontwikkeling, onduidelijke syllabus, eenmanssectie, rare schoolregels en duidelijkere informatie van SLO.

De laatste gesloten vraag betrof verbeterpunten voor de invulling van het SE op de eigen school. Er waren hierbij elf antwoordmogelijkheden waarbij de examencommissieleden meer vakjes aan konden kruizen (zie Bijlage 3, p29). Het meest genoemde verbeterpunt door examencommissieleden is het gebrek

aan tijd bij docenten (71,8%), gevolgd door gebrek aan tijd bij de examencommissie (54,1%) (zie bijlage 5, Tabel 5.17). Daarnaast noemt 31,8% van de respondenten te weinig regelgeving of sturing (op de eigen school). Het ontbreken van duidelijke afspraken binnen de school is volgens 29,4% van de respondenten een verbeterpunt. Onvrede over te veel ruimte voor invulling van het PTA door vaksecties komt naar voren bij 20,0%, en ontevredenheid over de afstemming met de vaksecties bij 18,8%. Verder noemen respondenten weinig deskundige vaksecties (11,8%), te weinig mogelijkheden voor bijscholing (10,6%), te veel regelgeving of sturing vanuit school (5,9%) en geen ruimte voor eigen invulling van het PTA en SE door vaksecties (3,5%).

Op de vraag wat het SE nog zou kunnen verbeteren hebben respondenten vooral behoefte aan meer tijd en middelen voor ontwikkeling, nakijken, overleg en analyse van SE's; dit is het meest genoemde verbeterpunt en omvat ook verzoeken voor structurele sectietijd, extra vergoeding en aangepast toetsbeleid voor specifieke leerstromen. Daarnaast is er regelmatig vraag naar duidelijke richtlijnen en beleid vanuit schoolleiding of landelijke instanties over inhoud, normering en verhouding SE/CE, en naar meer structureel overleg binnen vakgroepen en tussen scholen. Veel docenten willen een landelijke databank of itembank met kwalitatief goede vragen en opdrachten en geven daarnaast vaak aan behoefte te hebben aan bijscholing in toetsconstructie, vakdidactiek en exameneisen. Ook genoemd worden wensen voor aanpassing van onderwijsinhoud, meer inspiratie/uitwisseling en voorbeelden/materialen, en oplossingen voor uitdagingen door AI; enkele respondenten vragen om meer leidinggevende steun of kwaliteitschecks door meerdere beoordelaars.

3.1.8 Belangrijkste resultaten en conclusies uit het vragenlijst onderzoek

Het vragenlijstonderzoek heeft al met al een aantal inzichten opgeleverd. In deze paragraaf worden de belangrijkste hiervan puntsgewijs per thema samengevat.

Constructie

- Vrijwel alle docenten gebruiken schriftelijke toetsen; bij de meerderheid van de vakken bestaat het SE vooral uit schriftelijk afgenomen toetsen; het gebruik van niet-schriftelijke vormen is veelal vakgebonden, zoals mondelinge afnamen bij talen en praktijktoetsen bij cultuurvakken/vmbo.
- Toetsvorm wordt primair gekozen op basis van inhoud, secundair op praktische omstandigheden en gewoonte. Als bijkomende redenen worden gemak van grootschalige afname en gewoonte genoemd.
- De keuze voor oude CE-opgaven is vooral vanwege het gemak; docenten ervaren een snelle manier om inhoudelijk kwalitatief goede toetsen te maken. Ze erkennen dat dit wel soms het schooleigen karakter beperkt. Veel docenten gebruiken ook zelf ontwikkelde opgaven voor meer schooleigenheid.
- Examencommissies hebben vaak beperkt zicht op gebruikte bronnen en verhouding CE/SE.
- Samenwerking vindt vooral plaats in kleine duo's binnen vaksecties; brede samenwerking is zeldzaam.
- Kwaliteitsborging gaat vooral via collegiale beoordeling, beoordelingsmodellen; er is een beperkt gebruik van toetsmatrijzen.
- Toetsmatrijzen zouden inhoudsvaliditeit, consistentie en transparantie kunnen versterken.

Regels en richtlijnen

- SE-toetsen krijgen vaak verschillend gewicht in het PTA; weging wordt bepaald door hoeveelheid/inhoud stof, leerjaar, toetsvorm, moeilijkheidsgraad.

- Er zijn verschillen tussen schooltypen. Bij vmbo bb/kb is vaker gelijke weging; bij havo/vwo vaker variatie.
- Inhaal/herkansing heeft meestal andere opgaven, maar in vmbo vaak identiek aan oorspronkelijke toets. Zeker wanneer dat schriftelijke toetsen betreft, is dat zorgelijk omdat het betekent dat de leerlingen die inhalen/herkansen de opgaven al kunnen kennen.
- Kwaliteitscontrole vooraf door examencommissies is beperkt; de helft controleert niet structureel. Waar controle plaatsvindt, richt deze zich op aansluiting met lesstof, leerdoelen, taalniveau, moeilijkheid — vaak zonder vaste procedure.

Leerlingen

- Leerlingen weten meestal vooraf beoordelingswijze en toegestane hulpmiddelen, wat bijdraagt aan duidelijkheid en voorbereiding.
- Tijdsgebrek bij leerlingen komt in ongeveer 10% van de gevallen regelmatig voor; in bijna 50% zelden/nooit.

Beoordelen

- Docenten treffen veel praktische maatregelen om objectiviteit bij de beoordeling van de SE's te waarborgen. Dat kan gaan om de afstemming van de toets op de leerstof, om duidelijke beoordelingscriteria, of collegiale afstemming.
- Uniforme procedures voor normering en cijferbepaling ontbreken vaak; beoordeling en normbepaling per school en per toets zijn zodoende wisselend en vinden soms ad hoc plaats.
- Noodzaak voor gefundeerde richtlijnen voor score-naar-cijfer omzetting en cesuurbepaling.
- Tweede correctie (steekproefsgewijs) kan betrouwbaarheid vergroten.

Evaluatie en deskundigheid

- Evaluatie van SE's vooral binnen vaksecties en richt zich vooral op dekking eindtermen, kwaliteit opgaven, niveau.
- Structurele schoolbrede of externe evaluatie en gebruik van analysegegevens beperkt.
- Deskundigheid van docenten wordt hoog ingeschat, maar kennis over toetsmatrijzen is vaak minder. Daar ligt ruimte voor verbetering.
- Deskundigheidsbevordering vooral via interne kennisuitwisseling; formele training en gebruik actuele exameninformatie is er wel maar minder vaak.

Positieve punten en verbeterpunten

- Positief: grote vrijheid bij invullen PTA/SE, deskundigheid binnen vaksecties, interne afstemming en mogelijkheden voor bijscholing.
- Groot knelpunt: ervaren tijdgebrek voor toetsontwikkeling, nakijken, overleg en analyse. Dit is het meest genoemde verbeterpunt, zowel door docenten als door examencommissieleden.
- Behoefte aan structuur en ondersteuning: duidelijke afspraken/richtlijnen vanuit school voor weging, normering, verhouding SE/CE.
- Andere wensen: gedeelde itembanken, betere afstemming tussen secties en examencommissie, aandacht voor AI-impact, tweede correctie, en kwaliteitsverbetering.

3.2 Resultaten van de evaluatie van SE's

3.2.1 Inventarisatie van de gedeelde materialen

3.2.1.1 Aanwezigheid van materiaal

In totaal zijn er 18 verschillende pakketten ingediend afkomstig van 15 verschillende scholen. Wanneer een school SE's aanleverde van meerdere vakken of meerdere schooltypen, zijn de examens opgedeeld in twee pakketten en is dit aangegeven met een A en een B bij de pakketnaam. Op deze manier bevat elk pakket SE's van een afzonderlijk vak en schooltype. In Tabel 3.2.1.1 is het overzicht gegeven van de ingediende pakketten, inclusief bijbehorende vakken, schooltypen en het aantal SE toetsen. In de laatste kolom is aangegeven hoeveel praktische opdrachten er in elk pakket zitten, waarbij een toets als een praktische opdracht is gerekend als dit ofwel zo door de docent is aangegeven bij het betreffende SE, ofwel als het thema "Praktijkopdrachten" is ingevuld door de inhoudsdeskundigen en het SE dus als een praktische opdracht werd beschouwd door de beoordelaar.

Tabel 3.2.1.1 Overzicht van de gedeelde pakketten met SE toetsen

Pakket	Vak	Schooltype	Aantal SE toetsen	Aantal praktische opdrachten in pakket
Pakket 01	aardrijkskunde	havo	2	0
Pakket 02	wiskunde	vmbo[54]	7	0
Pakket 03	geschiedenis	havo	6	1
Pakket 04A	natuurkunde	havo	3	0
Pakket 04B	natuurkunde	vwo	3	0
Pakket 05	wiskunde	vmbo	10	1
Pakket 06	scheikunde	havo	6	3
Pakket 07A	Duits	havo	6	5
Pakket 07B	Duits	vwo	5	3
Pakket 08A	Engels	vmbo gl/tl	1	0
Pakket 08B	Nederlands	vmbo gl/tl	1	0
Pakket 09	Nederlands	vmbo	6	5
Pakket 10	Engels	vmbo kbl	6	5
Pakket 11	wiskunde B	havo	3	0
Pakket 12	Nederlands	vmbo tl	4	2
Pakket 13	natuurkunde	vwo	6	2
Pakket 14	Grieks	vwo	2	1
Pakket 15	biologie	vmbo kbl	1	1

Afkortingen bij vmbo kbl: kaderberoepsleerweg; gl: gemengde leerweg; tl: theoretische leerweg.

Er zijn in totaal 78 SE toetsen gedeeld. Dat betekent dus een gemiddelde van ruim 4 SE's per pakket. Het aantal SE's per pakket varieerde van één (bij drie pakketten) tot en met tien. In Bijlage 8 is het volledige overzicht gegeven van alle 78 gedeelde SE's.

[54] Bij de vmbo examens is een deel ingediend als een verzameling voor alle leerwegen, met al dan niet onderscheid tussen leerwegen binnen het pakket. Wanneer het pakket dat ingediend was een specifieke leerweg betrof, is dat toegevoegd bij het schooltype vmbo.

In Tabel 9.1 in Bijlage 9 zijn de aangeleverde materialen per SE toets gegeven. Deze informatie is beschikbaar over alle 78 toetsen heen en per pakket. De reden om dit ook per pakket aan te geven is dat er zulke verschillen in de aantallen toetsen per pakket zitten. Met tien toetsen in Pakket 05 heeft de docent of het docententeam een grotere invloed op de resultaten dan bijvoorbeeld de docenten of docententeams van pakket 08A, 08B en 15 met slechts één toets. De verschillen in percentages zijn over het algemeen niet groot. Het grootste verschil zit bij de informatie over het PTA, wat logisch is aangezien een PTA over een geheel pakket gaat en er geen PTA per toets is. Bij minder dan de helft van de pakketten is een PTA bijgevoegd.

Het is opvallend dat er bij één van de SE toetsen (1%; n=78) geen opgaven als document waren bijgevoegd. Dit betreft een toets waarbij enkel een voorblad met instructies en enkele formules is aangeleverd. De betreffende beoordelaar heeft dus maar een klein gedeelte van het beoordelingskader kunnen invullen voor dit SE en de rest van de vragen gemarkeerd als “niet van toepassing” (n.v.t.). Bij de meeste toetsen is een antwoord- of beoordelingsmodel aanwezig (88%; n=78), evenals een instructie of toelichting voor de kandidaat (85%; n=78). Verder blijkt dat er weinig toetsmatrijzen zijn aangeleverd, namelijk bij slechts 15% (n=78) van de toetsen. Ook is er slechts bij 14% (n=78) van de toetsen een alternatieve versie of herkansingsvariant bijgevoegd. Er is door geen enkele school een beoordelingsprocedure ofwel visie op toetsen gedeeld (beide 0%; n=78). Omdat er geen enkele beoordelingsprocedure is aangeleverd, kan in de resultaten ook niks worden beschreven over de criteria behorend bij dit onderdeel en zijn deze in Tabel 9.6 ingevuld als ‘n.v.t.’.

3.2.1.2 Aanvullende aangeleverde informatie per SE

Als we kijken naar de aanvullende aangeleverde informatie per SE (zie bijlage 9, Tabel 9.2) dan zien we dat bij 82% (n=78) van de SE toetsen de tijdsduur duidelijk vermeld staat bij de toets. In de meeste gevallen is er een aanwijsbare reden voor het ontbreken van de beschikbare tijdsduur, omdat er dan sprake is van praktische toetsen (zonder tijdsdruk). Bij één pakket ontbreekt de beschikbare tijdsduur bij allebei de toetsen zonder aanwijsbare oorzaak, en bij een ander pakket geldt dit voor vier van de in totaal zes toetsen.

In de meeste gevallen is het aantal vragen of opdrachten in de toets duidelijk aangegeven. Bij sommige toetsen is dit niet het geval, bijvoorbeeld omdat het een praktische opdracht is of omdat de opdracht in de vorm van een PowerPointpresentatie is aangeleverd en er in dit format geen ruimte is vrijgemaakt om expliciet duidelijk te maken hoeveel vragen/opdrachten in de toets zitten.

Er wordt in bijna één op de vijf SE’s niet aangegeven hoe van itemscores naar totaalscore te komen. Dit is meestal het geval bij praktische opdrachten. Het grootste deel van de 15% waarbij geen (minimaal en) maximaal te behalen puntenaantal vermeld staat, bestaat ook uit praktische opdrachten. Verder wordt duidelijk dat de toegestane hulpmiddelen in 87% (n=78) van de SE’s zijn vermeld. De overige 13% bevat onder andere een geschiedenistoets en twee scheikunde praktische opdrachten, waarbij je wel zou verwachten dat de toegestane hulpmiddelen relevant zijn om expliciet te benoemen. Het percentage bij toegestane hulpmiddelen op pakketniveau ligt hoger, wat kan worden verklaard doordat deze informatie verder in één pakket bij vijf van de zes SE’s Engels ontbreekt.

Tot slot blijkt dat de cesuur (grens voor voldoende/onvoldoende) slechts in één van de vier gevallen vermeld wordt. Er zijn drie pakketten, twee van één SE en één van vijf SE’s, waarin het consequent wordt vermeld. Voor de rest van de pakketten geldt dat het ofwel helemaal niet, ofwel bij slechts een deel van de SE’s wordt aangegeven.

3.2.2 Inhoudelijke beoordelingen

3.2.2.1 Vorm, inhoud, dekking en niveau

In Tabel 9.3 in Bijlage 9 staat een overzicht van de resultaten voor het thema “Vorm, inhoud, dekking en niveau”. Om de resultaten te duiden wordt gebruikgemaakt van de opmerkingen van de inhoudsdeskundigen bij de betreffende criteria, evenals de eerste open vraag “Geeft het SE de indruk dat er een eigen visie achter ligt? Waarop baseer je deze indruk?” en de tweede open vraag “Kan je een toelichting geven op je antwoorden op herkomst, opbouw en kwaliteit van de vragen in het SE?”. Bij de eerste open vraag is het belangrijk om op te merken dat het woord *visie* door beoordelaars verschillend is geïnterpreteerd: door sommige beoordelaars als de mate waarin de schoolidentiteit en onderwijsvisie terug te zien is in het SE, en door andere beoordelaars als een toetstechnische visie. In deze sectie worden enkele opvallende resultaten uitgelicht.

Allereerst konden bij het criterium “Dekt de toets-inhoud de leerdoelen uit het PTA volledig af?” slechts 34 SE’s worden beoordeeld, omdat in veel gevallen het PTA ontbrak of omdat er geen expliciete leerdoelen in het PTA staan. Van de beoordeelde SE’s kreeg 41% een positief oordeel. Beoordelaars geven regelmatig aan dat onderdelen uit het PTA ontbreken in de SE’s en/of dat een toets zich vooral op een deelthema concentreert. Voorbeelden zijn geschiedenistoetsen waarbij bepaalde tijdvakken nauwelijks aan bod komen en talige SE-pakketten waarbij onderdelen zoals ‘Fictie’ en ‘Spreken’ ontbreken. Hierbij moet wel opgemerkt worden dat we in veel gevallen ook niet alle SE’s ontvangen hebben, en dat onderdelen van het PTA ook in niet aangeboden SE’s behandeld zijn.

De verdeling van vragen over de voorgeschreven onderwerpen binnen de SE’s zelf werd in iets meer dan de helft van de gevallen als voldoende evenwichtig gezien (54%; n=57). De toelichtingen van de beoordelaars laten een aantal terugkerende knelpunten zien. Meerdere SE’s blijken sterk te focussen op één onderdeel, bijvoorbeeld veel algebra ten koste van meetkunde bij wiskunde en oververtegenwoordiging van ontbossing/biodiversiteit bij aardrijkskunde. Hierdoor blijven andere voorgeschreven onderwerpen onderbelicht. Ook ontbreken bij enkele toetsen relevante domeinen of zijn onderdelen slechts summier vertegenwoordigd. Verschillende beoordelaars geven aan dat het lastig is om de verdeling objectief te beoordelen bij de afwezigheid van een toetsmatrijs. Hier komen we later in de resultaten op terug (sectie “2e. Toetsmatrijs”).

Het beoordelen van de mate waarin een eigen schoolvisie mede vormgaf aan het SE was vaak lastig te beoordelen. Meerdere beoordelaars gaven aan de vraag over de schoolvisie niet te kunnen invullen, omdat er geen schoolvisie was aangeleverd. Van de 56 SE’s die op deze vraag beoordeeld waren was bij 36% de indruk dat er een eigen schoolvisie achter het SE lag. Bij het onderbouwen van hun indruk merken de beoordelaars op dat veel SE’s sterk zijn gebaseerd op de methode of op (oude) CE-opgaven: “Deze toets lijkt onaangepast van een methode te komen”, “De SE’s zijn duidelijk op de leest van oudere examens geschoeid. Een eigen visie durf ik niet te zeggen”, en: “Nee – het SE lijkt op vorm, onderwerp en vraagsoorten uit CE. (…) Er zijn veel vragen die of letterlijk uit een oud CE komen of die daarop gebaseerd zijn.” De eigen bedoeling en richting van de school wordt dus vaak gemist: “Het is een erg schools, op de boeken, gericht SE. Weinig diepgang en echte toepasbare kennis. De eigen ‘handtekening’ vind ik ontbreken.” Tegelijkertijd zijn er ook voorbeelden waarin bewust keuzes zichtbaar zijn die een visie kunnen weerspiegelen: een expliciete nadruk op bepaalde vaardigheden (bijvoorbeeld hogere eisen aan taalverzorging), keuze voor praktische opdrachten die een thema extra gewicht geven of de opbouw van SE’s van eenvoudig naar complex.

Ten slotte scoren onder meer de volgende drie criteria hoog: 92% van de beoordeelde SE’s (n=72) kreeg een positief oordeel voor de vraag of de vorm van het SE passend is gegeven het te meten doel, 93% (n=71) dat de toets aansluit bij het leerjaar en niveau van de opleiding en 86% (n=73) dat de moeilijkheidsgraad ook past bij dat leerjaar/niveau. Beoordelaars geven aan dat de gekozen vorm (schriftelijk voor kennis/leesvaardigheid, praktische opdrachten voor praktische vaardigheden) in de

meeste gevallen goed aansluit bij wat wordt gemeten. Ook lijkt de inhoud qua niveau overwegend te passen bij het curriculum en de opleidingsfase. Wel noemen beoordelaars enkele uitzonderingen, zoals dat een toets soms te veel gericht is op feitenkennis ("De toets blijft te veel hangen in memoriseren") en dat bij een deel juist hogere denkvaardigheden of een betere afstemming van cesuur/weging wenselijk zijn.

3.2.2.2 Transparantie en standaardisatie

Bij het thema "Transparantie en standaardisatie" gaat het erom dat leerlingen in de voorbereiding en tijdens de afname onder gelijke omstandigheden de SE's maken. De resultaten van de tien criteria voor dit thema staan in Tabel 9.4 in Bijlage 9. Van de zes criteria bij transparantie zijn er vijf voor bijna alle SE's ingevuld. Het laatste criterium is in 35% van de gevallen niet beoordeeld, wat mogelijk veroorzaakt is doordat dit het enige criterium was bij dit sub-thema dat door de inhoudsdeskundigen zelf moest worden ingevuld en door hen over het hoofd is gezien.

Uit de resultaten blijkt dat bij slechts 30% van de beoordeelde SE's (n=77) vermeld is wat voor voorbereiding bij deze vorm van toetsen past. Dit wil niet zeggen dat het bij de overige 70% nooit gecommuniceerd is aan de leerlingen, maar laat wel zien dat het niet expliciet op het SE zelf is vermeld. De puntentelling per item op de SE's wordt in 84% van de SE's (n=77) duidelijk vermeld, maar niet altijd bij alle opdrachten. Niet alle SE's geven daar ook bij aan hoe de totaalscore tot stand komt (73%; n=77). Naast deze totaalscore en puntentelling moet er een cesuur zijn in een toets. Deze is echter over het algemeen weinig gegeven (24%; n=78) en voor de leerlingen maar in 6% (n=77) van de gevallen bekend. De meeste leerlingen weten dus niet voordat zij het SE gaan maken vanaf welke score zij een voldoende hebben. Daarentegen zijn de hulpmiddelen (86%; n=76) en de instructie (90%; n=51) bij de schriftelijke toets of praktische opdracht in de meeste gevallen wel duidelijk voor de leerling. Beoordelaars geven bijvoorbeeld aan: "Per vraag is helder wat er moet gebeuren" en "Er staat duidelijk op het voorblad wat de toegestane hulpmiddelen zijn en hoe de vragen dienen te worden beantwoord." Het is opvallend dat juist bij de praktische opdrachten vaak de toegestane hulpmiddelen ontbreken, zoals bij scheikunde of natuurkunde.

De standaardisatie is op vier criteria beoordeeld. Het blijkt dat er nauwelijks instructies zijn gegeven over hoe het SE moet worden afgenomen (14%; n=77). Het valt op dat de volgende drie vragen die dieper ingaan op de instructies voor meer SE's zijn beantwoord (n=23) dan het aantal SE's waarvoor is aangegeven dat er instructies zijn. Beoordelaars zijn positief over de duidelijkheid van de instructies (87%) en of deze voor alle leerlingen naar redelijkheid gelijk zijn (100%). Een beoordelaar merkt bijvoorbeeld op: "Wat er staat, is inhoudelijk duidelijk". Bij sommige beoordelaars bestaat wel twijfel over de vergelijkbaarheid tussen leerlingen: "De instructies zijn wel oké, maar standaardisatie is lastig bij een open schrijfopdracht. Er zit hier en daar wat overlap tussen de stappen." Tot slot is het in bijna alle gevallen niet duidelijk wanneer er uitzonderingen mogen zijn (96%; n=25).

3.2.2.3 Vraagformulering en taalgebruik

In Tabel 9.5 (Bijlage 9) staat een overzicht van de resultaten voor het thema "Vraagformulering en taalgebruik", waarin middels 11 criteria is geëvalueerd in welke mate er onduidelijkheden in de vraagstelling gevonden worden, en het taalgebruik van de vragen passend is bij het schooltype waar het SE voor bedoeld is. De meeste SE's waren beoordeelbaar op alle criteria, behalve de laatste twee criteria over het gebruik van illustraties. Hier was 28% van de SE's niet beoordeelbaar, omdat er geen tabellen of grafieken in voorkwamen.

De lay-out werd overwegend als duidelijk, overzichtelijk en functioneel ervaren (86%; n=77). Meerdere toetsen bevatten wel onduidelijkheden rond de voortgangsindicatie: vragen lopen door naar de volgende pagina zonder waarschuwing en er ontbreekt soms een expliciete aanduiding dat de toets is afgelopen (67%; n=73). Verder werd een groot deel van de toetsen gescoord als grammaticaal correct en

helder (81%; n=75): “Meestal wel, paar uitzonderingen” en “Het Engels is niet helemaal perfect, maar wel oké.” Toch komen er ook heel wat fouten in de SE’s voor. Beoordelaars merken bijvoorbeeld op: “Soms zitten er slordigheidsfoutjes in”, “Er zitten een aantal spelfouten en formuleringsfouten in de toets”, en “Helaas is het Nederlands soms moeilijk”. Ook signaleren beoordelaars gevallen van onduidelijke instructiewoorden of verwijzingen binnen vragen (72%; n=75), wat zichtbaar wordt in opmerkingen als: “Wat wordt precies bedoeld met het stukje tussen haakjes?” en: “De vragen worden niet altijd helder gesteld (ontbreken van instructiewoorden.)”

Het criterium “Dubbele ontkenningen en verwarrende formuleringen” scoort hoog (89%; n=71), wat betekent dat veel toetsen dergelijke constructies vermijden. Waar problemen optreden, gaat het om concrete onduidelijkheden in woordgebruik of verwijzing. Een sterke kant van de onderzochte SE’s is daarnaast de zinslengte (95%; n=74) en de afstemming op de doelgroep (99%; n=75): vrijwel alle toetsen vermijden onnodig lange zinnen en het taalgebruik is bijna altijd passend voor de doelgroep.

De toegankelijkheid voor leerlingen met dyslexie of een NT2-achtergrond werd daarentegen niet altijd als positief ervaren. Deze is in 80% (n=74) van de gevallen positief beoordeeld, maar slechts in 68% van de gevallen op pakketniveau. Veel terugkomende aandachtspunten waren het inconsistent gebruik van opsommingsstijlen, cursief gedrukte tekstgedeelten, moeilijke woorden of formele en vrij lange formuleringen die moeilijker zijn voor NT2-lezers of dyslectische leerlingen. Dit hangt deels samen met het criterium voor overbodige of onnodige tekst (84%; n=74): sommige toetsen zijn heel talig en opmerkingen tonen dat bepaalde bronnen of lange teksten niet bijdragen aan beantwoording van vragen.

Tot slot blijkt dat, waar beoordeeld, afbeeldingen en tabellen in veel gevallen als functioneel en duidelijk worden ervaren (93%; n=56), maar dat er ook voorbeelden zijn van onduidelijke bijschriften, klein afgedrukte bronnen of niet-toereikende toelichtingen bij figuren. Ook komt het voor dat kleurgebruik in de kaarten verwarrend is of onjuist. In sommige gevallen zou extra illustratiemateriaal van toegevoegde waarde zijn ter ondersteuning.

3.2.2.4 Beoordelingsmodel en -procedure

In Tabel 9.6 van Bijlage 9 staat een overzicht van de resultaten voor het thema “Beoordelingsmodel en -procedure”. Er was niet bij alle SE’s een antwoord- of beoordelingsmodel bijgevoegd, waardoor het aantal beantwoorde vragen over de criteria varieert van 58-66 SE’s. Het criterium dat het hoogst wordt beoordeeld is of de beoordeelde criteria aansluiten bij de eindtermen en de vragen (85%; n=62). Dit suggereert dat bij veel toetsen de beoordeelde aspecten relevant zijn voor wat volgens het curriculum getoetst moet worden. Wel zijn regelmatig opmerkingen gemaakt over ontbrekende elementen of variatie in detailniveau. Ook zijn er enkele gevallen waarin beoordelaars aangeven dat onderdelen uit de eindtermen niet zijn vertaald naar concrete beoordelingscriteria. Ook de inhoudelijke aansluiting van het antwoord- of beoordelingsmodel wordt redelijk goed beoordeeld (78%; n=59), wat betekent dat in de meerderheid van de gevallen het beoordelingsmodel de gevraagde prestaties adresseert.

Op de overige vier criteria worden duidelijke tekortkomingen gesignaleerd. De beoordelingsmodellen bevatten in slechts 67% (n=64) van de gevallen alle correcte antwoorden of beoordelingscriteria. Uit de opmerkingen van de inhoudsdeskundigen wordt duidelijk dat dit samenhangt met in hoeverre het duidelijk is hoe alternatieve of deels juiste antwoorden worden beoordeeld (60% positief; n=58). “Niet alle mogelijke oplossingen zijn altijd beschreven”, “Geen voorbeeldantwoorden bij open vragen”, “Ik kan nergens lezen wat er allemaal aan antwoorden gegeven had kunnen worden en betwijfel of de docent zo’n overzicht heeft” en “Het is niet altijd duidelijk welke componenten een antwoord **in elk geval** moet bevatten en/of wanneer het geen (of misschien de helft van de) punten krijgt”.

Ook de transparantie en logica in de puntentoekenning wordt in slechts iets meer dan de helft van de gevallen positief beoordeeld (56%; n=66). Beoordelaars geven aan dat de transparantie vaak wel in orde is, maar dat er verschillen bestaan tussen de hoeveelheid werk die een vraag vereist en het aantal

punten dat daarvoor wordt toegekend, zoals:, “Complexe redeneringen wegen even zwaar als eenvoudige feitelijke antwoorden” en “Transparant wel, logisch verdeeld niet”. Tot slot blijkt de weging tussen opgaven niet passend en vertekenend voor het resultaat in 34% (n=59) van de SE’s. In toelichtingen wordt genoemd dat sommige onderdelen relatief zwaar of juist licht meewegen ten opzichte van hun inhoudelijke relevantie, bijvoorbeeld: “De weging hangt vooral op het aantal vragen en niet op de inhoudelijke relevantie; de open vragen aan het eind leveren maar 5 punten op terwijl die relevanter zijn voor taalproductie.” Tot slot zijn er, zoals reeds in paragraaf “1a. Aanwezigheid van materiaal” beschreven, geen resultaten beschikbaar voor de criteria bij “Beoordelingsprocedure” en kan hier niks over geconcludeerd worden.

3.2.2.5 Toetsmatrijs

Tabel 9.7 in Bijlage 9 laat zien hoe de toetsen beoordeeld zijn voor het thema “Toetsmatrijs”, waarbij een belangrijk gegeven is dat deze beoordeling slechts bij negen SE’s is gedaan, omdat in de andere gevallen de toetsmatrijzen en/of PTA’s ontbreken. Uit de gegevens blijkt dat bij 33% (n=9) van de gevallen de toetsmatrijs de toetsvragen op een duidelijke manier koppelt aan leerdoelen of eindtermen. In geen van de gevallen worden de beheersingsniveaus en weging volgens de inhoudsdeskundigen correct weergegeven in de toetsmatrijs. Ter illustratie: bij vier SE’s Nederlands, waarbij een toetsmatrijs is aangeleverd, is de beoordeling niet positief. De inhoudsdeskundige geeft aan dat ze de gedachtegang achter de matrijs niet begrijpt: “hoe is deze tot stand gekomen?” Ze stelt verschillende vragen over de keuzes die gemaakt zijn: “(…) Ook ‘tekstbegrip’ is als apart kopje opgenomen in de matrijs. Ik vraag me af of dit handig is. Het gaat immers om een leestoets waarbij het om “tekstbegrip” gaat. (…) De gedachte hierachter is me niet duidelijk. Op deze manier geeft de matrijs ook weinig sturing.” De vraag over de inhoudelijke aansluiting van de matrijs op het PTA en examenprogramma is slechts twee keer beantwoord, met “ja”. De inhoudsdeskundige heeft hier geen toelichting bij gegeven.

3.2.2.6 Validiteit en gelijke kansen

Tabel 9.8 (Bijlage 9) laat zien hoe de toetsen beoordeeld zijn voor het thema “Validiteit en gelijke kansen”, waarbij middels drie criteria is onderzocht of er voldoende verschillende vaardigheden of competenties aan bod komen en of deze inclusief genoeg gemeten worden. De meeste SE’s scoren positief op dit thema, maar er zijn enkele belangrijke en terugkerende aandachtspunten die een probleem kunnen vormen voor de validiteit en gelijke kansen.

Allereerst geven veel beoordelingen aan dat SE’s neutrale en acceptabele opgaven hebben, zonder een opvallende bias (88%; n=69). Een beoordelaar merkt bijvoorbeeld op: “Schrijftaken zijn heel realistisch waar ieder leerling zich wel iets bij kan voorstellen.” Toch worden bepaalde knelpunten regelmatig gesignaleerd, zoals culturele bias, met name bij geschiedenistoetsen. Er komen eurocentrische formuleringen voor, als er bijvoorbeeld wordt gesproken over “de ontdekking van Brazilië” of het routinematig gebruik van de termen als “Gouden Eeuw”, zonder nuancering. Verder bleken bronnen vaak eenzijdig van aard, waarover een beoordelaar bijvoorbeeld schrijft: “Alle bronnen zijn van “mannelijke westerse” afkomst. Weinig diversiteit.” Een andere beoordelaar komt tot dezelfde conclusie: “Alle bronnen zijn van mannelijke hand. Geen bronnen vanuit een ander cultuur o.i.d.”

Verder blijkt dat opgaven in veel gevallen zo zijn opgesteld dat mogelijke (fysieke/lichamelijke) belemmeringen een optimale prestatie niet in de weg staan (89%; n=63). Sommige beoordelaars hebben “ja” ingevuld, met de opmerking erbij dat de opgaven met passende aanpassingen wel toegankelijk gemaakt kunnen worden, bijvoorbeeld door toetsen mondeling af te nemen. Toch zijn er enkele terugkerende verbeterpunten: lay-out, te compact bladbeeld, klein of cursief lettertype en het gebruik van bepaalde tabellen of visuele bronnen verminderen de leesbaarheid en kunnen leerlingen met visuele beperkingen of dyslexie belemmeren. Zoals een beoordelaar formuleerde: “De lay-out zou belemmerend kunnen zijn voor kandidaten met een visuele beperking of met dyslexie.” Ook praktische opdrachten en

visuele taken blijken soms moeilijk uitvoerbaar voor leerlingen met lichamelijke beperkingen wanneer er geen valide alternatieve of aanpassingen worden aangeboden: “Dat is denk ik onmogelijk om hier te garanderen: iemand in een rolstoel kan misschien niet goed titreren[55] of aflezen, maar daar valt wel wat op te bedenken.”

Ten slotte is bij dit thema uitgevraagd in hoeverre er verschillende vaardigheden of competenties aan bod komen in de SE’s. In de meeste gevallen hebben de pakketten een positieve beoordeling gekregen, zonder verdere toelichting van de inhoudsdeskundigen (92%; n=61). Sommige SE’s toetsen meerdere competenties. Er zijn veel SE’s waarin slechts één vaardigheid wordt geadresseerd, zoals schrijven, maar beoordelaars geven aan dat andere vaardigheden zoals lezen en spreken dan wel aan bod komen in de andere SE’s in hetzelfde pakket.

3.2.2.7 Scoren en normeren

Tabel 9.9 in Bijlage 9 laat zien hoe de toetsen beoordeeld zijn voor het thema “Scoren en normeren”. Het onderdeel “Soort scoretoekenning per opgave” is met “ja” gescoord als deze gelijk was over de opgaven en met “nee” als de scoretoekenning verschilde over opgaven/criteria. Uit deze resultaten blijkt dat in 92% van de SE’s (n=72) de scoretoekenning niet uniform is over alle opgaven/criteria. In een substantieel deel van de toetsen (75%; n=56) wordt de scoretoekenning per opgave/criteria als helder en verdedigbaar beoordeeld, maar er is ook een substantieel deel waarbij dat niet het geval is. Verschillende beoordelaars geven specifieke voorbeelden van een onlogische of disproportionele puntentoekenning bij enkele opgaven, die het totaaloordeel van een leerling kan beïnvloeden.

Bij het onderdeel “Weging” is op eenzelfde manier gescoord als bij “Soort scoretoekenning per opgave”. Het beeld dat ontstaat lijkt sterk op dat van scoretoekenning: de toegepaste weging verschilt over opgaven en criteria (9%; n=70). In een groot deel van de toetsen wordt de weging als logisch en passend ervaren. In 31% van de SE’s (n=54), een iets hoger percentage dan bij scoretoekenning, wordt de weging als niet logisch of passend bij de dekking bevonden. Inhoudsdeskundigen wijzen op concrete voorbeelden bijvoorbeeld: “De toets *[lees: praktische opdracht voor het vak Engels]* weegt enkel mee terwijl grammaticatoetsen en de leestoets dubbel wegen. Dat is jammer, want de productieve vaardigheden komen niet terug in het CE en lezen wel.” En: “Enige toets over deze eindterm, heeft slechts weging 7%.”.

De gestelde cesuur kon bij slechts een klein deel van de SE’s beoordeeld worden (n=17). Binnen de set SE’s waar de cesuur wel beoordeeld kon worden, is aanzienlijke verbetering mogelijk: circa 40% van de gevallen werd door inhoudsdeskundigen als niet logisch of passend gezien. Citaten uit beoordelingen benadrukken de onduidelijkheid en ontbrekende onderbouwing: “Aan de scores is te zien dat de grote meerderheid van de leerlingen de toets voldoende heeft gemaakt. Ook zijn de eindtermen vrij algemeen en niet omgezet naar een “lat” die leerlingen moeten halen.” En: “Het is niet helemaal duidelijk wat de cesuur is voor het combinatiecijfer en voor de deelcijfers kon ik de cesuur aflezen uit de cijferlijst. Gezien de scores lijkt de cesuur en of de toets zelf voor idioom enigszins soepel en de cesuur voor lezen tamelijk streng.”

3.2.2.8 Praktijkopdrachten

Tabel 9.10 (Bijlage 9) laat de resultaten zien voor het thema “Praktijkopdrachten” en bijbehorende criteria. Er zijn 24 SE’s waarvoor alle criteria zijn beoordeeld. De drie criteria zijn voor 75-81% positief beoordeeld. De praktijkopdrachten worden in veel gevallen gezien als representatief voor een realistische beroeps- of leersituatie (n=26). Verschillende beoordelaars noemen dat opdrachten aansluiten bij dagelijkse of beroepsmatige situaties: “Het past in het leven van de kandidaten” en “De opdrachten zijn heel realistisch en ik denk dat vrijwel alle leerlingen vanuit hetzelfde startniveau aan de opdracht

[55] ‘Titreren’ is een scheikundige term en verwijst naar de handeling waarbij wordt bepaald wat de concentratie van een bepaalde stof in een oplossing is.

beginnen, voorkennis (of het ontbreken ervan) speelt nauwelijks een rol." Tegelijkertijd wijzen beoordelaars op opdrachten waarbij dit beter kan, bijvoorbeeld bij een praktische spreekopdracht voor het vak Engels: "Woordenschat en grammatica nuttig, maar niet heel relevant voor het meten van een productieve vaardigheid."

De beoordelingscriteria worden vaak als observeerbaar, objectief en inhoudelijk relevant gezien (n=24), met soms de kanttekening dat er wel een ervaren beoordelaar vereist is. Tegelijkertijd is er een terugkerende zorg over de objectiviteit en concrete formulering van bepaalde beoordelingscriteria. Een beoordelaar merkt op: "Objectief is wel ingewikkeld: in de rubric wordt gesproken over: "er wordt echt een betoog opgezet". Wanneer is het echt?" Anderen wijzen op de inherent subjectieve aard van sommige beoordelingen: "Beoordeling van schrijfvaardigheid blijft ondanks rubrics toch ook een subjectieve maat" en "Dat is erg moeilijk, objectiviteit bij mondeling. Twee beoordelaars is al fijn."

Ook het beoordelingsformulier wordt in veel gevallen als duidelijk en hanteerbaar in gebruik ervaren (n=27). Toch is er forse kritiek op enkele formulieren, onder andere op onduidelijke schaalindelingen en gebrek aan instructie voor beoordelaars: "De opdeling van de 0-3 en 1-3 scorepunten is niet helder", "Het geeft geen ruimte voor alternatieve oplossingen die er absoluut wel kunnen zijn", en "Het beoordelingsmodel is niet zo overzichtelijk. Er is geen beoordelaarsinstructie. Ook is er niet aangegeven hoe je van score tot cijfer komt." Meerdere beoordelaars noemen onder andere expliciet dat voorbeeldantwoorden de consistentie van de beoordeling zouden verhogen.

3.2.3 Relatie tussen CE en SE

Om de relatie tussen CE's en SE's te evalueren, is gekeken hoeveel opgaven op welke manier aan het CE te relateren zijn. Hiertoe is het totale aantal vragen in het SE geteld. Vervolgens hebben de inhoudsdeskundigen de opgaven geclassificeerd naar de volgende opdeling met betrekking tot de relatie met het CE. Dit is een opdeling naar vijf typen opgaven: (1) aantal vragen (vrijwel) letterlijk overgenomen uit het CE; (2) aantal vragen sterk gebaseerd op CE-opgaven; (3) aantal vragen licht geïnspireerd op CE-opgaven; (4) aantal originele vragen (wel gerelateerd aan CE-inhoud) en (5) aantal vragen over inhoud die niet in het CE gedekt wordt. Per SE is bepaald welk percentage van de opgaven tot welke van deze vijf typen gerekend kan worden. De meest opvallende resultaten worden uitgelicht in deze sectie.

In tabel 9.11 van Bijlage 9 staat een overzicht met de percentages per type vragen gemiddeld over 68 SE's; bij 10 SE's was een dergelijk opdeling slecht te maken, waardoor deze niet meegenomen zijn in de analyses. Het blijkt dat over al deze SE's gezien, er gemiddeld 31% van de vragen rechtstreeks wordt overgenomen uit CE's (n=68). Binnen het vmbo ligt dit percentage het hoogst (44%, n=32), een stuk hoger dan bij havo/vwo (19%, n=36). Uitgesplitst naar type schoolvak, blijkt dat er binnen de science-vakken gemiddeld de meeste inspiratie vanuit CE's wordt gehaald: 70% van de vragen in deze SE's is letterlijk overgenomen uit CE's, of is sterk of licht gebaseerd op CE's (n=34). Voor de gamma-vakken (54%, n=6) en talen (39%, n=28) ligt dit percentage wat lager. Over alle SE's gezien bestaat gemiddeld 19% originele vragen over CE inhoud en 26% originele vragen over inhoud die niet in het CE gedekt wordt. Dat betekent dat in totaal iets minder dan helft van de vragen echt origineel is (45%, n=68). Kijken we specifiek naar het percentage originele vragen over niet-CE inhoud, dan is die het hoogst bij de talige schoolvakken (45%, n=28). Dat valt mogelijk te verklaren doordat de talige CE's voornamelijk uit leesvaardigheid bestaan, en docenten in SE's andere vaardigheden aan bod willen laten komen. Over alle SE's heen komen vragen die sterk gebaseerd zijn op CE's (11%, n=68) of licht geïnspireerd door CE's (14%, n=68) het minste voor.

Verder uitgesplitst, is er een verdeling te maken in drie soorten SE's: SE's waarbij geen van een bepaald type vragen voorkomt (zie Bijlage 9, Tabel 9.12), SE's waarbij alle vragen van een bepaald type zijn (zie Bijlage 9, Tabel 9.13) en SE's waarbij een deel, maar niet alle vragen van een bepaald type zijn (zie Bijlage 9, Tabel 9.14). Over alle SE's gezien zit er in 46% van de SE's, dus dat is bijna de helft van de

SE's, geen enkele vraag die letterlijk is overgenomen uit een CE. Bij de talige vakken ligt dit percentage het hoogst (79%, n=28). In 16% van de SE's zijn alle vragen letterlijk overgenomen uit CE's en in 38% van de SE's is een deel, maar niet alle vragen letterlijk overgenomen uit CE's. Dit is het meest het geval bij de science-vakken (26%, n=34). Een SE dat volledig bestaat uit opgaven die letterlijk uit een CE komen, hebben we alleen binnen het vmbo gevonden (34%, n=32). Geen enkel havo/vwo SE dat we onderzocht hebben (0%, n=36) bestond geheel uit CE vragen. Op havo/vwo vonden we wel in iets meer dan de helft van de SE's dat een deel van de vragen (maar niet alle) letterlijk zijn overgenomen uit CE's (53%, n=36). Gelet op schooltypen, valt het op dat bij de gamma-vakken in 100% van de SE's het SE voor een gedeelte bestaat uit vragen die letterlijk zijn overgenomen uit CE's, en een deel niet.

Verder blijkt dat het over alle onderzochte SE's gezien nauwelijks voorkomt dat een SE enkel bestaat uit vragen die sterk gebaseerd zijn op CE's (3%, n=68), licht geïnspireerd zijn door CE's (1%, n=68) of originele vragen die gaan over CE-inhoud (6%, n=68). Deze typen vragen ontbreken ook veel in de onderzochte SE's. Opgaven die sterk gebaseerd zijn op het CE ontbreekt in 74% van de SE's (dus in 26% van de SE's komen dergelijke vragen wel voor). Voor opgaven die licht geïnspireerd zijn door het CE ontbreekt in 66% van 68 beschreven SE's (en zit dus in 34%). De helft van de onderzochte SE's bevat originele vragen over CE-inhoud, wat betekent dat ook in 50% van de SE's dergelijke vragen ontbreken (n=68). Als een van deze drie typen vragen voorkomen, beslaan deze dus vaak slechts een deel van het SE.

Ten slotte, als we kijken naar het type vraag dat origineel is en niet-CE inhoud dekt, dan zien we dat in 63% van de onderzochte SE's (n=68) een dergelijk type vraag niet voorkomt. In ongeveer één op de vijf van de SE's zijn juist alle vragen originele vragen over niet-CE inhoud (21%, n=68), waarbij er weinig verschil is tussen vmbo (19%, n=32) en havo/vwo (22%, n=36). Kijken we naar het type vak zien we wel een verschil: bij de talen zien we dat 39% (n=28) van de SE's geheel bestaat uit originele vragen over niet-CE inhoud, terwijl dat bij geen van onderzochte SE's van de gamma-vakken het geval was (0%, n=6). Bij 16% van de SE's bestaat een deel (maar niet het geheel) van het SE uit originele vragen over niet-CE inhoud. .

3.2.4 Open vragen: tips en tops voor de docenten

Naast de dichotome beoordelingen in het beoordelingskader, zijn er bij elk pakket vijf open vragen door de inhoudsdeskundigen beantwoord. De antwoorden op twee van de open vragen zijn reeds verwerkt in de beschrijving bij het thema "Vorm, inhoud, dekking en niveau". De overige drie vragen zijn met name gebruikt voor het schrijven van de terugkoppelingen aan de docenten. In de huidige sectie worden de antwoorden op de drie open vragen samengevat.

De derde vraag "Wat is er goed gegaan in dit SE of deze verzameling SE?" bevat uiteenlopende antwoorden, waarbij veel onderdelen uit het beoordelingskader terugkomen. Er wordt regelmatig genoemd dat de SE's inhoudelijk de relevante domeinen afdekken en dat het niveau over het algemeen passend is; veel toetsen en pakketten tonen een duidelijke opbouw van eenvoudige naar complexere vragen en spreiden onderwerpen gelijkmatig. Verder worden heldere voorbladen en instructies, realistische en herkenbare (schrijf)taken en een goede variatie in vraagtypen (reproductie, toepassing, praktische vaardigheden) positief gewaardeerd. Ook wordt bij meerdere SE's de thematische samenhang en de toegankelijkheid van de bronnen genoemd, evenals het gebruik van concrete nakijkmodellen of rubrics die de beoordeling ondersteunen.

De vierde vraag "Wat zijn de belangrijkste verbeterpunten?" en de vijfde vraag "Welke tip of tips zou je de samensteller(s) van het SE willen meegeven?" bevatten inhoudelijk veel overlap en hebben vaak eenzelfde strekking. Beoordelaars noemen als belangrijkste verbeterpunten de behoefte aan vollediger en concretere beoordelingsmodellen (inclusief voorbeeldantwoorden, bolletjesmodel met kapstokpunten of marges), een logischere en verdedigbare puntentoekenning en duidelijkere en eenduidiger

geformuleerde vragen, waarbij afhankelijkheden en onbedoelde hints worden vermeden. Verder wordt herhaaldelijk aangegeven dat lay-out en consistentie (zelfde lettertype, nette nummering, duidelijke paginaverdeling en voortgangsaanduidingen) aandacht behoeven, dat er te veel overlap of afhankelijke vragen voorkomen en dat sommige toetsen te lang of te tijdsintensief zijn voor de beschikbare afnametijd. Ook wordt opgemerkt dat de aansluiting op PTA/eindtermen en op de beoogde taxonomische niveaus explicieter in toetsmatrijzen vastgelegd zou moeten worden.

De tips van de beoordelaars sluiten hierop aan en zijn praktisch van aard: werk met een toetsmatrijs waarin leerdoelen, beheersingsniveaus en weging zijn vastgelegd; laat nieuwe of zelfgemaakte opgaven vooraf door collega's of sectieleden proeflezen; vermeld bij open vragen duidelijk welke onderdelen per sub-antwoord punten opleveren en beschrijf hoe deels juiste of alternatieve antwoorden worden beoordeeld; gebruik bij het beoordelen van met name praktische opdrachten waar mogelijk meerdere beoordelaars of kalibratiemomenten; houd bij gebruik van oud examenmateriaal vast aan bijbehorende correctievoorschriften of pas taal en context aan. Daarnaast wordt aanbevolen om tijdsindicaties te toetsen (tijdnoodanalyse) en bij schrijf- en mondelinge taken te werken met concrete rubric-instructies en voorbeeldantwoorden om consistentie en transparantie van de beoordeling te vergroten.

3.2.5 Belangrijkste resultaten en conclusies op basis van de evaluatie van SE's

De evaluatie van de SE heeft een aantal inzichten opgeleverd. In deze paragraaf worden de belangrijkste hiervan puntsgewijs per thema samengevat. Tot slot worden de resultaten van de open opgaven weergegeven waarbij de beoordelaars aangaven wat ze sterke punten van de beoordeelde SE's vonden, wat verbeterpunten waren, aangevuld met enige tips.

Vorm, inhoud, dekking en niveau

- De vorm en het niveau van veel SE's sluiten goed aan bij wat gemeten moet worden.
- Een aandachtspunt is wel dat de verdeling over de voorgeschreven onderwerpen beter zou kunnen; deze is bij een relatief deel va de SE's onevenwichtig, met duidelijke voorbeelden van over- en onderrepresentatie en onduidelijkheden.
- Een ander aandachtspunt is dat bij veel SE's moeilijk een duidelijke schoolvisie te herkennen is. De SE's zijn vaak geschoeid op de leest van oude CE's en op de methode.

Transparantie en standaardisatie

- Over het algemeen is er goede praktische informatie over uitvoering. Dat betekent dat de toegestane hulpmiddelen en instructies bij schriftelijke toetsen over het algemeen helder zijn voor leerlingen en beoordelaars.
- In een aantal specifieke gevallen bleek dat de instructies nog wel beter konden. Bij praktische opdrachten worden toegestane hulpmiddelen geregeld niet aangegeven.
- De scoring is transparant in die zin dat de puntentelling per vraag en in veel gevallen ook de totaalscore duidelijk vermeld staan.
- Vermelding van voorbereiding, cesuur en uitzonderingsbeleid ontbreekt vaak, waardoor leerlingen vooraf onvoldoende weten wat verwacht wordt en wanneer een toets voldoende is.
- Concrete instructies over het proces van de afname ontbraken in veel gevallen, wat standaardisatie bemoeilijkt. Het is overigens aannemelijk dat deze bij de meeste scholen zijn gestandaardiseerd over alle toetsen, waardoor deze per SE ontbrak.

Vraagformulering en taalgebruik

- Veel aspecten van vraagformulering en taalgebruik krijgen een positieve beoordeling, zoals een duidelijke en functionele lay-out, grammaticaal correcte en heldere formuleringen en de afstemming van het taalgebruik op de doelgroep.

- Iets minder positieve beoordelingen doen zich voor bij enkele concrete gebruiks- en toegankelijkheidsaspecten, zoals of het duidelijk is dat er nog vragen volgen in de toets of dat de laatste vraag is gesteld. Ook zijn er verbeteringen mogelijk als het gaat om de toegankelijkheid voor leerlingen met dyslexie of een NT2-achtergrond.

Beoordelingsmodel en -procedure

- De beoordelingscriteria sluiten over het algemeen goed aan bij de eindtermen en de opgaven. De inhoudelijke aansluiting van de beoordelingsmodellen wordt overwegend positief beoordeeld.
- Wel ontbreken er geregeld elementen in het beoordelingsmodel, waardoor het onduidelijk is hoe alternatieve of deels juiste antwoorden beoordeeld dienen te worden, en beoordelen deskundigen de transparantie en logica van puntentoekenning minder vaak als toereikend.
- Bij een deel van de SE's werd een disbalans gevonden in weging tussen opgaven. Deze kan tot vertekening van het resultaat leiden.

Toetsmatrijs

- In de beoordeelde verzameling SE's wordt nauwelijks gebruik gemaakt van toetsmatrijzen. Dit is ook overeenkomstig met de resultaten bij het vragenlijstonderzoek. Een nadeel van het beperkt gebruik van toetsmatrijzen is dat de koppeling met het PTA amper expliciet wordt gemaakt.
- In de gevallen dat er wel toetsmatrijzen zijn aangeleverd, is het gebruik ervan in de meeste gevallen nog niet aantoonbaar sturend en behulpzaam.

Validiteit en gelijke kansen

- Veel SE's zijn adequaat, maar er komen structurele risico's terug op het gebied van brondiversiteit, framing en de toegankelijkheid voor alle leerlingen die deze SE's moeten maken. Dit laatste is relevant als dit ook leerlingen mét een ondersteuningsbehoefte betreft.
- Relatief eenvoudige maatregelen – zoals divers brongebruik, expliciete bias-controles, standaarden voor lay-out en leesbaarheid, en het vastleggen van valide alternatieven voor praktische opdrachten – kunnen de validiteit en gelijke kansen aanzienlijk verbeteren.

Scoren en normeren

- De scoring en weging is in het grootste deel van de SE's op orde is, maar het komt ook nog wel te vaak voor dat de uniformiteit en transparantie in scoretoekenning en weging ontbreken. Dit heeft mogelijk ook te maken heeft met het structureel ontbreken van toetsmatrijzen.
- Het ontbreken van een cesuur of een niet-onderbouwde en onduidelijke cesuur zetten de vergelijkbaarheid tussen toetsen en eerlijke kansen voor alle leerlingen onder druk.

Praktijkopdrachten

- Praktische opdrachten worden overwegend positief beoordeeld wat betreft realistische aansluiting bij beroeps- of leersituaties, relevante beoordelingscriteria en hanteerbaarheid van beoordelingsformulieren.
- Er bestaan wel terugkerende aandachtspunten: formuleringen van criteria zijn soms onvoldoende concreet voor betrouwbare beoordelingen en een deel van de beoordelingsformulieren mist heldere schaalindelingen, instructie en voorbeeldantwoorden. Een systematische aanpak van deze punten kan de betrouwbaarheid en vergelijkbaarheid van beoordeling bij praktijkopdrachten verder versterken.

Relatie tussen CE en SE

- De resultaten laten zien dat een aanzienlijk deel van de SE's (een deel van de) inhoud direct ontleent aan het CE. Van de onderzochte SE's heeft gemiddeld een SE 31% van de vragen letterlijk uit CE overgenomen, met hogere percentages op het vmbo (44%) dan op havo/vwo

(19%). Science-vakken gebruiken het vaakst CE-materiaal (70% van de vragen), terwijl talen-SE's juist vaker originele vragen over niet-CE-inhoud bevatten (45%).

- Bijna de helft van de SE's bevat helemaal geen letterlijk overgenomen CE-vragen. Bij de praktische opdrachten, maar bij 16% is juist het hele examen hiervan afkomstig.
- 'Vragen die sterk of licht op CE's zijn gebaseerd', komen veel minder voor. Originele vragen vormen gemiddeld 45% van de vragen, waarvan ruim een kwart over niet in het CE gedekte inhoud gaat. De resultaten laten zien dat CE's veelvuldig worden gebruikt voor SE's, maar dat de mate waarin SE's aansluiten op het CE verschilt per schooltype en vakgebied.

Algemene beoordeling

Wat ging goed

- SE's dekken meestal relevante domeinen en hebben doorgaans een passend niveau.
- Duidelijke opbouw van eenvoudig naar complex, evenwichtige spreiding van onderwerpen.
- Heldere voorbladen/instructies, realistische en herkenbare (schrijf)taken en goede variatie in vraagtypen (reproductie, toepassing, praktische vaardigheden).
- Thematische samenhang, toegankelijke bronnen en het gebruik van concrete nakijkmodellen of rubrics worden gewaardeerd.

Belangrijkste verbeterpunten

- Behoefte aan volledieger en concreter beoordelingsmateriaal: voorbeeldantwoorden, bolletjesmodel/kapstokpunten, marges.
- Puntentoekenning moet logisch en verdedigbaar zijn; open vragen en afhankelijkheden duidelijker en zonder onbedoelde hints formuleren.
- Lay out en consistentie (lettertype, nummering, paginaverdeling, voortgangsindicatoren) verdienen aandacht.
- Te veel overlap of afhankelijke vragen en soms te lange toetsen voor de beschikbare afnametijd.
- Aansluiting op PTA/eindtermen en beoogde taxonomische niveaus explicieter vastleggen in toetsmatrices.

Aanbevelingen/tips voor samenstellers

- Werk met een toetsmatrijs waarin leerdoelen, beheersingsniveaus en weging zijn vastgelegd.
- Laat nieuwe of zelfgemaakte opgaven proeflezen door collega's/sectieleden.
- Geef per open vraag duidelijk aan welke onderdelen punten opleveren en hoe deels juiste of alternatieve antwoorden worden beoordeeld; gebruik voorbeeldantwoorden.
- Bij praktische opdrachten: meerdere beoordelaars of kalibratiemomenten inzetten.
- Bij gebruik van oud examenmateriaal: houd correctievoorschriften aan of pas taal/context zorgvuldig aan.
- Controleer en toets tijdsindicaties (tijdnoodanalyse) en gebruik concrete rubrics/instructies bij schrijf- en mondelinge taken voor meer consistentie en transparantie.

3.3 Resultaten van de vakverenigingen

3.3.1 Positionering van de vakverenigingen

Vakverenigingen hebben binnen het Nederlandse voortgezet onderwijs een belangrijke positie als verbinder tussen de onderwijspraktijk, de vakwetenschap en het beleid. Een vakvereniging is doorgaans een vrijwillige, landelijke organisatie die docenten en andere vakinhoudelijk geïnteresseerden, zoals

curriculumexperts en vakdidactici, samenbrengt. Dit gebeurt meestal per specifiek schoolvak (bijvoorbeeld wiskunde, of geschiedenis) maar kan ook voor een cluster van vakken georganiseerd zijn (bijv. de beeldende vakken en exacte vakken).

De vakverenigingen opereren onafhankelijk, maar zijn intrinsiek verbonden met de kwaliteitscyclus van het onderwijs. Zij vormen de georganiseerde stem van de vakdocenten in het veld en dragen collectief zorg voor de ontwikkeling en de borging van de vakinhoudelijke expertise in Nederland.

De bestaansgrond van vakverenigingen berust op drie primaire, elkaar versterkende functies, die gezamenlijk de maatschappelijke waarde bepalen. Ten eerste de vakinhoudelijke professionalisering waarin vakverenigingen continu de ontwikkeling van docenten faciliteren. Dit doen zij door het organiseren van studiedagen, congressen en nascholingsactiviteiten, en door het publiceren van vakbladen en leermiddelen. Hierdoor wordt de kennisbasis in de klas actueel en wetenschappelijk gefundeerd gehouden. Als tweede faciliteren ze informele netwerken en kennisdeling door een platform te bieden voor uitwisseling onder docenten. Docenten kunnen hierdoor expertise en *good practices* delen, wat direct bijdraagt aan de uniformiteit en het niveau van het onderwijs in het betreffende vakgebied in verschillende scholen. Als laatste fungeren de vakverenigingen als gesprekspartner voor beleidsmakers bij het opzetten en uitrollen van nieuw beleid. Ze leveren vanuit de praktijk gevraagde en ongevraagde adviezen over onderwerpen als de curriculumherziening, de inrichting van het CE en de uitwerking van onderwijsdoelen. Zij waarborgen hiermee dat beleid voldoende is geaard in de dagelijkse onderwijspraktijk en vakinhoudelijke eisen.

Stichting Cito heeft onderzocht waar mogelijkheden liggen voor een meer centrale coördinatie van de kwaliteitsimpuls voor het SE. Het is hierbij wel van belang dat het eigenaarschap bij vakdocenten blijft liggen. In dit kader zijn de vakverenigingen, als vertegenwoordigers van hun respectievelijke groepen vakdocenten, een potentiële belangrijke factor. Het is hierbij wel van belang op te merken dat de vakverenigingen geen volledige dekkingsgraad binnen hun vakgebied bezitten. Desondanks zijn zij, in tegenstelling tot informeel bestaande netwerken van docenten, de meest centrale aanspreekpunten. Dit bereik reikt vaak, bijvoorbeeld via hun nieuwsbrief, verder dan hun officiële ledenadministratie. Om deze reden hebben we een rondvraag onder de vakverenigingen georganiseerd om informatie te verzamelen over hun huidige rol en toekomstige bijdrage aan de verbetering van de kwaliteit van de SE's.

3.3.2 Activiteiten van de vakverenigingen

Dit rapport richt zich enkel op activiteiten van de vakverenigingen waarin de kwaliteit van de SE's in de toekomst aandacht kan krijgen. Besprekingen van het CE vallen hier zodoende buiten. Vakverenigingen organiseren (veelal gratis) studiedagen en jaarlijks of tweejaarlijks congresdagen, waarbij uiteenlopende onderwerpen met betrekking tot vakinhoud en vakdidactiek worden besproken. Incidenteel zijn er sessies over het SE, maar dit is geen terugkerend beleidsmatig onderwerp.

Veel vakverenigingen maken gebruik van één of meerdere vormen van online communicatiestromen. Dit kan bestaan uit Facebook- of LinkedIn-groepen, maar ook uit een eigen open of besloten web-omgeving, zoals een website of Google Drive. De mate waarin de uitwisseling van lesmaterialen wordt gefaciliteerd, verschilt per vakvereniging. In sommige gevallen gebeurt dit op basis van vraag en antwoord, terwijl in andere gevallen specifieke functionaliteiten zijn ingericht. De intentie is echter nooit expliciet gericht op het delen van materiaal van SE's, hoewel die mogelijkheid uiteraard bestaat.

Verder maken vakverenigingen frequent gebruik van nieuwsbrieven om hun leden, en in sommige gevallen ook niet-aangesloten leden, te informeren over de stand van zaken in hun vakgebied en aankomende evenementen. Sommige vakverenigingen geven aanvullend een vakblad uit. Ook in deze vormen van communicatie is het SE geen terugkerend onderwerp.

Tot slot beschouwen de verenigingen zichzelf vaak als vraagbaak of als verbinder van informele netwerken waaraan zij zelf niet actief deelnemen. Op deze manier brengen zij docenten met elkaar in

contact die daar behoefte aan hebben. Dit kan zijn omdat zij een eenpersoonssectie of een kleine sectie vertegenwoordigen en binnen hun eigen school onvoldoende overlegmogelijkheden hebben, of omdat er op regionaal niveau behoefte aan samenwerking bestaat.

3.3.3 De SE's als thema

Van de twaalf ondervraagde vakverenigingen geeft er slechts één aan expliciet een rol op zich te nemen in het bijdragen aan de kwaliteit van de SE's. Het Koninklijk Nederlands Aardrijkskundig Genootschap (KNAG) heeft als vakvereniging voor het vak aardrijkskunde in het verleden een set vragen ontwikkeld, die via de verkoop van USB-sticks onder hun leden is verspreid. Deze set is zeer goed ontvangen in het werkveld en heeft bovendien een (zeer bescheiden) inkomstenbron voor de vereniging gegenereerd. Na de vakvernieuwing is het KNAG voornemens om opnieuw een set SE-items uit te brengen.

Negen van de overige elf vakverenigingen gaven aan op dit moment geen aandacht aan het SE te besteden, waarbij sommigen zich nog een overleg of workshop van een medewerker van Stichting Cito, of een kennis-sessie tijdens een congres, konden herinneren. Drie vakverenigingen gaven aan actiever bezig te zijn met activiteiten rondom het delen van lesmateriaal, het koppelen van vragen vanuit docenten aan de kwaliteit en diversiteit van het SE, en zeggen zich bewust te zijn van hun eigen rol in het bewaken van het kwaliteitsbesef rondom de SE's. Daarnaast waren drie van de twaalf gesproken bestuurders op basis van het huidige onderzoek van plan om het thema SE's binnen hun vereniging te agenderen.

3.3.4 De kwaliteit van de SE's

Voor zover de vakverenigingen inzicht hebben in de dagelijkse praktijk van de SE's, is hun oordeel (op één uitzondering na) eerder negatief dan positief. De uitzondering betreft de vakvereniging voor geschiedenis, die signaleert dat docenten het bijzonder leuk vinden om zich bezig te houden met hun SE's. De overige vakverenigingen zijn minder optimistisch en constateren vooral dat docenten, waar mogelijk, gebruikmaken van oude CE's vanuit Stichting Cito, gecombineerd met invloeden vanuit methodetoetsen en af en toe enkele eigen bijdragen. De kwaliteit van de methodetoetsen wordt echter door de respondenten als onvoldoende beschouwd. Bovendien wordt door vrijwel alle vakverenigingen opgemerkt dat docenten te kampen hebben met hoge werkdruk en dat het voorbereiden van een SE veel tijd in beslag neemt. Daarbij geven zij aan dat de startkwalificaties van de individuele docent sterk variëren op het gebied van toetsconstructie. Het algemene beeld, onder andere ingegeven door enkele bestuurders die ook actief zijn bij de lerarenopleidingen, is dat er binnen die opleidingen nauwelijks aandacht is voor examinering. Dit maakt de toetsdeskundige basis volgens de vakverenigingen onder de beroepsgroep uiterst gering. Diverse initiatieven om het delen van SE's te bevorderen zijn gestrand door een gebrek aan animo en terughoudendheid van docenten om hun eigen werk te delen.

3.3.5 Vakverenigingen en hun blik op SE's

Vanzelfsprekend ligt op dit moment de focus van de vakverenigingen sterk op de aanstaande vakvernieuwing. Collectief zien zij een aanzienlijke reeks taken op zich afkomen, waarbij zij zich, op basis van de interviews, bewust leken te worden van de kansen die deze nieuwe taken bieden. De vakvereniging van muziek kijkt bijvoorbeeld uit naar een uniform curriculum, waarmee de huidige tweespalt tussen een 'nieuw' en 'oud' systeem verleden tijd wordt.

Docenten stellen rondom de vakvernieuwing vaker de vraag aan hun vakvereniging hoe het SE in de toekomst moet worden vormgegeven. Het terugslageffect van een vernieuwd curriculum en CE zal ook

impact hebben op het SE, waardoor de tijd rijp lijkt voor een kwaliteitsimpuls. De vakverenigingen signaleren hier de aanleiding van dit onderzoek, dat er veel variatie in de kwaliteit van de SE's zit, en het verschil in opvattingen tussen scholen, maar zelfs ook binnen scholen over toetsen. De visie en koers van een schoolbestuur kunnen daarbij doorslaggevend zijn. Eveneens leeft de opvatting dat docenten, ondanks hun behoefte aan meer sturing, ook sterk gehecht zijn aan hun autonomie binnen het SE. Enerzijds is er behoefte aan kwaliteitscriteria en antwoorden op vragen zoals 'wat toets ik en hoe?' of 'op welke eindtermen richt ik mij als ik ze niet allemaal kan of mag toetsen?', anderzijds willen zij hun eigen vrijheid behouden. Een aantal vakverenigingen ziet echter mogelijkheden voor brede en landelijke samenwerking op gebieden waar vakoverstijgende winst te behalen valt.

3.3.6 Ondersteuningsbehoefte van vakverenigingen

Hoewel dit niet het primaire onderwerp van de interviews was, is het verschil in organisatiestructuur van de vakverenigingen opvallend. Van de twaalf gesprekken zijn er zes daadwerkelijk verenigingen, terwijl de overige zes een vereniging zijn die zich ook als vakbond heeft georganiseerd (zie ook tabel 3.4.1 in hoofdstuk 3 Methodologie). Het belangrijkste inhoudelijke onderscheid ligt in de actieve rol die een vakbond speelt in de arbeidsvoorwaarden van hun leden. Organisatorisch leidt dit tot hogere lidmaatschapskosten, maar ook tot hogere inkomsten voor de vereniging, waardoor de structuur over het algemeen stabieler is. Voor sommige verenigingen was het een bewuste keuze om de vermenging van vakinhoud en arbeidsvoorwaarden niet te combineren.

In de context van deze korte interviews lijken de meeste vakverenigingen bereid te zijn een actieve rol te spelen bij het leveren van een kwaliteitsimpuls aan de SE's. Deze ambitie stuit echter voorlopig op praktische uitdagingen in termen van tijd en financiering. In dit verband wordt door de vakverenigingen gekeken naar het ministerie van OCW voor een financiële impuls om de slagkracht van de verenigingen te vergroten. Opvallend genoeg lijkt het onderscheid in organisatiestructuur tussen verenigingen en vakbonden geen direct verschil te maken in de mogelijkheden tot uitbreiding van activiteiten. Sommige besturen maken gretig gebruik van gepensioneerde collega's, maar de meeste besturen opereren op basis van vrijwilligheid, vaak in de avonduren naast het docentschap. De vakverenigingen zijn daarom toe aan een professionaliseringsslag als zij een actieve bijdrage aan de SE's willen leveren. Hoewel de vakvernieuwing onzekerheid met zich meebrengt over hoe het onderwijs zich de komende jaren zal ontwikkelen en hoe dit zal doorwerken in het SE, biedt het echter voor de vakverenigingen in samenwerking met SLO vooral ook een kans om positie in te nemen. Ten slotte geven enkele verenigingen aan dat zij binnen het bestuur over onvoldoende toetstechnische kennis beschikken om dit onderwerp geheel zelfstandig te kunnen behandelen. In dit kader kan mogelijk enige vorm van ondersteuning vanuit Stichting Cito gewenst zijn.

4. Discussie en conclusie

In dit rapport hebben we de vakinhoudelijke en toetstechnische kwaliteit onderzocht van SE's in het voortgezet onderwijs. Deze onderzoeksvraag is beantwoord via vragenlijsten en toetsing van SE's, met een speciaal voor dit onderzoek ontwikkeld beoordelingskader, gebaseerd op bestaande instrumenten (zoals toegelicht in hoofdstuk 2) en op de benadering van bewijsgericht valideren[56]. Dit beoordelingskader richtte zich op mogelijke validiteitsbedreigers en betrouwbaarheid van de SE's, maar ook op aanvullende kwaliteitseisen, zoals de relatie tussen SE en CE en de wijze waarop SE's zijn samengesteld. Het vragenlijstonderzoek is uitgevoerd onder zowel vakdocenten als leden van examencommissies. Naar aanleiding van de mogelijke rol die vakverenigingen zouden kunnen gaan spelen volgens docenten zijn ook interviews gehouden met vakverenigingen om te beoordelen welke potentiële, toekomstige rol vakverenigingen voor zich zien in de ondersteuning van de kwaliteit van SE's.

Wat volgt is een discussie van wat er goed gaat, wat er beter kan, een methodologische reflectie en tot slot een set aan aanbevelingen.

4.1 Wat gaat er goed?

De beoordelingen van 78 SE's door inhoudsdeskundigen middels een beoordelingskader leveren een gematigd positief beeld op. De basisopzet van de SE's is overwegend goed. De docenten in de vragenlijsten zijn positief over de deskundigheid van hun vaksecties, en over dat ze veel vrijheid krijgen voor hun eigen invulling van het PTA en het SE. Docenten zorgen ervoor dat de toetsen het eindniveau reflecteren, waarbij zij feedback vragen van collega's. Dit blijkt ook uit de resultaten van het beoordelingskaderonderzoek, waar in de meerderheid van de gevallen de vorm en opzet van een SE goed passen bij het leerjaar, niveau en te meten doel, en de moeilijkheid als passend wordt beoordeeld. Ook de aangeleverde praktische opdrachten worden door onze toetsdeskundigen overwegend als realistisch en relevant beoordeeld, aansluitend bij beroeps- of leersituaties, en bevatten meestal relevante beoordelingscriteria en een goede hanteerbaarheid van beoordelingsformulieren - al blijkt overigens wel uit het vragenlijstonderzoek dat bij praktische opdrachten het eindniveau minder vaak als referentiepunt wordt gebruikt dan bij schriftelijke toetsen, wat een risicofactor kan zijn voor de validiteit van deze praktijktoetsen. Binnen vrijwel alle scholen is er sprake van uniforme toetsing voor leerlingen van hetzelfde examenjaar: zij maken dezelfde toetsen en opdrachten. Inhaaltoetsen bevatten meestal dezelfde inhoud, al dan niet met andere opgaven.

Ook de praktische en taalkundige kwaliteit scoren vaak goed. Zo worden de lay-out en leesbaarheid vaak positief beoordeeld, is het taalgebruik vrijwel altijd afgestemd op de doelgroep en worden onnodig lange zinnen consequent vermeden. Instructies voor de leerlingen en toegestane hulpmiddelen zijn vaak aanwezig en duidelijk en de puntentelling per item is veelal inzichtelijk gemaakt. Verder vertonen de vragen weinig culturele of fysieke bias en komen in veel SE's meerdere vaardigheden aan bod. Docenten geven aan dat ze inschatten dat hun leerlingen doorgaans geen tijdsdruk ervaren bij de SE's en meestal goed geïnformeerd zijn over de beoordelingscriteria, vooral bij schriftelijke toetsen. Beoordelingen worden veelal gebaseerd op een vastgesteld cesuurpercentage of een vooraf bepaald aantal scorepunten, afhankelijk van de moeilijkheidsgraad van de toets.

[56] Kane, M. T. (2013). The argument-based approach to validation. *School Psychology Review*, *42*(4), 448 - 457. https://doi.org/10.1080/02796015.2013.12087465

4.2 Wat kan er beter?

Al met al zijn de SE's (toetsen en opdrachten) op veel punten best goed op orde, zoals ook in de vorige paragraaf is aangegeven. Er is echter ook ruimte voor verbetering. Een aantal toetstechnische zaken waar op gelet kan worden, is al genoemd. Een belangrijk verbeterpunt voor de SE's als geheel blijft echter dat ze hun eigen karakter beter kunnen benutten: geef meer ruimte aan de eigen visie van de school en maak minder gebruik van opgaven uit het CE en uit methodes. Dit sluit aan bij eerdere onderzoeken naar het SE. In deze paragraaf komen ook enkele andere suggesties aan de orde.

4.2.1 Het gebruik van CE's en lesmethodemateriaal bij de SE's

Uit het onderzoek onder docenten en examencommissieleden blijkt dat om de kwaliteit en vergelijking met het CE te waarborgen vaak gebruik wordt gemaakt van oude CE's en lesmethodemateriaal als basis voor de SE-opdrachten. Dit blijkt ook uit de beoordeling van aangeleverde SE's. Veel toetsen lijken sterk te leunen op methode-materiaal of op (oude) CE-opgaven, waardoor een expliciete 'handtekening' van de school vaak ontbreekt, hoewel er ook voorbeelden zijn van bewuste keuzes (bijv. nadruk op taalverzorging of bepaalde thema's) die wél een eigen accent tonen. Geen van de betrokken scholen hebben beoordelingsprocedures en toetstechnische visies aangeleverd. Het is ondertussen heel goed mogelijk dat deze procedures en visies wel op scholen leven, en dat docenten er naar handelen. Het zou een indicatie kunnen zijn dat dat deze procedures en toetstechnische visie vaak niet duidelijke documenten betreft, waarbij het voor docenten evident is dat deze als zodanig gedeeld kunnen worden.

Daarnaast is aan de beoordelaars gevraagd in hoeverre de SE's de indruk geven dat er een eigen visie van de school achter zit. Deze blijkt in de meerderheid van de beoordeelde SE's niet expliciet zichtbaar; slechts een minderheid wekt de indruk dat er een duidelijke eigen visie of onderwijskundig accent wordt ingezet.

Deze praktijk van het gebruik van oudere CE's en bestaand lesmateriaal helpt om te zorgen dat een zekere mate van kwaliteit en vergelijkbaarheid wordt gewaarborgd en om leerlingen doelgericht voor te bereiden op het CE. Dit laatste kan curriculumvernauwing veroorzaken, en daarnaast raken praktische vaardigheden mogelijk onderbelicht doordat de toetsen vooral schriftelijk zijn ingericht. Daarmee lopen SE's mogelijk het risico minder representatief te zijn voor het volledige curriculum en de gewenste leerdoelen binnen de schoolcontext. Daarnaast heeft het als nadeel dat het SE daardoor minder aansluit bij een bredere schoolvisie, wat een belangrijk kenmerk van SE's is.

4.2.2 Rol examencommissie

De onderzoeksresultaten benadrukken het belang van een expliciete rol voor de examencommissie bij de kwaliteitsborging van SE's. Hoewel examencommissies op veel scholen formeel bestaan, blijkt dat hun inhoudelijke betrokkenheid bij de toetsconstructie en de controle op gebruik van toetsmatrijzen beperkt is. Toetsmatrijzen, die het verband tussen leerdoelen en toetsvragen systematisch in kaart brengen, worden niet consequent ingezet, zo blijkt uit zowel het vragenlijstonderzoek onder docenten en examencommissieleden als uit het beoordelingskaderonderzoek. Uit dit laatste onderzoek bleek dat toetsmatrijzen zelden werden aangeleverd of niet aantoonbaar effectief werden gebruikt, waardoor de expliciete koppeling tussen opgaven, leerdoelen en weging vaak ontbreekt. Omdat een instrument als een toetsmatrijs duidelijk bijdraagt aan de validiteit en transparantie van toetsen, is het ontbreken ervan een aandachtspunt. Het versterken van de rol van examencommissies, onder meer door hen een meer

actieve positie te geven bij toetsontwikkeling en kwaliteitscontrole, kan daarom de kwaliteit van SE's verbeteren.

Opgemerkt moet worden dat examencommissies geenszins door wet- en regelgeving verplicht worden om naar de opgaven van de SE's te kijken. Examencommissies zouden in de invulling van hun rol er echter wel iets mee kunnen doen, ook als dit geen verplichting is. Als de examencommissie de inhoudelijke borging niet zelf op zich neemt, zou het goed zijn als ze een proces inrichten dat deze borging wel garandeert.

4.2.3 Beoordeling en normering

Wat betreft de beoordeling en normering van SE's blijken er wisselende praktijken te zijn. Veel docenten hanteren een vaste cesuur (bijvoorbeeld 70% correct voor een voldoende), maar er zijn weinig vaste richtlijnen op schoolniveau over het omzetten van scores in cijfers. De logica van puntentoekenning en weging is daarnaast niet altijd consequent of verdedigbaar, en de beoordelingsmodellen missen regelmatig details (voorbeeldantwoorden, richtlijnen voor deels correcte/alternatieve antwoorden), wat de vergelijkbaarheid en betrouwbaarheid van beoordelingen onder druk zet. Praktische opdrachten scoren over het algemeen goed op realisme, maar sommige vragen ook om scherpere rubrics en kalibratie om de betrouwbaarheid van de beoordelingen te verhogen. Ook wordt uit de data duidelijk dat tweede correctoren vrijwel niet worden ingezet bij schriftelijke toetsopgaven. Hierdoor bestaat het risico dat de objectiviteit en betrouwbaarheid van de beoordelingen niet optimaal zijn gewaarborgd.

4.2.4 Transparantie

Daarnaast zijn er voor sommige SE's enkele terugkerende aandachtspunten omtrent toegankelijkheid, bijvoorbeeld voor leesbaarheid en een duidelijke en functionele lay-out. Verder blijkt de transparantie van sommige SE's op enkele cruciale punten onvoldoende. Vanuit de aangeleverde materialen door scholen was niet op te maken dat de benodigde voorbereidingen voor leerlingen, de cesuur en afname-instructies expliciet zijn vastgelegd of gecommuniceerd, wat de voorspelbaarheid en gelijkheid van behandeling beperkt. Het is daarmee niet met zekerheid te zeggen dat de transparantie daadwerkelijk niet op orde was, of dat dit te verklaren is doordat scholen niet alle relevante informatie aangeleverd hebben, of alles documenteren wat zij doen. De resultaten uit de beoordelingen van de aangeleverde SE's geven namelijk een ander beeld dan het vragenlijstonderzoek, waarin docenten juist aangeven dat leerlingen volgens hen goed op de hoogte zijn van beoordelingscriteria.

4.2.5 Rol vakverenigingen

Uit het vragenlijstonderzoek onder docenten en examencommissieleden blijkt dat hoewel de tevredenheid over de deskundigheid van vaksecties hoog is, de samenwerking tussen docenten bij het maken van SE's meestal beperkt blijft tot kleine samenwerkingen binnen vaksecties, voornamelijk vanwege gebrek aan tijd. Breder sectie-overleg, interdisciplinaire afstemming of samenwerking met examencommissies is zeldzaam. Deze beperkte afstemming kan leiden tot een fragmentarische toetspraktijk en minder uniforme kwaliteit binnen scholen. Uit de interviews met vakverenigingen komt daarom ook naar voren dat de vakverenigingen een belangrijke rol kunnen spelen in het faciliteren van professionele netwerken. Deze netwerken, ook de informele, kunnen bijdragen aan de professionalisering van docenten en het verbeteren van de kwaliteit van SE's. Er ligt dus ruimte voor verbetering in het mobiliseren van het netwerk en de agenda van de vakverenigingen met betrekking tot de kwaliteit van de

SE's. Het is raadzaam om verder te onderzoeken in hoeverre vakverenigingen met name vakoverstijgende thema's kunnen oppakken om zo overhead en logistiek te beperken.

4.3 Methodologische reflectie

Een belangrijk uitgangspunt van dit onderzoek was dat deelname volledig vrijwillig moest zijn. Dit waarborgde de ethiek, maar bracht methodologische beperkingen met zich mee die de interpretatie van de resultaten beïnvloeden.

Gegevens werden verzameld via opt-in, waardoor respons- en selectiebias waarschijnlijk zijn. De steekproef is hierdoor niet aselect. De steekproef is tevens niet representatief voor alle scholen, vakken en leerwegen; sommige provincies en schooltypen (bijv. vwo) zijn oververtegenwoordigd. Ondanks wervingsinspanningen was deelname afhankelijk van intrinsieke motivatie, wat de controle over samenstelling en omvang beperkte. De respons was voldoende voor beschrijvende analyses, maar te beperkt voor gedetailleerde analyses naar achtergrondvariabelen, waardoor generaliseerbaarheid onzeker blijft. Bij de evaluatie van de SE's speelt daarbij ook dat er sprake is van een clustering van ingediende SE's. De 78 ingezonden SE's zijn afkomstig uit 18 pakketten van 15 scholen. Hierdoor kunnen individuele scholen, vakken of leerwegen disproportioneel wegen; kleine subgroepen verhinderen betrouwbare vergelijkingen tussen niveaus of vakken. Het is daarmee niet gelukt een representatief beeld te verkrijgen van de kwaliteit van de SE's. Zowel de zelfselectie als het beperkt aantal scholen spelen hier een rol. De opgehaalde informatie kan echter gelezen worden als een indicatie van hoe het gesteld is met de kwaliteit: wat zien we bij scholen die het aandurven hun SE's vrijwillig te laten beoordelen? Wat hierbij ook opgemerkt kan worden is dat de uitkomsten niet substantieel afwijken van eerdere vergelijkbare onderzoeken met meer gegevens[57]: de SE's laten voldoende kwaliteit zien, maar er zijn duidelijke kwaliteitsverschillen op bepaalde criteria tussen scholen en tussen vakken.

Om deelname laagdrempelig te houden, was de vragenlijst compact. Dat maakte kernobservaties mogelijk, maar beperkte het inzicht in de achterliggende rationale van toetskeuzes en werkwijzen. Een kwalitatieve vervolgstudie (diepte-interviews of focusgroepen) is nodig om onduidelijke of opvallende bevindingen te verklaren, zoals dat 14% van examencommissieleden het PTA niet kent en 11,8% de visie op toetsing niet kent. Een ander punt bij de vragenlijst is de kans op sociaal wenselijke antwoorden. Hoewel anonimiteit beloofd werd, kan sociale wenselijkheid de zelfrapportage kleuren, vooral over deskundigheid. Toch geeft de rangorde van comfort bij taken waardevolle informatie; de beoordeling van ingediende SE's biedt aanvullende reflectie.

Bij de ingediende SE-pakketten zagen we dat relevante documenten of context vaak ontbraken. Geen enkele school leverde een expliciete toetstechnische visie aan en slechts 39% van de pakketten bevatte een PTA[58]. Dit bemoeilijkt het geven van gerichte verbeteradviezen wanneer respondenten aangeven iets niet te weten. Een ander punt wat dat bij de beoordeling van SE's grotendeels dichotome (aanwezig/niet) criteria gebruikt zijn. Dat was efficiënt, en maakte de taak haalbaar, gestructureerd en vergelijkbaar over de beoordelaars, maar verminderde nuance. Beoordelaars konden opmerkingen plaatsen, maar deden dat weinig; open vragen werden wel door iedereen ingevuld. Beoordelaars kregen instructies en konden bij onduidelijkheden contact opnemen met onderzoekers en collega's. Interbeoordelaarsbetrouwbaarheid is niet formeel vastgesteld.

[57] Cito (2008). Het SE in het voortgezet onderwijs. Verslag van een onderzoek naar de kwaliteit van het SE bij de vakken Engels, Nederlands, biologie en wiskunde. Arnhem: Cito. Samengevat in: Kuhlemeier, H. , & Van Til, A. (2016). *Toetsen op School: Voortgezet onderwijs*. Arnhem: Cito.

[58] Deze PTA's zijn niet meer los van inzending door de scholen opgezocht.

Het deelonderzoek onder vakverenigingen kan vooral exploratief en oriënterend genoemd worden. Hoewel veel verenigingen zijn gesproken, was de vraagstelling beperkt; de resultaten zijn indicatief en vormen een basis voor vervolgonderzoek naar hun rol rond SE's.

Concluderend kan gesteld worden dat het vrijwillige karakter ethische zorgvuldigheid bevorderde, maar dat ten koste ging van representativiteit, diepgang en statistische robuustheid. De resultaten zijn vooral beschrijvend en verkennend. Een aanbeveling voor toekomstig onderzoek is een gestructureerde steekproef, aanvullende kwalitatieve dataverzameling, uitgebreidere documentatie en fijnmazigere beoordelingsschalen om meer generaliseerbare en verklarende inzichten te verkrijgen. Het is hierbij wel van belang dat er een meer verplichtend karakter aan de deelname zit, of dat er andere middelen beschikbaar zijn om respondenten te verleiden hieraan deel te nemen. Voor het onderzoek dat hier is uitgevoerd heeft Stichting Cito zeer veel kanalen aangewend en veel stappen genomen, maar zonder verplichting of beloning lijkt dit de maximale deelnamebereidheid.

4.4 Conclusies en aanbevelingen

Dit onderzoek biedt inzicht in de vakinhoudelijke en toetstechnische kwaliteit van een selectie van SE's. Op basis van onze steekproef blijkt dat er zowel stevige fundamenten als duidelijke aandachtspunten zijn. Er gaat veel goed: de basisopzet, vormgeving, taalkundige en praktische kwaliteit en uniformiteit binnen leerjaren scoren vaak goed, docenten ervaren vaksecties als deskundig, en toetsen stemmen doorgaans met het beoogde eindniveau overeen. Ook is positief dat docenten meestal de SE's samen met een collega maken of een andere collega naar hun SE laten kijken.

Echter, de huidige praktijk is veelal gericht op voorbereiding op het CE en wordt gekenmerkt door beperkte vormen van samenwerking tussen vakdocenten (voornamelijk in duo's), onderbenutting van toetsmatrijzen of andere hulpmiddelen en wisselende normeringspraktijken. Verbetering van de rol van examencommissies, versterking van vaksecties, bredere samenwerking tussen vakdocenten binnen en buiten de vaksecties, en benutting van de positie en netwerken van vakverenigingen bieden aanknopingspunten voor een versterking van de kwaliteit van SE's. De gesignaleerde knelpunten en mogelijkheden vormen een duidelijke aanzet tot vervolgonderzoek en ondersteuningstrajecten, die nodig zijn om het SE de bedoelde rol te laten spelen in het Nederlandse voortgezet onderwijs.

Op basis van het uitgangspunt van het onderzoek en de verzamelde bevindingen formuleren wij tot slot een aantal aanbevelingen. Deze presenteren we puntsgewijs en per doelgroep.

4.4.1 Aanbevelingen voor vakdocenten

- *Normeringspraktijk verstevigen*: Het is aan te raden de cesuur (grens voldoende/onvoldoende) expliciet vast te stellen en ook kort te motiveren. Onduidelijke of ontbrekende cesuren verminderen de vergelijkbaarheid en transparantie. Daarbij zou er ook voor elke opdracht/SE een duidelijk beoordelingsmodel aanwezig moeten zijn met voorbeeldantwoorden of kapstokpunten. Dit vermindert subjectiviteit en maakt het nakijken consistenter.
- *Standaardiseren*: Het helpt als elk SE een standaard voorblad heeft met daarop de tijdsduur, toegestane hulpmiddelen, hoe de totaalscore wordt berekend en wat de cesuur is. Het helpt ook de rest van de lay-out, nummering en voortgangsaanduiding schoolbreed te standaardiseren en te controleren op helderheid en toegankelijkheid. Er kunnen bijvoorbeeld vaste opmaakregels worden afgesproken en indien mogelijk kan er één collega op leesbaarheid checken. Het helpt ook standaard te werken met een(zelfde) toetsmatrijs of ander hulpmiddel om vooraf helder in kaart te brengen wat er getoetst wordt, hoe dit wordt gedaan en waarom.

- *Kalibratie en collegiale toetscontrole*: Vóór de definitieve inzet van nieuwe of aangepaste SE's is het aan te raden dat vakdocenten een collegiale toetscontrole (peer review) uitvoeren. Zo kunnen eventuele fouten, onduidelijkheden en mogelijke bias nog uit het SE gehaald worden. Binnen een vaksectie kunnen bijvoorbeeld toetsen uitgewisseld worden en korte feedback gegeven worden op helderheid, dekking en weging. De beoordeling van open vragen en praktische opdrachten wordt betrouwbaarder wanneer kalibratiemomenten worden ingezet tussen vakdocenten van dezelfde sectie en waar mogelijk wordt gewerkt met een tweede, onafhankelijke beoordelaar.
- *Variatie in toetsingsvormen*: Er bleek binnen dit onderzoek veel nadruk te liggen op schriftelijke toetsen. Het is wenselijk kritisch na te denken of andere toetsvormen ook nuttig kunnen zijn om aanvullende vaardigheden te toetsen, maar die niet of nauwelijks aan bod komen in het CE, maar wel onderdeel van het curriculum en van belang zijn voor leerlingen.
- *Beweging richting minder CE-opgaven*: we constateren dat er door vakdocenten binnen de SE's veel CE-vragen worden gebruikt. De voornaamste reden hiervoor is dat de docent zo gemakkelijker en/of sneller toetsen van hoge kwaliteit kan maken (67%, zie bijlage 4, tabel 4.5). Iets minder dan een derde van de docenten geeft aan dat het beperken van verschillen met de cijfers van het CE een reden is voor het gebruik van oude CE-opgaven. Wat betreft de eerste reden zou dit gedeeltelijk kunnen worden opgelost wanneer er voor de vakdocenten meer bronnen zouden zijn om opgaven uit te putten, bijvoorbeeld gefaciliteerd door de school of door de vakvereniging (zie aanbevelingen schoolleiding en vakvereniging). Wat betreft de tweede reden hebben we nu niet direct een concrete aanbevelingen. We vinden het echter wel belangrijk om hier te benoemen.

4.4.2 Aanbevelingen voor examencommissies

- *Reserveer tijd en organiseer gerichte bijscholing*: In haar controlerende rol ziet de examencommissie erop toe dat de jaarplanning aantoonbaar ruimte bevat voor toetsconstructie, normering en normeerbijeenkomsten. Voor een goede uitvoering van deze taak moet de examencommissie kunnen vaststellen welke hiaten er zijn tussen de benodigde en de huidige kennis en vaardigheden, zodat zij gericht kan worden bijgeschoold.
- *Tracht meer grip te krijgen op toetstechnische kwaliteit:* Hoewel examencommissies wettelijk niet verplicht zijn alle SE-opgaven te bekijken, kunnen zij via een controlesysteem mogelijk wel grip krijgen op de kwaliteit van toetsinhoud. De commissie stelt een toetsreview-protocol vast waarin staat welke toetsen/items worden gecontroleerd (bijv. steekproef van 10–20% of alle belangrijke SE-kernen), welke criteria gelden (eenduidige beoordelingscriteria, aansluiting bij het PTA, toetsmatrijs, niveau en helderheid van opgaven, betrouwbaarheid van beoordelingsvoorschriften) en binnen welke termijn de review plaatsvindt.
- *Versterk zicht- en regiefunctie van de examencommissie*: Zet in op een proactieve stimulerende functie van examencommissies. Zij kunnen helpen meer eenheid bij de SE's op een school aan te brengen door een korte, helder geformuleerde visie op toetsing en concrete basisafspraken (cesuren, omzettingstabellen, inzet tweede beoordelaar). Maak deze intern bekend en maak de taak- en verantwoordelijkheidsverdeling binnen de examenorganisatie expliciet. Toetskwaliteit is een vast agendapunt; besluiten over toetsing en genomen maatregelen worden kort gemotiveerd en vastgelegd in toetsdossiers zodat de commissie de brede kwaliteitsbewaking kan aantonen (inclusief naleving wettelijke eisen). Periodieke collegiale of externe peerreviewen worden georganiseerd en de bevindingen worden vertaald naar verbeteracties en gerapporteerd in het jaarverslag.

4.4.3 Aanbevelingen voor schoolleiding

- *Schoolbrede visie op toetsing en schoolexaminering*: Wanneer dit nog niet het geval is, helpt het wanneer de schoolleiding een schoolbrede visie op toetsing en schoolexaminering formuleert en communiceert. Een dergelijk plan zorgt voor samenhang in keuzes voor toetsvormen, frequentie en weging.
- *Structureel tijd voor toetsontwikkeling*: Tijdgebrek is de oorzaak van veel onvolledige toetsdocumentatie en inconsistenties. Daarom raden wij schoolleidingen aan structureel tijd in het rooster of sectieoverleg te reserveren voor toetsontwikkeling, collegiale review en kalibratie.
- *Reserveer tijd en faciliteer gerichte bijscholing voor docenten en examencommissieleden:* Zorg in de jaarplanning voor structureel vrijgemaakte tijd komt om bijscholing voor docenten en examencommissieleden te bevorderen.
- *Normeringspraktijk verstevigen*: De schoolleiding kan via de examencommissie de aanwezigheid van kwalitatief goede beoordelingsprocedures bewaken. Het helpt verder wanneer de schoolleiding korte, praktische scholing stimuleert rondom rubrics, toetsmatrijzen, cesuurbepaling en toegankelijk toetsen. Zo krijgen de vakdocenten de tools die ze nodig hebben voor consistente beoordeling en inclusie.
- *Faciliteren centraal archief SE*: Het helpt als alle beoordelingsmodellen, matrijzen en afname-instructies in toegankelijke dossiers worden opgeslagen en er duidelijke regels zijn omtrent de opslag ervan (bijvoorbeeld een sectiemap of gedeelde drive met duidelijke bestandsnaam). De schoolleiding kan een centraal en doorzoekbaar archief introduceren voor alle SE-materialen, wat hergebruik en naleving van regelgeving vergemakkelijkt en ervoor zorgt dat zowel de examencommissies als vaksecties eenvoudig toegang hebben tot deze materialen.
- *Monitoring*: De schoolleiding zou verder, bijvoorbeeld via de examencommissie, enkele eenvoudige kwaliteitsindicatoren kunnen monitoren (bijvoorbeeld % SE's met toetsmatrijs, % SE's met beoordelingsmodel, kalibraties).
- *Stimulering gebruik toetsmatrijs of andere hulpmiddelen:* Het helpt wanneer vanuit de schoolleiding het gebruik van toetsmatrijzen wordt gestimuleerd, bijvoorbeeld door een format beschikbaar te stellen en minimale eisen vast te leggen die bij elk SE horen. Vanuit de schoolleiding kunnen ook richtlijnen opgesteld worden voor het hergebruik van CE-materiaal.
- *Vastleggen minimale toegankelijkheidseisen*: Vanuit de schoolleiding kunnen minimale toegankelijkheidseisen worden opgesteld (layout, lettergrootte, alternatieven voor praktische opdrachten) om onbedoelde uitsluiting van bijvoorbeeld NT2-leerlingen, leerlingen met dyslexie of leerlingen met fysieke beperkingen te voorkomen.
- *Kalibratie en collegiale toetscontrole:* De schoolleiding kan kalibratiemomenten verplicht stellen en steekproeven laten uitvoeren door de sectie of de examencommissie voor en na afname.

4.4.4 Aanbevelingen voor vakverenigingen

- *Structurele aandacht voor SE's:* Zorg als beroepsgroep voor blijvende aandacht voor de kwaliteit van schoolexamens. Laat het onderwerp regelmatig terugkomen in communicatie, kennisdeling en professionalisering, zodat het bewustzijn levend blijft en kwaliteitsverbetering duurzaam wordt opgepakt. Dit kan bijvoorbeeld via themasessies, publicaties of het delen van goede praktijkvoorbeelden.
- *Ontwikkel voorbeeldsets SE-items*: Vakverenigingen ontwikkelen en delen collectief voorbeeldsets van SE-vragen. Het gaat hierbij om opgaven die door experts binnen de vakvereniging zijn beoordeeld en goedgekeurd als van voldoende kwaliteit, vergelijkbaar met de aanpak van het KNAG. Dit helpt

docenten op weg en zorgt voor een kwaliteitsstandaard binnen het vak. Begin met een pilot waarin 10 tot 20 representatieve vragen worden verzameld of samengesteld. Na de evaluatie en mogelijk het uittesten van deze opgaven, kunnen deze centraal worden opgeslagen in een zogenaamde gedeelde itembank. Het gebruik kan dan ook verder gemonitord worden. Zo hoeft niet elke vakdocent het wiel opnieuw uit te vinden en kan collectief gekomen worden tot kwalitatief goede opgaven, zonder afhankelijkheid van CE's en methodetoetsen.

- *Veilige digitale omgeving voor materiaaluitwisseling en effectieve communicatie*: Richt een besloten, gebruiksvriendelijk digitaal platform in voor het veilig delen, beoordelen en ontwikkelen van SE-materiaal. Combineer dit met een vaste SE-rubriek in nieuwsbrieven, vakbladen en congresprogramma's om zichtbaarheid en betrokkenheid te vergroten. Zorg voor duidelijke gebruiks- en auteursrechtregels om vertrouwen te waarborgen.
- *Gerichte scholing en train-de-trainer traject***:** Ontwikkel een flexibel scholingsaanbod rondom toetsconstructie en SE-kwaliteit, gecombineerd met een train-de-trainer model binnen vakverenigingen. Zo vergroot je de toetsdeskundigheid duurzaam en op schaal. Start met het selecteren en trainen van een kleine groep docenten die vervolgens collega's kunnen ondersteunen.

Dankwoord en verantwoording

Onze dank gaat uit naar iedereen die een bijdrage heeft geleverd aan dit rapport. Dat betreft de respondenten op de vragenlijsten, de docenten die schoolexamens ter beschikking hebben gesteld, en de geïnterviewden mensen van de vakverenigingen; de beoordelaars van de schoolexamens; iedereen die bijgedragen heeft aan de instrumentontwikkeling (vragenlijst en beoordelingsmodel); iedereen die kritisch meegelezen heeft aan (delen van) het rapport; en iedereen die op een andere manier een bijdrage heeft geleverd.

Dit onderzoek is uitgevoerd door Team Onderwijskundig binnen CitoLab (Stichting Cito)

Opzet van het onderzoek:	Bas Hemker, Max van Haastrecht, Britt van Vliet
Vragenlijstonderzoek:	Feline de Wit, Max van Haastrecht, Bas Hemker, Tessa de Vries, Britt van Vliet
Beoordeling schoolexamens:	Anneke Dubbeld, Tessa de Vries, Bas Hemker, Max van Haastrecht
Onderzoek vakverenigingen:	Teun Aalbers
Tekstbijdragen:	Anke Brons, Anneke Dubbeld, Bas Hemker, Emmelien van der Scheer, Feline de Wit, Kimberley Lek
Redactie:	Bas Hemker, Kimberley Lek, Emmelien van der Scheer