

ONGERUBRICEERD

Defence, Safety & SecurityKampweg 55
3769 DE Soesterberg
Postbus 23
3769 ZG Soesterbergwww.tno.nl

T +31 88 866 15 00

TNO-rapport**TNO 2020 R10174****Tussentijdse evaluatie van de e-screener
'psychische gesteldheid van
wapenverlofaanvragers'**

Datum	13 februari 2020
Auteur(s)	TNO Defence, Safety & Security
Aantal pagina's	16
Aantal bijlagen	niet van toepassing
Opdrachtgever	Ministerie van Justitie en Veiligheid
Projectnaam	E-screener
Projectnummer	060.32357

Alle rechten voorbehouden.

Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook, zonder voorafgaande toestemming van TNO.

Indien dit rapport in opdracht werd uitgebracht, wordt voor de rechten en verplichtingen van opdrachtgever en opdrachtnemer verwezen naar de Algemene Voorwaarden voor opdrachten aan TNO, dan wel de betreffende terzake tussen de partijen gesloten overeenkomst.

Het ter inzage geven van het TNO-rapport aan direct belanghebbenden is toegestaan.

© 2020 TNO

ONGERUBRICEERD

Inhoudsopgave

1	Inleiding	3
1.1	Aanleiding	3
1.2	Ontwikkeling van de e-screener	4
2	Validatie van de e-screener	5
2.1	Doelstelling	5
2.2	Onderzoeksvragen	5
2.3	Aanpak.....	6
2.4	Deelnemers aan het onderzoek	7
2.5	Conclusies	8
2.6	Aanbevelingen	9
3	Aanpassingen en evaluatie na ingebruikname	11
3.1	Aanpassingen voorafgaand aan de implementatie	11
3.2	Tussentijdse evaluatieresultaten na invoering	12
3.3	Conclusie en aanbevelingen	15

1 Inleiding

Het Ministerie van Justitie en Veiligheid (J&V) heeft TNO verzocht om een samenvattende rapportage met betrekking tot het uitgevoerde onderzoek naar de 'e-screener psychische gesteldheid van wapenverlofaanvragers'. Dit rapport beschrijft achtereenvolgens wat voor instrument de e-screener is, hoe dit door TNO is gevalideerd, welke aanpassingen er zijn gedaan en op welke wijze door TNO wordt gemonitord en geëvalueerd hoe de implementatie van de e-screener in de praktijk verloopt. Hierbij wordt gebruik gemaakt van het door TNO uitgevoerde validatieonderzoek uit 2017 (513 deelnemers) en van daadwerkelijke wapenverlofaanvragers in de periode 1 oktober 2019 t/m 4 februari 2020 (1.666 deelnemers). Dit leidt tot conclusies over de waarde van de e-screener als meetinstrument en op welke wijze dit dient te worden gebruikt.

1.1 Aanleiding

Naar aanleiding van het schietincident in Alphen aan de Rijn (2011) is het onderzoek onder de aanvragers van een wapenverlof aangescherpt. In lijn met de vuurwapenrichtlijn (EU2017/853) is in juli 2019 een wijziging van de Wet Wapens en Munitie geïmplementeerd¹. Uitgangspunt in deze wet is dat vuurwapens in principe verboden zijn. De lat voor de afgifte van een wapenvergunning is door de Wetgever hierbij zeer hoog gelegd: 'reeds bij geringe twijfel zal de beoordeling van de aanvraag negatief zijn'.

Onderdeel van de aanvraagprocedure is de 'e-screener psychische gesteldheid van wapenverlofaanvragers'. Dit is een digitale vragenlijst waarvan de resultaten gebruikt worden bij het risicotaxatieproces of iemand een wapen kan worden toevertrouwd. Naast de e-screener informatie wordt hiertoe verdere informatie verkregen uit een antecedentenonderzoek, een door de aanvrager verstrekt inlichtingenformulier, het nagaan van opgegeven referenten, en bezoek aan huis in het kader van de kluiscontrole. De e-screener is bedoeld als vervanging van het WM-32 formulier. Zowel de *nieuwe* aanvragers (ongeveer 5.000 per jaar) als *bestaande* wapenverlofhouders (ongeveer 60.000 in totaal) moeten verplicht meewerken aan de e-screener. Bij een negatieve uitslag op de e-screener, dat wil zeggen een 'ongunstige' uitslag waar mogelijke risicofactoren zijn geïdentificeerd, kan de aanvrager in administratief beroep gaan. Voor wat betreft de aspecten die in de e-screener aan de orde komen kan een beroep onderbouwd worden middels het leveren van contra-indicaties in de vorm van bijvoorbeeld medische of psychologische verklaringen.

De e-screener is in oktober 2019 door de Politie in gebruik genomen. Hier is een langdurig ontwikkel- en validatietraject aan vooraf gegaan. Het Trimbos-instituut heeft de e-screener in eerste aanleg ontwikkeld en TNO is verantwoordelijk voor de validatie, (tussentijdse) evaluatie en bijstelling van de e-screener. De e-screener wordt individueel en onder toezicht op een besloten locatie afgenomen. De uitslag van de e-screener wordt op een later moment door de Politie aan de aanvrager teruggekoppeld.

¹ Zie voor een uitgebreidere toelichting: Kamerbrief 'Implementatie e-screener' (7 oktober 2019) en Kamerbrief 'Bijstelling uitvoeringspraktijk e-screener' (29 oktober 2019).

1.2 Ontwikkeling van de e-screener

In opdracht van de Ministeries van Volksgezondheid, Welzijn en Sport en van Justitie en Veiligheid heeft het Trimbos-instituut de e-screener ontwikkeld. De eerste digitale versie was gereed in 2013. Tijdens deze ontwikkelfase zijn de doelstellingen van de e-screener als volgt vastgesteld:

- Het – uitsluitend en alleen – screenen van de *psychische* gesteldheid van een aanvrager voor een wapenvergunning;
- Het geven van advies inzake risicotaxatie dat de persoon het wapen niet oneigenlijk zal gebruiken ('Is aan deze persoon een vuurwapen toe te vertrouwen?');
- Onderdeel kunnen zijn van het behandelingsproces van de gehele wapenvergunningaanvraag.

Op basis van literatuuronderzoek en de input van diverse experts in enkele workshops heeft het Trimbos-instituut de belangrijkste risicofactoren voor mogelijk wapenmisbruik geïdentificeerd (bijvoorbeeld suicidaliteit, psychose, impulsiviteit, middelengebruik). Vervolgens zijn voor het meten van deze risicofactoren bestaande, reeds gevalideerde schalen uit beschikbare vragenlijsten in de literatuur en gerelateerde beroepspraktijk gezocht en opgenomen in de e-screener. Voor elke schaal is een afkappunt bepaald op basis van de betrokken handleidingen, de literatuur en/of expertoordelen. Eén van de schalen meet 'sociale wenselijkheid': de mate waarin een aanvrager de antwoorden te positief of te 'braaf' en daarmee niet helemaal naar waarheid heeft ingevuld.

Bij gebrek aan empirische data op het gebied van wapenmisbruik is met behulp van een specifieke methode, de 'conjunctmeting'-methode, door experts (o.a. psychologen, psychiaters, deskundigen bij politie) vastgesteld bij welke (combinaties van) uitslagen op de risicofactoren, gebaseerd op de afkappunten, het verlenen van een wapenvergunning als (te) risicovol beoordeeld zou kunnen worden. Dit heeft in 2013 geleid tot een eerste normering van de e-screener. De testuitslag wordt gegeven in de vorm van een advies met als opties 'Rood' (negatief of afwijzend advies) en 'Groen' (positief advies).

Zoals elk meetinstrument moest ook de e-screener gevalideerd en gekalibreerd worden, voordat deze in gebruik genomen kon worden, door de e-screener af te nemen bij verschillende groepen deelnemers. Dit validatieonderzoek is op verzoek van het Ministerie van J&V in 2017 uitgevoerd door TNO (zie hoofdstuk 2).

2 Validatie van de e-screener

TNO heeft in het validatieonderzoek voor de dataverzameling bij een representatieve steekproef uit de Nederlandse bevolking samengewerkt met bureau GfK (inmiddels deels overgenomen door de firma Ipsos). Het Trimbos-instituut heeft TNO ten behoeve van het onderzoek de benodigde achtergrondinformatie en advies gegeven over de e-screener. De volgende brancheverenigingen hebben een bijdrage geleverd in de vorm van deelname van wapenverlofhouders aan het onderzoek: Koninklijke Nederlandse Schietsport Associatie (KNSA), Koepel Nederlandse Traditionele Schutters (KNTS), Nederlandse Vereniging ter Bevordering en Instandhouding van Wapenverzamelingen Edouard de Beaumont, en de Koninklijke Nederlandse Jagersvereniging. Dit onderzoek is uitgevoerd in de periode juni 2017 tot en met februari 2018.

In dit hoofdstuk wordt toegelicht hoe dit onderzoek is uitgevoerd, op welke wijze dit heeft bijgedragen aan de totstandkoming van een kwalitatief hoogwaardig instrument en hoe het instrument, gegeven de doelstelling van de wetgever, ingezet dient te worden.

2.1 Doelstelling

De doelstelling van het onderzoek was tweeledig:

- Het vaststellen van de geschiktheid van de e-screener als meetinstrument – gegeven de geïdentificeerde risicofactoren – om af te nemen bij aanvragers van een wapenvergunning.
- Het identificeren van mogelijk noodzakelijke of gewenste aanpassingen aan de e-screener, bijvoorbeeld met betrekking tot schalen, items, antwoordmogelijkheden, of de beslisregels om de einduitslag van de e-screener te bepalen.

2.2 Onderzoeksvragen

Gegeven de doelstelling stonden in het onderzoek drie onderzoeksvragen centraal:

2.2.1 *Wat is de psychometrische kwaliteit van de e-screener?*

De psychometrische kwaliteit heeft betrekking op de e-screener als meetinstrument: wordt het juiste gemeten en wordt dat op een betrouwbare manier gemeten? Het betreft de interne consistentie (samenhang tussen de items, behorende bij een bepaalde schaal), de mate waarin het instrument robuust is voor sociaal wenselijke antwoorden, en de constructvaliditeit (de logische samenhang met andere schalen die dezelfde betekenis hebben). Aanpassen, toevoegen of verwijderen van bepaalde schalen en/of items kunnen manieren zijn om de psychometrische kwaliteit van het instrument indien nodig te verbeteren.

2.2.2 *Hoe kan de e-screener worden gekalibreerd?*

De afkappunten en beslisregels in de eerste versie van de e-screener zijn in het onderzoek van het Trimbos-instituut vooraf bepaald, zonder te beschikken over data van daadwerkelijke kandidaten. De kalibratie van de e-screener betreft het 'inregelen' of aanpassen van deze afkappunten en beslisregels op basis van de daadwerkelijke afname van de e-screener bij verschillende groepen, die naar verwachting een verschillend antwoordpatroon zouden moeten hebben.

2.2.3 Wat is de praktische bruikbaarheid van de e-screener?

In het onderzoek is tevens aandacht besteed aan de praktische bruikbaarheid van de e-screener. De gebruikers van het instrument zijn politiefunctionarissen en de aanvragers van een wapenvergunning. Voor wat betreft de politiefunctionarissen is het van belang dat zij de uitkomsten goed kunnen interpreteren en kunnen inpassen in het te volgen beoordelingsproces voor het aanvragen van een wapenvergunning. Voor de aanvragers is het belangrijk dat het afnemen van de vragenlijst en het ontvangen van de uitslag geen praktische noch mentale problemen oplevert.

2.3 Aanpak

Het onderzoek heeft zich beperkt tot de kwaliteit van de e-screener als meet-instrument, ofwel de *interne validiteit*. Worden de risicofactoren op een betrouwbare en juiste manier gemeten waarmee inzicht verkregen kan worden in de psychische gesteldheid van personen om daarmee het risico op wapenmisbruik zo veel mogelijk uit te sluiten? Het onderzoek heeft zich niet gericht op de *externe* of *predictieve validiteit*, namelijk in welke mate het instrument correct voorspelt of iemand misbruik zou maken van zijn wapen. Dit vereist longitudinaal onderzoek op basis van (toekomstige) incidenten met wapenmisbruik. Omdat dergelijke incidenten zeer weinig voorkomen, is zo'n onderzoek vrijwel niet uitvoerbaar, zeker niet in de fase waarin de e-screener nog niet in gebruik is genomen.

Als alternatief is een andere methodologie gebruikt om de e-screener te valideren en te kalibreren, namelijk het contrast tussen een *laagrisicogroep* met verwachte uitslag 'Groen' (positief advies), en een *hoogrisicogroep* met verwachte uitslag 'Rood' (negatief of afwijzend advies; zie paragraaf 2.4 hoe deze groepen zijn vastgesteld). De gedachte hierbij is dat zowel onterecht 'Groen' (geen ingeschat risico, terwijl iemand wel risico-verhogende kenmerken heeft) als onterecht 'Rood' (wel ingeschat risico, terwijl iemand geen risico-verhogende kenmerken heeft) in ongeveer gelijke mate ongewenst is.

Ook andere (meer beleidsmatige) keuzes zouden gemaakt kunnen worden waarbij bijvoorbeeld meer nadruk komt te liggen op het vermijden van onterecht 'Groen' en minder op het vermijden van onterecht 'Rood'. Hierbij dient men zich echter te realiseren dat een *afname* van onterecht 'Groen' altijd leidt tot een *toename* van onterecht 'Rood'. In dit onderzoek had TNO te weinig informatie over kosten of ernst van de gevolgen om op een verantwoorde manier een andere verhouding te kiezen dan het gekozen maximale contrast. Hoewel de statistiek op basis van het gekozen maximale contrast goede handvatten biedt voor (het aanpassen van) de beslisregels, zijn afwegingen daarom ook afhankelijk van beleidsmatige keuzes en praktische uitvoering, bijvoorbeeld op hoeveel risico op 'onterecht groen' het Ministerie van Justitie en Veiligheid en de politie het instrument afgesteld willen hebben, en hoeveel aanvullend onderzoek in het aanvraagproces kan worden uitgevoerd, naast de e-screener.

Het onderzoek heeft in verband met het vertrouwelijke karakter in een aantal sessies onder toezicht op diverse locaties bij GfK plaatsgevonden; afname via internet (op afstand) was niet toegestaan. Naast de e-screener is steeds ook een vragenlijst met zes extra schalen afgenomen. Deze zes schalen, op basis van vakliteratuur, komen in betekenis overeen met bepaalde schalen in de e-screener.

Schalen met dezelfde betekenis zouden derhalve sterk met elkaar moeten samenhangen omdat ze hetzelfde concept meten. Hiermee kon de (convergente) validiteit beter worden onderzocht.

2.4 Deelnemers aan het onderzoek

Ter beantwoording van de onderzoeksvragen is de e-screener in 2017 eenmalig afgenomen bij 513 deelnemers. De keuze van dit totale aantal proefpersonen is gemaakt op basis van een statistische methodiek in combinatie met een aantal randvoorwaarden zoals de periode waarin de data verzameld moesten worden (september – november 2017), de praktische organisatie in verband met verplichte afname op locatie, en de kosten zoals vergoedingen aan de deelnemers.

De proefpersonen zijn onderverdeeld in subgroepen:

- **Hoogrisicogroep** (N=107): een groep deelnemers met risicoverhogende kenmerken op mogelijk wapenmisbruik zoals een psychiatrisch verleden of agressief gedrag.
- **Laagrisicogroep**: deze bestaat uit twee type deelnemers:
 - Representatieve groep (N=303): een steekproef uit de Nederlandse bevolking, 95% mannen en evenredig verdeeld over algemene kenmerken als leeftijd en opleiding, in overeenstemming met aanvragers van een wapenvergunning, maar zonder de risicoverhogende kenmerken.
 - Wapenverlofhouders (N=103): bestaande bezitters van een wapenvergunning.

De hoogrisicogroep en de representatieve groep zijn geworven uit een database van GfK met mensen die zich bereid hebben verklaard om aan (markt)onderzoeken deel te nemen. Om de hoogrisicogroep te kunnen selecteren uit deze database is eerst een voorlopige steekproeftrekking gehouden. Hierbij werd een korte vragenlijst afgenomen waarmee risicoverhogende kenmerken werden gemeten, de *pre-screener* genoemd (respons: 54.025 deelnemers uit de totale database met 102.827 mensen). De groep wapenverlofhouders is geworven via de brancheverenigingen.

Hierbij dient te worden opgemerkt dat met de pre-screener voor een deel vergelijkbare risicoverhogende kenmerken als met de e-screener zelf werden gemeten. Dit betekent dat de hoogrisicogroep niet geheel onafhankelijk gekozen is en daardoor niet het meest optimale criterium of 'de absolute waarheid' is, maar op dit moment was dit het beste wat voorhanden was om de e-screener zo goed mogelijk te kalibreren. Ook hebben alle deelnemers op vrijwillige basis deelgenomen aan het onderzoek, waardoor er mogelijk filtering door zelfselectie heeft plaatsgevonden. Betrokken experts (bijvoorbeeld politiemedewerkers, het Ministerie van Justitie en Veiligheid) hebben aangegeven dat aan deelnemers in de hoogrisicogroep geen wapenvergunning verleend zou mogen worden, en de einduitslag 'Rood' zouden moeten krijgen. Voor de laagrisicogroep geldt dat zij voornamelijk einduitslag 'Groen' zouden moeten krijgen.

Ter beantwoording van de onderzoeksvraag over de praktische bruikbaarheid is een werksessie gehouden met 10 politievertegenwoordigers uit de diverse eenheden. Daarnaast heeft TNO een evaluatiegesprek gehouden met een groep wapenverlofhouders direct na afloop van hun e-screener afname. De wapenverlofhouders hebben hun ervaringen en meningen met elkaar en met TNO gedeeld. Daarnaast zijn alle deelnemers in de gelegenheid gesteld om opmerkingen over de e-screener te noteren die zijn samengevat door TNO.

2.5 Conclusies

Uit de onderzoeksresultaten is gebleken dat de e-screener als instrument meet wat het zou moeten meten. De schalen zijn voldoende betrouwbaar en valide bevonden voor wat betreft de beschikbare data in 2017. Wel bleken er aanpassingen nodig te zijn aan de afkappunten en beslisregels voordat de e-screener geïmplementeerd zou kunnen worden.

2.5.1 *Psychometrische kwaliteit*

Het onderzoek naar de psychometrische kwaliteit van de e-screener heeft uitgewezen dat de interne consistentie van de schalen voldoende is. De in de literatuur gedefinieerde grenswaarden op Cronbach's alfa werden gehaald. Het was daarom niet nodig om items te herformuleren, te verwijderen, of toe te voegen. Dit resultaat was overeenkomstig de verwachting, omdat de e-screener gebruik maakt van bestaande, reeds gevalideerde schalen. Voor de meeste schalen is ook de convergente validiteit voldoende gebleken. Uit de analyses bleek dat de e-screener schalen over het algemeen logisch met elkaar samenhangen en dat eveneens doen met de tijdens het onderzoek door TNO toegevoegde schalen. Met name één van de zes toegevoegde schalen bleek duidelijk onderscheid te maken tussen de laag- en de hoogrisicogroep. TNO heeft aanbevolen om deze extra schaal toe te voegen aan de schalen die reeds door het Trimbos-instituut waren geselecteerd.

Uit de resultaten van het onderzoek bleek ook dat sommige wapenverlofhouders – die een intrinsiek belang hebben bij een positieve uitkomst van de e-screener – geneigd waren om de e-screener sociaal wenselijk in te vullen. Dit had echter bij deze onderzoeksgroep geen vertekende invloed op de scores die zij hebben gegeven op de (andere) risicofactoren. Dit komt omdat deze groep geen reden had om sociaal wenselijk te antwoorden: de uitkomst had voor hen geen consequenties. Bij de daadwerkelijke aanvragers in oktober 2019 bleek dit echter wel het geval te zijn (zie paragraaf 3.2.1).

2.5.2 *Kalibratie van de e-screener*

Uit de data-analyses bleek dat op basis van de beslisregels de hoogrisicogroep vaker 'Rood' scoorde dan de laagrisicogroep, maar dat de eerste versie van de e-screener erg 'streng' stond afgesteld (leidde tot een relatief hoog percentage 'Rood'). In totaal had 81% van de deelnemers binnen de hoogrisicogroep in eerste instantie de einduitslag 'Rood'. Dit hoge percentage 'Rood' is volgens verwachting omdat deze groep naar de mening van experts in principe niet in aanmerking zou moeten komen voor een wapenvergunning (criterium 'geringe twijfel' is daar zeker van toepassing). Een percentage van 100% zou eigenlijk wenselijk zijn, maar is niet realistisch vanwege de onoverkomelijke meetfouten die er altijd zijn. Binnen de laagrisicogroep was het percentage deelnemers met einduitslag 'Rood' weliswaar

een stuk lager, namelijk 54% onder de door GfK geworven representatieve groep, en 44% onder de groep wapenverlofhouders, maar nog steeds hoog. Dit zou betekenen dat bijna de helft van de bestaande groep hun wapenverlof zou moeten inleveren. Dat is een onwaarschijnlijk hoog percentage en maatschappelijke gezien onacceptabel. Met andere woorden, de beslisregels in de eerste versie van de e-screener hebben niet geleid tot het gewenste en verwachte resultaat.

Nader onderzoek door TNO liet zien dat de door het Trimbos-instituut aangezochte experts erg voorzichtig zijn geweest in hun inschattingen over het risico op mogelijk wapenmisbruik bij de diverse combinaties van risicofactoren (de profielen). Dit is in principe in lijn met het wettelijk kader dat reeds bij 'geringe twijfel' geen wapenvergunning verleend mag worden, maar in de praktijk zou dit leiden tot een onrealistisch hoog percentage afwijzingen. Daarbij hebben de experts zich in vrij sterke mate laten leiden door een veronderstelde mate waarin de e-screener sociaal wenselijk is ingevuld, waardoor ze extra streng zijn geweest.

Bovendien is onderzocht in welke mate de afkappunten goed gekozen waren. Uit 'ROC'-analyses, waarmee is bekeken met welke afkappunten het contrast tussen de laag- en de hoogrisicogroep het grootste wordt, bleek dat er een betere manier was om de afkappunten te kiezen dan deze in de eerste versie van de e-screener waren opgenomen.

2.6 Aanbevelingen

Vanuit de conclusies zijn op basis van het onderzoek in 2017 een aantal aanbevelingen gedaan om de e-screener te verbeteren alvorens deze te implementeren.

2.6.1 Schalen

Op basis van de onderzoeksresultaten zijn enkele aanbevelingen gedaan voor het aanpassen van formuleringen van items, zodat ze eenduidiger zijn en er meer variatie ontstaat in de antwoorden. Daarnaast is aangeraden om een extra risicofactor, die mogelijk wijst op een instabiele persoonlijkheid, toe te voegen. Deze schaal, die in het onderzoek aanvullend is getest, betreft een ander type risicofactor (niet-klinisch), en bleek goed te differentiëren tussen de laag- en de hoogrisicogroep.

2.6.2 Afkappunten en beslisregels

Op basis van een vergelijking van alternatieven zijn concrete aanbevelingen gedaan voor het aanpassen van afkappunten en beslisregels. Hierbij is benadrukt dat er nog wel een monitorings- en evaluatiefase nodig is na ingebruikname van de e-screener omdat de daadwerkelijke groep aanvragers kan afwijken van de onderzoeksgroep. Op basis van de monitorings- en evaluatiefase dienen wellicht afkappunten en beslisregels te worden aangepast (gekalibreerd).

2.6.3 Weergave van de uitkomsten

Ten behoeve van de praktische bruikbaarheid door politiemedewerkers is gediscussieerd over de wijze waarop de uitkomsten van de e-screener moet worden getoond in het systeem. Het werd bovendien aanbevolen om de politie-medewerkers en andere betrokkenen die gaan werken met de e-screener een training te geven over het gebruik van het instrument en de betekenis van de

diverse schalen en uitkomsten. Dat zou hen helpen bij het gebruik en de interpretatie van de uitkomsten van de e-screener.

2.6.4 *Inzet van de e-screener in de praktijk*

De correcte werking van de e-screener wordt niet alleen wordt bepaald door de psychometrische kwaliteit, maar ook door de manier waarop deze wordt gebruikt en toegepast. Zo heeft een vragenlijst als screeningsinstrument per definitie een zekere mate van onbetrouwbaarheid in zich, omdat elke vragenlijst door een aanvrager sociaal wenselijk kan worden ingevuld. Ook kunnen er begrips- of taalproblemen voorkomen. Daarnaast is het invullen van een vragenlijst een momentopname; de situatie rondom de wapenverlofaanvrager en zijn psychische gesteldheid kunnen over de tijd veranderen. De e-screener moet daarom altijd onderdeel zijn van een uitgebreidere procedure, waarin op basis van meer informatie dan alleen de resultaten van de e-screener wordt besloten om iemand al dan geen wapenvergunning te verlenen. Bovendien verdienen aanvragers die de e-screener extra sociaal wenselijk ingevuld blijken te hebben, specifieke aandacht bij het vervolgonderzoek door de politie.

Bovendien blijft de e-screener beperkt tot een screening op de *psychische gesteldheid*, terwijl ook andere risicofactoren een rol kunnen spelen, zoals bijvoorbeeld een crimineel verleden of radicalisering en extremisme. Deze risicofactoren² maken echter wel deel uit van de in de wetgeving genoemde risicofactoren en moeten op een andere manier geïdentificeerd worden.

2.6.5 *Continue evaluatie*

Op de langere termijn is het zinvol om de data van de e-screener (geanonimiseerd) te blijven opslaan in een database, zodat men op gezette tijden kan blijven monitoren en controleren of de e-screener naar behoren werkt. Dit betreft zowel de selectie van schalen als de afkappunten en beslisregels.

Het zou aan te bevelen zijn als in de toekomst incidenten of bijna-incidenten van wapenmisbruik (binnen het legaal bezit van vuurwapens) worden beschreven en bewaard blijven, zodat deze gekoppeld kunnen worden aan de ingevulde e-screener data van de daders en nader kunnen worden geanalyseerd.

² In de CWM: in punt 1.2 als 'Specifieke kenmerken van de aanvrager' genoemd.

3 Aanpassingen en evaluatie na ingebruikname

Een volgende fase betrof het doorvoeren van aanpassingen aan de eerste versie van de e-screener op basis van de aanbevelingen uit het onderzoek van TNO in 2017, gevolgd door een fase van monitoring en evaluatie in de eerste periode na ingebruikname van de e-screener. Hiertoe is door het Ministerie van Justitie & Veiligheid een ontwikkel- en implementatieproject gestart in juni 2019 en dit loopt de komende jaren nog door. In het project zijn naast het Ministerie van Justitie & Veiligheid (DGPOL, Justis) en de politie, de volgende drie partijen betrokken:

- Ipsos is verantwoordelijk voor de technische implementatie en hosting van de e-screener software (tweede versie op een nieuw platform);
- TNO doet onderzoek naar de afstelling of normering van de e-screener, en verzorgt (tussentijdse) evaluaties;
- Lamark regelt de afname van de e-screener onder toezicht op locatie.

3.1 Aanpassingen voorafgaand aan de implementatie

Er zijn diverse aanpassingen uitgevoerd aan de e-screener zoals deze door het Trimbos-instituut initieel was opgezet, voordat het instrument per 1 oktober 2019 in gebruik is genomen.

3.1.1 Technische aanpassingen

Los van gewenste functionele aanpassingen diende de e-screener opnieuw te worden geprogrammeerd. De versie van het Trimbos-instituut was op een inmiddels verouderd software platform ontwikkeld in 2013. In de tweede versie is zoveel als mogelijk de *look and feel* hetzelfde gehouden als in de eerste versie.

3.1.2 Schalen

De aanbevelingen uit het TNO onderzoek met betrekking tot de schalen zijn door de opdrachtgever volledig overgenomen en in de nieuwe versie als zodanig geïmplementeerd. Dit houdt in dat enkele formuleringen zijn aangepast en de extra persoonlijkheidsschaal is toegevoegd. Bovendien is de volgorde waarin de diverse schalen aan bod komen verbeterd op basis van de eerdere gebruikerservaringen.

3.1.3 Afkappunten en beslisregels

Als hulpmiddel voor het aanpassen van de afkappunten en beslisregels is een simulatietool in Excel gemaakt, om de diverse varianten met elkaar te kunnen vergelijken. Er kan hiermee worden bekeken hoe de einduitslagen van de 513 deelnemers in het onderzoek veranderen bij andere instellingen van de afkappunten, de knock-outs en bepaalde combinaties van risicofactoren (gewogen som). Bij het inregelen is steeds een zo groot mogelijk contrast tussen de laagrisicogroep en de hoogrisicogroep als criterium gebruikt (zie paragraaf 2.3).

Een set potentieel geschikte varianten is voorgelegd aan de vertegenwoordigers van het Ministerie van Justitie en Veiligheid (DGP&V, Justis) en de Nationale Politie in twee bijeenkomsten (*i.c.* op 1 juli 2019 en op 12 augustus 2019). Meerdere argumenten hebben meegespeeld in de besluitvorming. De keuzes zijn in de eerste plaats gebaseerd op de statistische analyses, maar de beslissingen kunnen niet puur en alleen op basis van de statistiek genomen worden. Ook inhoudelijke afwegingen speelden een rol, zoals het besluit om bij bepaalde psychiatrische

problematiek *altijd* een negatief advies door de e-screener te laten geven. Steeds is het uitgangspunt geweest dat een aanvrager of verlofhouder reeds 'bij geringe twijfel' een negatief advies krijgt, conform de Wet Wapens en Munitie³.

Dit heeft geleid tot de uiteindelijke afkappunten en beslisregels zoals deze zijn geïmplementeerd in de tweede versie van de e-screener. Er wordt een combinatie van verschillende soorten beslisregels gebruikt, waarbij zowel combinaties van risicofactoren als een score op één enkele risicofactor kunnen leiden tot de einduitslag 'Rood' op de e-screener. Hierdoor wordt er vanuit meerdere invalshoeken gekeken naar de mate van risico op de verschillende risicofactoren. Binnen de groep geteste kandidaten wordt als het ware steeds een andere 'dwarsdoorsnede' gemaakt. Bovendien worden daarmee *alle* risicofactoren meegenomen, ook al tellen sommige factoren bijvoorbeeld in een eerder toegepaste regel minder zwaar mee.

3.2 Tussentijdse evaluatieresultaten na invoering

Vanaf 1 oktober 2019 is gestart met het invullen van de e-screener door nieuwe aanvragers én bestaande verlofhouders. TNO heeft voortdurend gemonitord hoe dit proces verliep en hoe de aanvragers hebben gescoord op basis van analyses van de data. De belangrijkste doelstelling van de tussentijdse evaluaties, die op het moment van verschijnen van deze rapportage nog steeds voortduren, is om te monitoren of de aanvragers in de e-screener vergelijkbaar scoren als de onderzoeksgroep uit 2017 (zie paragraaf 2.4), omdat hierop de afkappunten en beslisregels zijn gebaseerd. Indien dit niet zo is, geeft dit te zijner tijd mogelijk aanleiding tot aanpassingen.

De eerste tussentijdse evaluatie heeft half oktober 2019 plaatsgevonden (N=271). Vervolgens is dit herhaald half november (N=1.235), half december (N=1.428) en begin februari (N=1.666). In oktober is gestart met het testen van relatief oudere en bestaande wapenverlofhouders, en was het aantal geteste mensen het grootst (N=1.085). Vanaf 1 november zijn hoofdzakelijk *nieuwe* aanvragers getest, aangezien het testen van bestaande wapenverlofhouders vanaf dat moment door de Minister voor twee jaar was opgeschort⁴. Hierdoor is gedurende de periode van 1 november tot en met 4 februari ook het aantal geteste aanvragers sterk gedaald (N=581).

3.2.1 Sociale wenselijkheid

Veruit het grootste verschil in scores tussen de daadwerkelijke aanvragers en de onderzoeksgroep uit 2017 betreft de score op sociale wenselijkheid. Waarschijnlijk speelt mee dat de aanvragers met name in oktober 2019 bang waren dat zij hun wapenverlof zouden kwijtraken. Deze consequentie was er niet voor de deelnemers aan het onderzoek in 2017. Het afkappunt op sociale wenselijkheid is daarbij zo extreem gekozen dat bij een score 'Rood' op sociale wenselijkheid het vrijwel zeker is dat de aanvrager zichzelf positiever of 'braver' heeft voorgedaan dan hij in werkelijkheid is, en dat hij de vragenlijst (deels) niet naar waarheid heeft ingevuld. Bij sociale wenselijkheid zijn de antwoorden op de e-screener als geheel onbetrouwbaar en er kan dan geen goede risicotaxatie worden gedaan.

³ Zie Kamerbrief 'Implementatie e-screener' (7 oktober 2019).

⁴ Zie Kamerbrief 'Bijstelling uitvoeringspraktijk e-screener' (29 oktober 2019).

Bij de eerste 271 aanvragers was het percentage met een extreme score op sociale wenselijkheid maar liefst 17%. Dit percentage is met name sinds november 2019 veel lager geworden. Van de aanvragers die na 1 november zijn getest, betreft dit nog geen 4%. Voor de totale groep die vanaf oktober is getest (N=1.666) komt dit percentage uit op gemiddeld 11%. Dit percentage is nog altijd veel hoger dan bij de in 2017 geteste laagrisicogroep (0,3%), hoogrisicogroep (0%) en wapenverlofhouders (2%).

We zien in de daling van de mate van sociale wenselijkheid een verschil tussen bestaande wapenverlofhouders (vooral getest in oktober) en de nieuwe aanvragers (vooral getest vanaf november) die nog geen wapenverlof hebben, en voor wie de consequenties wellicht minder groot zijn. In de huidige data kan geen onderscheid worden gemaakt tussen nieuwe aanvragers en bestaande wapenverlofhouders, dus het is niet volledig zeker of dit alleen hierdoor wordt verklaard. Ook de communicatie over sociale wenselijkheid in de pers vanaf oktober zal hebben bijgedragen tot afname van sociaal wenselijk ingevulde vragenlijsten.

We zien ook een trend in leeftijd: van de daadwerkelijk geteste 70-plussers had 21% een extreme score op sociale wenselijkheid. Dit relatief hoge percentage hangt waarschijnlijk deels samen met het feit dat deze oudere generatie (bestaande wapenverlofhouders) vooral in oktober 2019 is getest. Bij de (deels nieuwe) aanvragers jonger dan 70 jaar ligt dit percentage significant lager namelijk op 7%.

Er is in de data gecontroleerd of de antwoorden op de andere schalen in de e-screener ook extra 'braaf' zijn ingevuld. De aanvragers met einduitslag 'Rood' vanwege sociale wenselijkheid blijken op bijna alle schalen significant lager te scoren dan de overige aanvragers die hier minder hoog op scoren. Deze resultaten bevestigen dat de schaal voor sociale wenselijkheid goed zijn werk doet als meting van betrouwbaarheid van de gegeven antwoorden.

3.2.2 *Einduitslag e-screener*

Gebaseerd op het totale aantal aanvragers (zowel bestaande wapenverlofhouders als nieuwe aanvragers) dat de e-screener heeft gedaan tot en met 4 februari 2020 (N=1.666), had 93% de einduitslag 'Groen' (geen risicofactoren⁵) en 7% de einduitslag 'Rood' (wel risicofactoren); deze 7% betreft in totaal 118 aanvragers. Bij vertaling naar de totale groep van 60.000 betekent 7% dat 4.200 aanvragers de einduitslag 'Rood' zouden krijgen. Het percentage weigeringen van een wapenverlof was voorheen 4%. Dit is een verschil van 3% in vergelijking met de situatie waarin de e-screener wordt gebruikt. Absoluut gezien is dit een substantieel aantal aanvragers binnen de totale groep (1.800 extra) bij wie wordt geconstateerd dat er mogelijk risicoverhogende kenmerken aanwezig zijn in hun psychische gesteldheid.

⁵ Bij deze 93% zitten ook aanvragers met een extreme score op sociale wenselijkheid, die voorheen de einduitslag 'Rood' kregen, ook als er geen andere risicofactoren waren geïdentificeerd. De maatregelen voor aanvragers met uitsluitend een extreme score op sociale wenselijkheid zijn later echter aangepast (zie Kamerbrief 'Bijstelling uitvoeringspraktijk e-screener', 29 oktober 2019). Een einduitslag 'Groen' betekent in dit rapport: er zijn geen risicofactoren geïdentificeerd.

Screenen op deze kenmerken is ook altijd het uiteindelijke doel van de e-screener geweest in het kader van weigering van een wapenverlof bij 'de geringste twijfel'. Een deel van deze aanvragers zal alsnog een wapenverlof kunnen krijgen op basis van contra-indicaties.

De groep aanvragers die sinds oktober 2019 is getest, is vergeleken met de onderzoeksgroep uit 2017. Binnen de laagrisicogroep in dit onderzoek (inclusief de geteste wapenverlofhouders) had maar liefst 26% einduitslag 'Rood'; dit percentage is significant hoger dan onder de daadwerkelijke aanvragers sinds oktober 2019 (7%). Indien we de wapenverlofhouders uit 2017 als aparte groep bekijken, dan is het verschil met de daadwerkelijke (recent geteste) aanvragers veel minder groot: slechts 9% zou destijds de einduitslag 'Rood' krijgen. Dat is 2% meer dan de huidige daadwerkelijke aanvragers, maar dit verschil is niet significant.

Over de tijd beschouwd blijft het percentage met einduitslag 'Rood' vanwege mogelijke risicofactoren vrij constant. Het percentage 'Rood' in de eerste groep in oktober 2019 is 7%, en in de tweede groep, vanaf 1 november 2019, is dat 8%. Deze percentages zijn niet significant verschillend. Hetzelfde geldt voor de leeftijds-groepen met als uitzondering de 80-plussers: hiervan heeft 16% einduitslag 'Rood', terwijl dit percentage voor de generaties daaronder significant lager is (7%).

Er zijn diverse redenen voor het feit dat de daadwerkelijke aanvragers positiever scoren dan de onderzoeksgroep in 2017. Voor de aanvragers hangt er daadwerkelijk iets vanaf, namelijk een mogelijke weigering van een wapenverlof. Dit verhoogt de kans op sociaal wenselijk ingevulde antwoorden, en dat is uit de eerste evaluaties ook gebleken (zie paragraaf 3.2.3). Dit kan bijvoorbeeld de 2% verschil verklaren met de in 2017 geteste wapenverlofhouders. Daarnaast geldt dat er al periodieke controles plaatsvinden onder degenen die al jaren een wapenverlof hebben. Juist die oudere generatie is gestart met het invullen van de e-screener. Het feit dat deze groep positiever scoort dan de gemiddelde Nederlander is vanwege hun historie dan ook niet vreemd. Dat is ook al aangetoond in het onderzoek in 2017. Het is op dit moment nog niet duidelijk of wapenverlofhouders daadwerkelijk een ander profiel hebben dan de gemiddelde Nederlander. Het kan zijn dat zij bijvoorbeeld minder vaak een risicovolle psychische gesteldheid hebben, juist omdat zij een wapen bezitten. Dit zal in de toekomst moeten blijken als er meer data beschikbaar komen, ook van *nieuwe* aanvragers en *jongere* generaties.

3.2.3 Scores op de risicofactoren

Er is ook bekeken hoe de daadwerkelijke groep aanvragers sinds oktober 2019 scoort op de afzonderlijke risicofactoren die gezamenlijk leiden tot de einduitslag van de e-screener. De reden voor einduitslag 'Rood' is voor 18 mensen gebaseerd op relatief duidelijk aanwijsbare psychiatrische problematiek zoals suïcidaliteit, specifiek medicijngebruik in combinatie met psychotische ervaringen of een psychiatrische opnamegeschiedenis. Vooraf is bepaald dat deze aanvragers niet zomaar een wapenverlof kunnen krijgen, en deze aanvragers worden er dus terecht uit gehaald. Dit geldt ook voor enkele aanvragers die op relatief veel risicofactoren hoog scoren die niet op het WM32-formulier stonden. Deze aanvragers zouden voorheen moeilijker herkend worden, en hier voorziet de e-screener dus in een toegevoegde waarde ten opzichte van de oude situatie.

In lijn met de einduitslag geldt dat de daadwerkelijke aanvragers op de meeste risicofactoren lager (minder risicovol) scoren dan de onderzoeksgroep uit 2017, maar wel ongeveer vergelijkbaar met de in 2017 geteste wapenverlofhouders.

3.3 Conclusie en aanbevelingen

3.3.1 Conclusie

In algemene zin kan worden geconcludeerd dat de e-screener als meetinstrument werkt zoals deze zou moeten werken, voor zover de data dit tot dusver laten zien. De e-screener signaleert de mogelijke aanwezigheid van risicofactoren bij een subgroep van de aanvragers zoals bedoeld. Aanvragers met een extreme score op sociale wenselijkheid worden goed herkend. Afgezien van de hogere mate van sociale wenselijkheid, zijn de scores van de daadwerkelijke aanvragers redelijk vergelijkbaar met die van de onderzoeksgroep uit 2017, met name met de indertijd geteste wapenverlofhouders. Er zijn dan ook op basis van de eerste resultaten geen aanwijzingen dat de e-screener op dit moment te streng of juist te soepel staat afgesteld. Ook is er op basis van de scoreverdelingen geen directe aanleiding tot bijstelling van beslisregels, afkappunten of andere inhoudelijke onderdelen van de e-screener (zoals schalen, items, en antwoordcategorieën).

Bij een toenemend aantal nieuwe aanvragers is de verwachting dat het percentage aanvragers met einduitslag 'Rood' nog iets zal toenemen, omdat zij mogelijk een iets ander profiel hebben dan de bestaande wapenverlofhouders bij wie al jarenlang controles worden uitgevoerd, en omdat de neiging tot sociale wenselijkheid na verloop van tijd minder bleek te worden. Hiermee zullen de verschillen in scores van de daadwerkelijke aanvragers en de in 2017 geteste wapenverlofhouders nog wat kleiner worden.

In de komende tijd wordt gemonitord of de scoreverdelingen op de diverse risicofactoren veranderen en in het najaar van 2020 wordt een uitgebreide evaluatie gehouden. Het kan zijn dat er op grond daarvan wel aanpassingen nodig blijken op basis van meer data en meer opgedane ervaring in de praktijk.

3.3.2 Aanbevelingen

De aanbevelingen hebben vooral te maken met een verbeterde afstemming tussen de mogelijkheden en de beperkingen van de e-screener als meetinstrument, en het gebruik hiervan in de praktijk. De wijze waarop de e-screener wordt ingezet in het aanvraagproces van een wapenverlof bepaalt in belangrijke mate de waarde van het instrument.

Ten eerste dienen de maatregelen die op basis van de uitkomsten van de e-screener worden genomen in verhouding te staan tot de betekenis van de beslisregels en de factoren die in de e-screener worden gemeten. Voor wat betreft sociale wenselijkheid zijn er inmiddels maatregelen genomen om het proces beter af te stemmen op de betekenis van deze schaal. Sociale wenselijkheid is enkel en alleen een indicator voor de (on)betrouwbare wijze waarop een aanvrager heeft geantwoord, en staat los van de daadwerkelijke risicofactoren in de psychische gesteldheid van de aanvrager. De maatregel om bij een extreme score op sociale wenselijkheid de psychische gesteldheid op een andere manier te evalueren is daarom passend. Sociale wenselijkheid is wel een zeer belangrijke indicator gebleken, dus moet vooral onderdeel van de e-screener blijven.

Echter, ondanks deze indicator kan het nog steeds zijn dat aanvragers, zowel degenen met einduitslag 'Groen' als 'Rood', de vragenlijst extra sociaal wenselijk hebben ingevuld, ook al hebben zij niet de meest extreme score hierop behaald. Elke vragenlijst heeft een zekere mate van onbetrouwbaarheid in zich en kan sociaal wenselijk worden ingevuld. Dit geldt ook voor de e-screener, waarbij dit bij de ene schaal gemakkelijker gaat (bijvoorbeeld middelengebruik) dan bij een minder eenduidige schaal (zoals een persoonlijkheidskenmerk). Hier moet te allen tijde rekening mee gehouden worden gedurende het proces van de wapenverlof-aanvraag. Een einduitslag 'Rood' dient, juist vanwege de neiging tot sociale wenselijkheid, wel serieus te worden genomen aangezien de gegeven antwoorden eerder 'braver' zijn dan 'ernstiger' in vergelijking met de werkelijkheid.

Tegelijkertijd dient er te worden gestreefd naar het terugdringen van sociale wenselijkheid bij het invullen van de e-screener. Hiertoe moeten de aanvragers in het vervolg in het digitale systeem aanvinken dat zij de vragenlijst naar waarheid zullen invullen, anders kunnen ze niet verder. Deze boodschap moet ook in andere vormen van communicatie naar de aanvragers toe worden benadrukt. Indien de e-screener minder sociaal wenselijk ingevuld wordt, dan neemt de waarde van het instrument in het aanvraagproces van een wapenverlof in sterke mate toe.

Vanwege de beperkingen in de betrouwbaarheid alsmede het gebrek aan evidentie voor de predictieve validiteit van de e-screener dient de besluitvorming omtrent het verlenen van een wapenverlof op meer informatie gebaseerd te zijn dan uitsluitend de uitkomsten van de e-screener. Er zijn meer risicofactoren dan die de e-screener meet, zoals bijvoorbeeld radicalisering, extremisme of een crimineel verleden. Dit is ook op deze manier ingeregeld door middel van een antecedentenonderzoek, een inlichtingenformulier, het nagaan van opgegeven referenten, en huisbezoek in de procesaanvraag voor een wapenverlof. Daarnaast kunnen de uitkomsten op de e-screener zelf mogelijk niet altijd de juiste conclusie geven, want er zullen er altijd aanvragers blijven met een onterechte einduitslag 'Groen' of onterechte einduitslag 'Rood'. Het is dus goed dat de aanvragers contra-indicaties kunnen aanleveren bij een negatieve uitkomst op de e-screener om alsnog een wapenverlof te verkrijgen.

Deskundigen (bijvoorbeeld een psycholoog of psychiater) die bij het leveren van contra-indicaties door de aanvragers worden betrokken hebben nu ook al inzage in de risicofactoren waarop negatief is gescoord, maar dit wordt nog verder verbeterd. Hiertoe vindt een technische aanpassing plaats aan de e-screener, waarbij de scores op elke risicofactor ook in percentielscores worden weergegeven in een nieuwe versie van de bij de einduitslag behorende toelichting. Tevens wordt een tekstuele toelichting per risicofactor toegevoegd, zodat een deskundige inzicht heeft in de betekenis van elke risicofactor en hoe deze in de e-screener wordt gemeten. Het zou heel nuttig zijn als in een database systematisch wordt bijgehouden hoe de diverse deskundigen oordelen over het e-screener profiel per aanvrager. Gebruikmakend van expertoordelen als extern criterium kan de e-screener in de toekomst nog beter worden gekalibreerd en afgesteld.