

De evaluatie van verkeerseducatieprogramma's

Dr. J. Mesken

R-2011-8

De evaluatie van verkeerseducatieprogramma's

Aanbevelingen voor effectmeting en een voorstel voor een verkort
meetinstrument

Documentbeschrijving

Rapportnummer:	R-2011-8
Titel:	De evaluatie van verkeerseducatieprogramma's
Ondertitel:	Aanbevelingen voor effectmeting en een voorstel voor een verkort meetinstrument
Auteur(s):	Dr. J. Mesken
Projectleider:	Dr. J. Mesken
Projectnummer SWOV:	06.2.3
Trefwoord(en):	Education; evaluation (assessment); efficiency; method; traffic; safety; quality; measurement; Netherlands; SWOV.
Projectinhoud:	Dit rapport beschrijft hoe de effectiviteit van verkeerseducatieprogramma's kan worden geëvalueerd, als onderdeel van de PDCA-cyclus. Het rapport gaat ook kort in op de ontwikkeling van een verkeerseducatieprogramma. Daarnaast bevat het een voorstel om een database met vragen op te zetten, waarmee het makkelijker wordt om evaluaties uit te voeren.
Aantal pagina's:	32 + 2
Prijs:	€ 10,-
Uitgave:	SWOV, Leidschendam, 2011

De informatie in deze publicatie is openbaar.
Overname is echter alleen toegestaan met bronvermelding.

Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV
Postbus 1090
2260 BB Leidschendam
Telefoon 070 317 33 33
Telefax 070 320 12 61
E-mail info@swov.nl
Internet www.swov.nl

Samenvatting

Om de kwaliteit van verkeerseducatieprogramma's te verbeteren, is het belangrijk dat deze programma's goed worden geëvalueerd. Nu gebeurt dat te weinig. Dit rapport bevat adviezen die uitvoerders en ontwikkelaars van verkeerseducatieprogramma's kunnen helpen om hun programma's te evalueren. Het rapport gaat in op de inhoudelijke effectmeting en dus niet op de evaluatie van procesmatige aspecten van educatieprogramma's.

De evaluatie van verkeerseducatieprogramma's is onderdeel van een groter proces van kwaliteitszorg: de zogenaamde PDCA-cyclus (PLAN, DO, CHECK, ACT). In de ontwikkelingsfase van het programma moet gekeken worden naar het gedrag en de doelgroep waarop het programma zich richt, de leerdoelen en de didactische aanpak. De evaluatie is alleen zinlijk als dit goed is gedaan, omdat dan duidelijk is welke leerdoelen geëvalueerd worden.

De evaluatie maakt deel uit van de CHECK-fase. Het uiteindelijke doel van de meeste verkeerseducatieprogramma's is om het aantal verkeersslachtoffers te reduceren. Omdat het niet mogelijk is educatieprogramma's direct te relateren aan het aantal verkeersslachtoffers, wordt bij de evaluatie meestal gekeken naar het effect op variabelen waarvan bekend is dat ze samenhangen met verkeersveiligheid, zoals risicogedrag. Om een programma te kunnen relateren aan bijvoorbeeld risicogedrag, is het belangrijk een goed onderzoeksdesign voor de evaluatie te kiezen. Vaak zijn er alternatieve verklaringen te vinden voor een effect dat gevonden wordt en het is zaak deze uit te sluiten. Een goed onderzoek bestaat uit een voor- en een nameting in een experimentele groep en een controlegroep. Als de evaluatie optimaal is uitgevoerd en er is geen positief effect of alleen een negatief effect van het programma te zien, dan is het tijd om opnieuw de ontwikkelingsfase van het programma te doorlopen.

Omdat programma's vaak niet voldoende geëvalueerd worden, bevat dit rapport een voorstel voor een laagdrempelige methode die ontwikkelaars en uitvoerders kan helpen en motiveren om hun programma's te evalueren. Het voorstel is om een database met clusters van vragen op te zetten. Deze clusters zouden moeten worden opgebouwd naar onderwerp (bijvoorbeeld alcohol, veilig fietsen) en naar competentie (bijvoorbeeld kennis, attitude).

Summary

The evaluation of traffic education programmes; Recommendations for measuring the effects and a proposal for an abbreviated measuring instrument

A thorough evaluation is important for the quality improvement of traffic education programmes. Presently, this is not done sufficiently. This report contains advice that can help developers and users of traffic education programmes to evaluate their programmes. The report focuses on measuring the effects regarding content, and therefore does not go into the evaluation of procedural aspects of education programmes.

The evaluation of traffic education programmes is part of a larger process of quality assurance: the so-called PDCA cycle (PLAN, DO, CHECK, ACT). In the development phase of the programme, the programme's target group and their behaviour, the educational goals, and the didactic approach must be taken into account. The evaluation is only worthwhile when this is done thoroughly, because it then becomes clear which educational goals are being assessed.

The evaluation is part of the CHECK phase. The reduction of the number of road casualties is the ultimate purpose of most traffic education programmes. As it is impossible to directly relate educational programmes to the numbers of road casualties, the evaluation generally assesses the effect on variables that are known to be related to road safety, for example high-risk behaviour. To be able to relate a programme to, for instance, high-risk behaviour it is important to choose an adequate research design for the evaluation. Often alternative explanations can be given for an observed effect and it is essential to exclude these. A good study consists of a before and an after measurement in an experimental and in a control group. Once the evaluation has been optimally carried out and no positive effect or only a negative effect can be established, the time has come to go through the development phase of the programme once more.

Because programmes are often not evaluated sufficiently, this report proposes an accessible method that can help and motivate developers and users to evaluate their programmes. The proposal is to build a database with clusters of questions. These clusters should be constructed on the basis of subject (e.g. alcohol, safe cycling) and competence (e.g. knowledge, attitude).

Inhoud

1. Inleiding	7
1.1. De PDCA-cyclus	7
1.2. Checklist voor algemene evaluatie	9
1.3. Effectmeting	10
1.4. Leeswijzer	10
2. De ontwikkeling van het programma	11
2.1. Gedrag	11
2.1.1. Bewuste overtredingen	12
2.1.2. Beveiligingsmiddelen	12
2.1.3. Statusonderkenning	12
2.2. Doelgroep	13
2.3. Leerdoelen	13
2.4. Didactische aanpak	13
3. Evaluatiemethoden	15
3.1. Gedragsobservaties	15
3.2. Zelfrapportage	15
4. De evaluatie	17
4.1. Grootte van de steekproef	17
4.2. Het design	18
4.2.1. Voor- en nameting	18
4.2.2. De controlegroep	19
4.2.3. Eisen aan de controlegroep	20
4.2.4. Bedreigingen van de validiteit	21
4.3. Analyse van de resultaten	22
5. Het bijstellen van het programma	24
5.1. De evaluatie laat een positief resultaat zien	24
5.2. De evaluatie laat geen resultaat zien	24
5.3. De evaluatie laat een negatief resultaat zien	24
6. Naar een verkort meetinstrument	25
6.1. De constructie van een itembank	25
6.2. Het construeren van een meetinstrument	26
6.3. Karakterisering en selectie van vragen	27
7. Conclusies	28
7.1. De kwaliteit van educatieve programma's	28
7.2. Het belang van een goede evaluatie	28
7.3. Voorwaarden voor een goede evaluatie	28
7.4. Een verkort meetinstrument	29
Literatuur	30
Bijlage	33
Bedreigingen van de validiteit	

1. Inleiding

In Nederland is een scala aan programma's voor verkeerseducatie voorhanden. Deze richten zich op allerlei onderwerpen (veilig fietsen, alcohol in het verkeer, bromfietzers, openbaar vervoer) en op allerlei doelgroepen (kinderen, jongeren, beginnende automobilisten, ouderen). Het Kennisplatform Verkeer en Vervoer heeft al deze programma's verzameld en gerubriceerd, wat heeft geresulteerd in de Toolkit Permanente Verkeerseducatie (PVE): een overzicht van instrumenten die op een zinvolle wijze voor verkeerseducatie kunnen worden ingezet. Momenteel bevinden zich zo'n tachtig programma's in de toolkit. Het gaat daarbij in principe om alle onderwerpen en alle doelgroepen, hoewel het merendeel van de verkeerseducatieprogramma's zich richt op kinderen en jongeren. Vaak worden deze programma's zeer gewaardeerd door scholen, ouders en de doelgroep zelf. Maar in hoeverre dragen deze programma's nu bij aan de verbetering van de verkeersveiligheid in Nederland? Om die vraag goed te kunnen beantwoorden, zullen deze programma's geëvalueerd moeten worden. Dit rapport helpt ontwikkelaars en uitvoerders van verkeerseducatieprogramma's (zoals Regionale Organen Verkeersveiligheid (ROV's), onderwijsbegeleidingsdiensten en uitgevers) om zo'n evaluatie uit te (laten) voeren. Ontwikkelaars kunnen door een goede evaluatie hun product verbeteren en uitvoerders krijgen hierdoor meer zicht op de effecten van hun activiteiten. Het rapport is in principe niet bedoeld voor personen die de feitelijke uitrol van het programma voor hun rekening nemen, zoals leraren en vrijwilligers, hoewel een groot deel voor de geïnteresseerde leek waarschijnlijk goed te volgen is.

Dit rapport biedt handvatten voor de evaluatie van educatieprogramma's op indicatoren voor verkeersveiligheid. Met een dergelijke evaluatie kan worden bekeken wat het effect van een programma is op onder meer kennis, houding, gedrag, risicoperceptie, zelfreflectie en (zelfgerapporteerde) ongevallen van de deelnemers aan het programma. Zo'n evaluatie op effectiviteit maakt deel uit van een groter proces om de kwaliteit van educatieprogramma's te verbeteren. Dit inleidende hoofdstuk bespreekt het proces van kwaliteitszorg waarvan evaluatie op effectiviteit onderdeel uitmaakt. Ook gaat het in op de manier om de algemene kwaliteit van verkeerseducatieprogramma's te meten (*Paragraaf 1.2*) en op de specifieke effectmeting die in dit rapport centraal staat (*Paragraaf 1.3*).

1.1. De PDCA-cyclus

Hoewel kwaliteit sterk afhankelijk is van de inhoud van een product, zou de kwaliteit ook in het *proces* verankerd moeten zijn. Een bekende methode voor structureel kwaliteitsmanagement is de PDCA-cyclus (PLAN, DO, CHECK, ACT). Deze paragraaf bespreekt de verschillende stappen in deze cyclus zoals die worden toegepast in verkeerseducatieprogramma's.

PLAN-fase: de ontwikkeling van het programma

In de PLAN-fase wordt de inhoud van het educatieprogramma ontwikkeld. Het is belangrijk daarbij de vraag te stellen op welk gedrag het programma zich richt en of dit gedrag een relatie heeft met verkeersveiligheid. Ook moeten in deze fase leerdoelen worden geformuleerd. Binnen de Toolkit

PVE is een groot aantal educatieve programma's al met dit oog bekeken en is een onderverdeling gemaakt in niveaus van gedragingen. Dit kan een hulpmiddel zijn bij het ontwikkelen van een programma of bij het toetsen van een voorgenomen idee.

Vervolgens is het van belang om te bekijken welke processen ten grondslag liggen aan het gedrag dat door het programma moet worden beïnvloed. Zo kan het zijn dat kinderen geen verlichting op de fiets voeren omdat zij het gevaar van rijden zonder licht niet inzien (risicoperceptie), maar het is bijvoorbeeld ook mogelijk dat status of groepsdruk een rol speelt. In dat laatste geval heeft het minder zin om in een programma de risico's zichtbaar te maken, want die kennen de kinderen al. Het is dan beter om kinderen weerbaar te maken tegen groepsdruk. In de PLAN-fase moet ook een didactische aanpak worden gekozen. Op welke manier probeert het programma het gedrag te beïnvloeden? Hierbij kunnen verschillende principes uit de onderwijskunde gebruikt worden. In de educatieve 'weetlat' van Veilig Verkeer Nederland (VVN) zijn deze principes duidelijk en toegankelijk samengevat.

DO-fase: de uitvoering van het programma

In de DO-fase wordt het programma daadwerkelijk uitgevoerd. In deze fase moet duidelijk worden of het programma in de praktijk toepasbaar is. Er is in deze fase echter nog geen sprake van formele evaluatie. Bij de implementatie van het programma moet worden nagedacht over de vraag welke personen en organisaties allemaal een rol hebben bij de uitvoering. De school is meestal een voor de hand liggende organisatie, maar vaak zijn er ook andere organisaties bij de uitvoering betrokken (zoals de politie). Ook ouders kunnen een rol spelen bij de uitvoering; de ouders kunnen bijvoorbeeld betrokken worden via opdrachten of discussie. Binnen een organisatie kunnen verschillende personen betrokken zijn, in het geval van de school bijvoorbeeld leerkrachten en verkeersouders.

CHECK-fase: evaluatie van proces en inhoud

In de CHECK-fase wordt de effectiviteit van het programma in relatie tot de leerdoelen onderzocht: de evaluatie die het onderwerp is van dit rapport. Daarnaast wordt bekeken of het programma uitgevoerd wordt zoals beoogd, en wordt naar het proces en de waardering gekeken. Dit gebeurt meestal in de vorm van een procesevaluatie waarin op een meer kwalitatieve wijze wordt gevraagd naar de mening van deelnemers, uitvoerders en andere betrokkenen.

ACT-fase: bijstelling van het programma

De ACT-fase wordt vaak vergeten bij de ontwikkeling en de evaluatie van educatieve programma's. Toch is juist deze fase een essentieel onderdeel van goede kwaliteitszorg. In de ACT-fase wordt bekeken wat kan worden geleerd uit de CHECK-fase. Is er aanleiding om het programma te herzien? Als dat het geval is, komt men weer terug in de PLAN-fase en kan worden bekeken welke verbeteringen in de aanpak mogelijk zijn. Een voorbeeld is het dodehoekprogramma van het Regionaal Orgaan verkeersveiligheid Fryslân (ROF), dat naar aanleiding van het Effecten van Verkeerseducatie Onderzoek (EVEO) is herzien.

Door zorgvuldig alle stappen in het kwaliteitsproces te doorlopen, kunnen zwaktes in een educatieprogramma duidelijk worden en kunnen hiervoor

oplossingen worden gevonden. Zo verbetert de kwaliteit van het programma. Evaluatie speelt daarin een belangrijke rol. Omdat kwaliteitszorg een cyclisch proces is, zou een programma idealiter telkens opnieuw de PDCA-cyclus moeten doorlopen.

Een voorbeeld: stel dat er een educatief programma is dat, gebaseerd op verkeersveiligheidsonderzoek, niet de juiste benadering kiest (bijvoorbeeld een slipcursus die het gevoel van controle benadrukt). Dan zou dit programma op basis van het oordeel van verkeersveiligheidsexperts als ongeschikt beoordeeld worden. Als de ontwikkelaar het programma echter kan aanpassen op de onderdelen die dat behoeven, doordat hij zeer zorgvuldig een proces- en effectevaluatie heeft uitgevoerd en ook alle andere stappen heeft doorlopen, dan kan uit dit proces toch een goed product ontstaan. Zo is het in het voorbeeld van de slipcursus denkbaar dat deze wel geschikt zou zijn indien de cursus ervoor zorgt dat deelnemers zich realiseren hoe weinig controle ze eigenlijk hebben. Het educatieve element is dan niet 'uit een slip geraken', maar voorkomen dat men in een slip raakt, bijvoorbeeld door niet de weg op te gaan als het glad is.

Op deze manier zorg je ervoor dat ook ideeën waarvan mensen misschien in het begin zeggen dat ze weinig kans van slagen hebben, toch de mogelijkheid hebben om door het doorlopen van het proces de kwaliteit te bewijzen. Een optimale kwaliteit bereik je dus niet door het enkele oordeel van een of meerdere experts, maar door het inzichtelijk maken van de stappen die zijn genomen om de kwaliteit te borgen.

1.2. Checklist voor algemene evaluatie

De kwaliteit van verkeerseducatieprogramma's kan in de CHECK-fase van de PDCA-cyclus op verschillende manieren worden onderzocht. Dit rapport beschrijft een manier om het effect dat programma's hebben te evalueren, maar daarnaast kan een programma bijvoorbeeld ook worden geëvalueerd op proces, deskundigheid van betrokkenen en didactische kwaliteit. Vissers (2010) beschrijft een checklist die ingaat op alle aspecten van educatieprogramma's en dus niet alleen op het effect. Deze checklist zou kunnen worden doorlopen om de algemene kwaliteit van een educatief programma vast te stellen. De volgende onderwerpen komen in de checklist aan de orde:

- het te beïnvloeden gedrag;
- de doelgroep;
- concrete, meetbare leerdoelen;
- didactische uitgangspunten en passende technieken en strategieën;
- inhoud en vormgeving van de materialen;
- toetsing en evaluatie binnen het programma;
- organisatie van het programma;
- implementatie van het programma;
- procesevaluatie: inventarisatie van gebruikerservaringen;
- effectmeting: monitoring en evaluatie van de effecten van het programma op geaggregeerd niveau.

Voor elk van deze onderwerpen is een aantal vragen geformuleerd. Ontwikkelaars en uitvoerders van verkeerseducatieprogramma's zouden deze checklist kunnen gebruiken om vast te stellen of ze met hun

programma op de goede weg zijn. Ook zou een subsidieverstrekker de checklist kunnen gebruiken om een afweging te maken tussen verschillende educatieve programma's, zodat de subsidie uiteindelijk terecht komt bij de programma's die het effectiefst zijn. Dit is ook een reden om programma's goed te evalueren. Tot slot is het mogelijk om een onafhankelijk keurmerk in te stellen, waarbij experts betrokken zijn die een oordeel geven over de score van een programma op de checklist.

Zie voor meer informatie over deze checklist Vissers (2010).

1.3. Effectmeting

Effectmeting van de inhoud van verkeerseducatieprogramma's is een belangrijk element in het proces van kwaliteitszorg. De meeste projecten voor verkeerseducatie hebben uiteindelijk tot doel de verkeersveiligheid te verbeteren. Dat betekent dat elk project zou moeten leiden tot een afname van het aantal verkeersongevallen. Het is belangrijk om te kunnen vaststellen in hoeverre een educatieproject in staat is om aan dit doel een bijdrage te leveren. Vaak is het echter niet mogelijk om een direct effect op het aantal verkeersongevallen vast te stellen (Twisk, Vlakveld & Commandeur, 2007). Dat komt doordat ongevallen (gelukkig) niet vaak voorkomen en de ongevallen die plaatsvinden, het resultaat zijn van een complexe verzameling van factoren. Wel kan worden onderzocht in hoeverre een educatieprogramma invloed heeft op factoren waarvan we uit ander onderzoek weten dat die een relatie hebben met verkeersveiligheid (zie bijvoorbeeld Dragutinovic & Twisk, 2006). Zo weten we bijvoorbeeld dat het dragen van gordels de kans op ernstig letsel vermindert (SWOV, 2010). Wanneer een educatieprogramma aantoonbaar effectief is in het verhogen van de gordeldracht, dan kan aangenomen worden dat dit programma de verkeersveiligheid ten goede komt. Op deze manier kan dus op indirecte wijze worden bepaald of een project een bijdrage levert aan het verbeteren van de verkeersveiligheid.

1.4. Leeswijzer

In dit rapport staat de vraag centraal op welke manier het inhoudelijke effect van een educatief programma kan worden vastgesteld. Hoewel het zwaartepunt in dit rapport bij evaluatie achteraf ligt, bevat *Hoofdstuk 2* ook enkele overwegingen over de ontwikkeling van het programma. In *Hoofdstuk 3* wordt stilgestaan bij de keuze van een evaluatiemethode. *Hoofdstuk 4* gaat in op het feitelijk uitvoeren van de evaluatie, de analyse van de resultaten en de problemen waarmee men in deze fase te maken kan krijgen. *Hoofdstuk 5* bespreekt de eventuele bijstelling van het programma op basis van de resultaten van de evaluatie. *Hoofdstuk 6* beschrijft ten slotte een voorstel voor een verkort meetinstrument voor evaluatie. Hiermee moet het in de toekomst gemakkelijker worden om een evaluatie uit te voeren. *Hoofdstuk 7* bevat een samenvatting en conclusies.

2. De ontwikkeling van het programma

Dit hoofdstuk gaat in op de ontwikkeling van een verkeerseducatieprogramma, de PLAN-fase uit de PDCA-cyclus. In de meeste gevallen zal het programma echter al ontwikkeld zijn voordat een organisatie aan de slag gaat met de evaluatie. Indien de evaluatie een al volledig uitontwikkeld programma betreft, kan deze stap dus worden overgeslagen. Dit hoofdstuk is alleen relevant voor programma's die nog ontwikkeld moeten worden, en voor programma's waarvan na evaluatie blijkt dat ze bijgesteld moeten worden (de ACT-fase). Dit hoofdstuk gaat achtereenvolgens in op het gedrag waarop het programma zich zou moeten richten, de doelgroep, de leerdoelen en de didactische aanpak.

2.1. Gedrag

Verkeerseducatieprogramma's hebben meestal¹ tot doel het gedrag van mensen in het verkeer te veranderen. Bepaalde gedragingen kunnen immers riskant zijn en de kans op een verkeersongeval vergroten. Maar welke gedragingen zijn dat? En op welk gedrag zouden educatieprogramma's zich moeten richten om een maximaal effect te behalen? De samenhang tussen gedragingen en verkeersveiligheid is inmiddels voor een groot aantal gedragingen wetenschappelijk onderzocht. Daarnaast zijn er aspecten waarvan we aannemen dat ze samenhangen met verkeersveiligheid, maar dit niet zeker weten. Kennis van verkeersregels is zo'n aspect. Er is geen onderzoek bekend dat kennis over verkeersregels rechtstreeks in verband brengt met verkeersonveiligheid. We kunnen dus formeel niet bewijzen dat betere kennis van de verkeersregels tot een lagere kans op een ongeval leidt. Toch moeten we hieruit niet concluderen dat kennis van de verkeersregels niet belangrijk is. We kunnen het belang van die kennis namelijk theoretisch afleiden. Neem als voorbeeld een aanrijding tussen een fietser en een auto, waarbij de fietser onterecht geen voorrang verleent aan de auto. Dat de fietser geen voorrang verleent, kan komen doordat hij niet weet dat hij voorrang had moeten verlenen. Het kan ook zijn dat de fietser de voorrangsregels wel kent, maar bijvoorbeeld de auto niet ziet of bewust geen voorrang verleent omdat hij denkt dat hij er nog wel voorlangs kan. Alleen die gevallen waarbij een fietser wel voorrang kan en wil verlenen maar niet weet dat hij dit zou moeten doen, kunnen door kennis van de verkeersregels worden voorkomen. Kortom, kennis van verkeersveiligheid is een belangrijke voorwaarde voor veilig gedrag (als verkeersdeelnemers de regels niet kennen, kunnen zij ze ook niet toepassen), maar zeker geen garantie voor veilig gedrag. Overigens kan het ook voor gevaarlijke situaties zorgen wanneer men té zeer uitgaat van het eigen gelijk omdat men bijvoorbeeld de voorrangsregels goed kent en daaraan vasthoudt; men is dan minder flexibel.

Bij het ontwikkelen van een verkeerseducatieprogramma is het belangrijk dat duidelijk is welk gedrag het programma moet beïnvloeden en wat de relatie is van dat gedrag met verkeersveiligheid. Het vervolg van deze paragraaf gaat in op enkele voorbeelden van gedragingen waarvan de

¹ Soms is het doel vooral procesmatig van aard en richt het programma zich bijvoorbeeld op agendasetting: verkeersveiligheid hoger op de (beleids)agenda krijgen.

relatie met verkeersonveiligheid is aangetoond. Er worden vooral gedragingen genoemd die door educatie kunnen worden beïnvloed.

2.1.1. *Bewuste overtredingen*

Een eerste categorie gedragingen die door educatie kunnen worden beïnvloed, zijn bewuste verkeersovertredingen. In onderzoek naar verkeersgedrag wordt onderscheid gemaakt tussen fouten en overtredingen. Bij fouten gaat het om onbewust verkeerd of afwijkend gedrag en verkeerd of afwijkend gedrag als gevolg van een inschattingfout. Bij overtredingen gaat het om bewust verkeerd of afwijkend gedrag. Wetenschappelijke onderzoeken naar verkeersgedrag van automobilisten tonen aan dat mensen die veel bewuste overtredingen rapporteren, niet per se ook de mensen zijn die veel fouten rapporteren (zie bijvoorbeeld Mesken, Lajunen & Summala, 2002; Parker et al., 1995; Reason et al., 1990). Fouten en overtredingen zijn dus twee afzonderlijke categorieën die een andere relatie met ongevalsbetrokkenheid hebben. Het blijkt dat vooral bewuste overtredingen ongevalsbetrokkenheid kunnen verklaren.

Bij verkeersgedrag van kinderen is dit onderscheid niet gevonden (Twisk, Vlakveld & Commandeur, 2006). Uit onderzoek blijkt dat kinderen die vaak kleine vergissingen of inschattingfouten maken, ook vaak bewuste overtredingen rapporteren. Al deze gevaarlijke gedragingen hangen dus samen en hebben een relatie met (zelfgerapporteerde) ongevalsbetrokkenheid.

Er zijn drie typen bewuste overtredingen waarvan de relatie met verkeersonveiligheid duidelijk is aangetoond: te hard rijden door automobilisten (Aarts & Van Schagen, 2006), rijden onder invloed van alcohol (Blomberg et al., 2005) en bellen op de fiets of in de auto (Goldenbeld, Houtenbos & Ehlers, 2010; Caird et al., 2008). Bij dit laatste gedrag maakt het nauwelijks uit of mensen handsfree of handheld bellen.

2.1.2. *Beveiligingsmiddelen*

Educatieprogramma's kunnen zich ook richten op het gebruik van beveiligingsmiddelen zoals helmen en autogordels. Het gebruik van een beveiligingsmiddel kan ongevallen niet voorkomen, maar de afloop van een ongeval kan hierdoor wel minder ernstig zijn. Het veiligheidseffect van het dragen van een helm is in diverse studies aangetoond, zowel voor fietsers (Thompson et al., 2004) als voor bromfietzers (Huijbers & Van Kampen, 1985) als voor motorrijders (Liu et al., 2007). Ook het dragen van autogordels reduceert de kans op ernstig letsel aantoonbaar (Evans, 1996).

2.1.3. *Statusonderkenning*

Tot slot kan verkeerseducatie zich richten op statusonderkenning. De laatste tijd is er steeds meer bewijs dat veilige verkeersdeelname valt of staat met een realistische inschatting van zowel de eigen vaardigheden als de complexiteit van de verkeerssituatie. Vooral de inschatting van de eigen vaardigheden laat nogal eens te wensen over, hoewel het uitmaakt of de zelf ingeschatte vaardigheden worden vergeleken met een beoordeling door een instructeur of met een gemiddelde andere automobilist (zie bijvoorbeeld Mynttinen et al., 2009a; Mynttinen et al., 2009b; Tronsmoen, 2008).

Educatieprogramma's zouden dus tot doel kunnen hebben om weggebruikers zich beter bewust te maken van hun eigen feilbaarheid. Dit soort programma's lijken effectiever te zijn dan bijvoorbeeld vaardigheidstrainingen, waarvan de effecten nooit sluitend zijn aangetoond (zie Clarke, Ward & Truman, 2005).

2.2. Doelgroep

Hoewel een verkeerseducatieprogramma nuttig kan zijn voor alle verkeersdeelnemers, zijn er groepen waarvan we weten dat deze in het bijzonder baat hebben bij educatie. Dit geldt vooral voor weggebruikers die voor het eerst in een bepaalde rol deelnemen aan het verkeer. Het is bekend dat jonge, beginnende automobilisten een grotere kans hebben om bij een ongeval betrokken te raken. Dit verhoogde risico geldt ook voor jonge voetgangers die voor het eerst alleen naar school lopen, beginnende fietsers en beginnende bromfietzers. Hier moet bij de ontwikkeling van verkeerseducatieprogramma's rekening mee gehouden worden.

Daarnaast is het van belang om na te gaan of de doelgroep ook het gedrag vertoont waarop het programma is gericht. Een bromfietsprogramma op een school waar nauwelijks bromfiets wordt gereden, zal weinig effectief zijn.

2.3. Leerdoelen

Leerdoelen geven weer wat een deelnemer na het volgen van een programma moet weten en/of kunnen. Alleen als er leerdoelen voor een programma zijn geformuleerd, kan achteraf in een evaluatie worden vastgesteld of deze doelen gehaald zijn. Vaak is het zo dat leerdoelen impliciet blijven tijdens het ontwikkelen van een programma of dat leerdoelen in te algemene bewoordingen worden geformuleerd. In samenspraak met de regio's zijn door Vissers et al. (2005) leerdoelen voor verschillende leeftijdsgroepen ontwikkeld. Bij het ontwikkelen van een nieuw programma kan Vissers et al. (2005) als naslagwerk dienen en kan bekeken worden of het programma aansluit bij de leerdoelen voor de betreffende leeftijdsgroep.

2.4. Didactische aanpak

De keuze voor een type didactische aanpak is voor een groot deel afhankelijk van de persoonlijke voorkeur van de docent. Voor verkeerseducatie is nog geen aparte vakdidactiek ontwikkeld. Wel is bewezen dat bepaalde benaderingswijzen contraproductief zijn. Zo blijkt bijvoorbeeld (ook uit andere toepassingsgebieden dan het verkeer) dat confronterende boodschappen zonder gedragsalternatieven niet goed werken (Ruiter, Abraham & Kok, 2001).

Vissers (2010) noemt een aantal uitgangspunten voor de manier waarop verkeerseducatie in de praktijk zou moeten worden vormgegeven:

- Integreer theorie en praktijk.
- Leg de nadruk op zelflerend vermogen.
- Leer kennis, inzicht en begrip in een betekenisvolle context aan.
- Maak gebruik van informeel leren.
- Besteed aandacht aan alle niveaus van onveilig handelen.

- Besteed aandacht aan sociale, communicatieve en morele aspecten van verkeersdeelname.

Er zijn dus diverse aanknopingspunten om tot een goede didactische aanpak te komen. Ongeacht de gekozen aanpak is het belangrijk om in het kwaliteitsproces de didactische aanpak te expliciteren en deze consistent in het hele programma door te voeren.

3. Evaluatiemethoden

Als een verkeerseducatieprogramma is ontwikkeld (PLAN-fase) en vervolgens uitgevoerd (DO-fase), dient de effectiviteit van het programma te worden onderzocht (CHECK-fase; zie ook *Paragraaf 1.1*). Het is in de evaluatie vooral belangrijk om te kijken naar de variabelen waarop een effect wordt verwacht. Een verkeerseducatieproject heeft in het algemeen tot doel de verkeersveiligheid positief te beïnvloeden. In het ideale geval zou bij de evaluatie van een verkeerseducatieproject dan ook gekeken moeten worden naar het effect daarvan op het aantal verkeersongevallen. Omdat onderzoek naar ongevallen te ingewikkeld is (zie ook *Paragraaf 1.3*), ligt het voor de hand om te kijken naar andere indicatoren van verkeersonveiligheid, zoals gedrag en voorspellers van gedrag. Dit hoofdstuk gaat in op deze twee evaluatiemethoden.

3.1. Gedragsobservaties

Het effect van een educatieprogramma op feitelijk gedrag kan worden vastgesteld door het betreffende gedrag te observeren. Een (fictief) voorbeeld is de evaluatie van een programma dat het dragen van de helm door bromfietzers moet bevorderen. De effectiviteit van dit programma kan worden vastgesteld door het aantal helmdragers onder bromfietzers op scholen te tellen, zowel op scholen die het programma wel en scholen die het niet aanbieden, en zowel voor als nadat het programma is uitgevoerd (zie *Hoofdstuk 4* voor onderzoeksdesigns). Dit is de meest directe manier van meten, omdat men het te beïnvloeden gedrag hierbij daadwerkelijk observeert.

Vaak zijn zulke directe gedragsobservaties echter niet mogelijk. In het Effecten van Verkeerseducatie Onderzoek (EVEO) is bijvoorbeeld getracht om het effect van een dodehoekprogramma voor kinderen te onderzoeken. Het doel van dit educatieprogramma was te voorkomen dat jonge fietsers zich in de dode hoek van vrachtauto's bevinden. Het bevatte simpele regels, onder andere over de plaats waar fietsers zich bij verkeerslichten moeten opstellen wanneer er een vrachtauto aanwezig is. Idealiter zou men in dit geval willen onderzoeken of de deelnemers aan het programma werkelijk veiliger gedrag vertonen. Hiervoor zou men eigenlijk grote groepen kinderen in het verkeer moeten volgen en moeten observeren hoe de kinderen omgaan met deze situatie als die zich voordoet. Het moge duidelijk zijn dat dit vanwege de tijdsinvestering en praktische bijkomstigheden bijna onmogelijk is. In dit geval kan men uitwijken naar andere maten. In het EVEO-project is gekozen voor een maquetteopstelling: een gesimplificeerde versie van de werkelijkheid waarbinnen verkeersopgaven moeten worden opgelost (zie ook Twisk, Vlakveld & Commandeur, 2007).

3.2. Zelfrapportage

Wanneer gedragsobservaties niet mogelijk zijn, kan men ervoor kiezen om vragenlijsten te gebruiken waarin gevraagd wordt naar gedrag of naar indicatoren van gedrag. In het geval van het programma dat het dragen van de bromfietshelm moet bevorderen, kan men bijvoorbeeld de vraag stellen:

“Tijdens hoeveel van de afgelopen tien ritten op de bromfiets heb je je helm gedragen?”

Ook wanneer men het effect van een programma op psychologische variabelen wil onderzoeken, is men aangewezen op zelfrapportage. Een grote voorlichtingscampagne gericht op rijden onder invloed kan bijvoorbeeld (mede) tot doel hebben de houding ten opzichte van rijden onder invloed te veranderen. Deze houding, of attitude, is niet aan de buitenkant waarneembaar. Men zal deze attitude dus moeten afleiden door een aantal vragen te stellen.

Bij het uitvoeren van vragenlijstonderzoek moet met verschillende zaken rekening gehouden worden. Ten eerste is het belangrijk om na te gaan of de vragen geschikt zijn voor de doelgroep. Lange, complexe zinnen of te talige vragenlijsten zijn over het algemeen minder geschikt voor jonge kinderen. Het is dan beter om bijvoorbeeld gebruik te maken van beeldmateriaal in de vorm van foto's of video's. Zo is voor de evaluatie van een dodehoek-programma in Gelderland gebruikgemaakt van videobeelden, waarbij deelnemers op een stopknop moesten drukken als zij vonden dat het te gevaarlijk werd. Ook dit is een vorm van zelfrapportage, alleen wordt niet gemeten aan de hand van een vraag maar aan de hand van directe reacties.

Een ander lastig punt is de balans tussen enerzijds betrouwbaar meten en anderzijds de lengte van de vragenlijst. Vanuit onderzoeksoogpunt moet een bepaalde gedraging op verschillende manieren in de vragen aan bod komen om betrouwbare uitspraken te kunnen doen. Als men bijvoorbeeld de houding van bromfietzers ten opzichte van helmgebruik wil meten, dan kan men de stelling “Het is belangrijk om een helm te dragen” voorleggen, maar beter is om een combinatie van vragen te stellen die allemaal ongeveer hetzelfde meten, maar iets anders gesteld zijn. Bijvoorbeeld: “Het is belangrijk om een helm te dragen”; “Als het niet druk is, hoef ik mijn helm niet te dragen”; “Als ik mijn helm draag voel ik me veilig”. Op die manier is de uitkomst betrouwbaarder dan wanneer slechts één vraag gesteld wordt. Hierdoor kan de vragenlijst echter behoorlijk lang worden, zeker wanneer meerdere attitudes gemeten moeten worden. Een educatieprogramma voor bromfietzers kan zich bijvoorbeeld richten op helmdracht, opvoeren en risicogedrag. Daarnaast zal men niet alleen iets willen weten over de attitude, maar misschien ook over risicoperceptie, het (zelfgerapporteerde) gedrag, gevoeligheid voor groepsdruk enzovoort. Als men meerdere vragen over elk onderdeel wil stellen om betrouwbaar te meten, kan de vragenlijst onhandelbaar lang worden. Het kan dan gebeuren dat de doelgroep niet meer gemotiveerd is om de vragenlijst in te vullen. *Hoofdstuk 6* biedt een oplossing voor dit probleem.

4. De evaluatie

Wanneer een meetmethode is gekozen, kan het evaluatieonderzoek daadwerkelijk worden uitgevoerd. Dit hoofdstuk gaat achtereenvolgens in op de grootte van de steekproef, het belang van het design (voor- en nameting, controlegroep, random toewijzing) en mogelijke problemen met het leggen van oorzaak-gevolgrelaties.

4.1. Grootte van de steekproef

Met de evaluatie wil men over het algemeen niet alleen een uitspraak doen over het effect van het programma op één groep of één school. Vaak gaat het om het effect van het programma op de gehele doelgroep, als deze hele doelgroep het programma zou volgen. Neem als voorbeeld het al eerder aangehaalde programma voor bromfietzers. Natuurlijk is het nuttig om te weten of op school x het percentage bromfietzers dat de helm draagt is toegenomen als gevolg van het programma. Maar óók zal men willen weten of het programma hetzelfde effect zal hebben als het ook op andere scholen wordt uitgezet. Met andere woorden: men zal willen generaliseren naar de doelgroep en daarmee willen vaststellen of het effect dat in de steekproef is gevonden ook van toepassing is op de totale populatie.

Er zijn verschillende tabellen en formules beschikbaar om de benodigde steekproefgrootte uit te rekenen. Deze formules houden rekening met de omvang van de populatie, met het vereiste betrouwbaarheidsniveau en met de vraag of er vooraf al iets gezegd kan worden over de mate van variabiliteit in het onderzochte populatiekenmerk. Een simpele formule (Cochran, 1963) is:

$$N = \frac{Z^2 p(1-p)}{E^2}$$

Z is een constante en is afhankelijk van het gewenste betrouwbaarheidsniveau (bij een gewenst betrouwbaarheidsniveau van 95% is Z 1,96); p is de mate van variabiliteit van het kenmerk in de populatie (0,5 indien onbekend) en E is het betrouwbaarheidsinterval (de maximale spreiding van de resultaten die je nog wilt accepteren).

Een voorbeeld: als we het effect van het bromfietsprogramma vaststellen, willen we voor 95% zeker zijn dat het gevonden effect niet op toeval berust. Daarnaast willen we een maximale spreiding in de resultaten van 3%. We weten niet hoe de verdeling in de populatie bromfietzers is tussen personen die wel en niet de helm dragen en dus gaan we uit van maximale variabiliteit: 50%. De benodigde steekproefgrootte is dan:

$$N = \frac{1,96^2 \cdot 0,5(1-0,5)}{0,03^2} = 1.067$$

Als we iets meer spreiding accepteren in de resultaten, neemt de vereiste steekproefgrootte snel af. Bij een foutmarge van 5% in plaats van 3% is nog maar een omvang van 385 nodig.

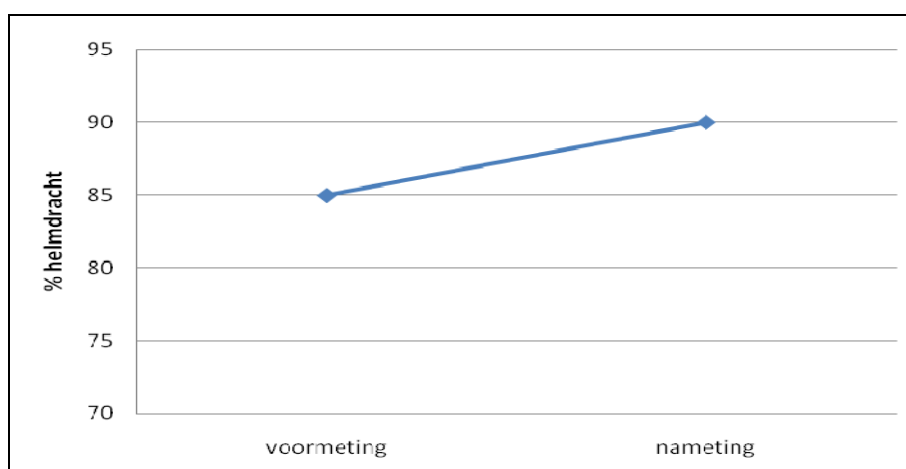
De besproken formule betreft de steekproefgrootte die nodig is om te kunnen generaliseren. Hiermee wordt de kans vergroot dat een effect dat in de steekproef wordt gevonden, ook in de populatie bestaat. Het is ook mogelijk dat een effect in de populatie wel bestaat, maar in de steekproef niet wordt gevonden. Om vast te stellen hoe groot de steekproef moet zijn om een effect dat in de populatie bestaat, ook in de steekproef te vinden, zou men een zogenaamde power-analyse moeten uitvoeren. Omdat hiervoor ook kennis over de te gebruiken statistische toetsen nodig is, blijft deze discussie hier buiten beschouwing.

4.2. Het design

Het ideale onderzoeksdesign voor een effectmeting is een design met een voor- en nameting en een controlegroep. Hieronder wordt toegelicht waarom dat zo is.

4.2.1. Voor- en nameting

Het uitvoeren van een voor- en nameting ligt wellicht voor de hand: dat 90% van de bromfietzers de helm altijd draagt, zegt weinig als men niet weet hoe hoog dit percentage was voorafgaand aan het programma. Vaak wordt dus ook een voormeting uitgevoerd (zie *Afbeelding 4.1*).



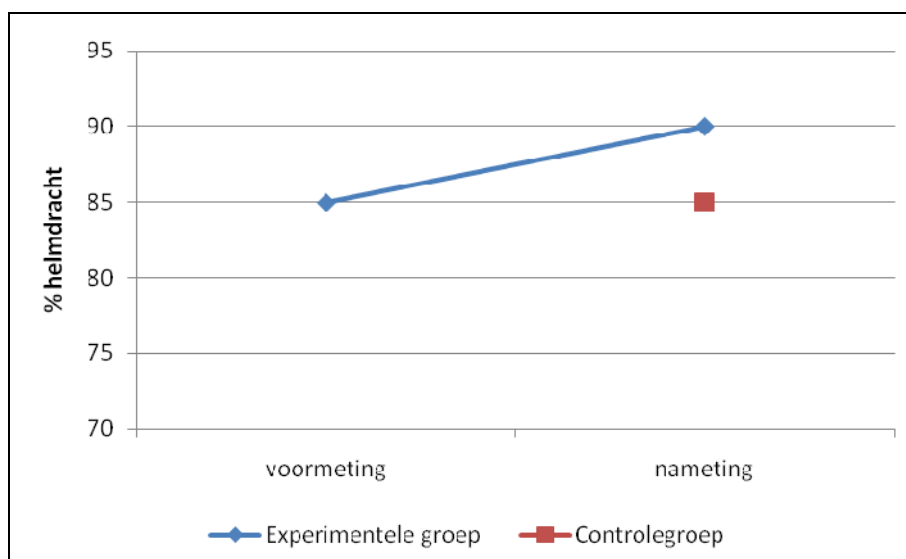
Afbeelding 4.1. *Illustratie van een mogelijk resultaat bij een onderzoek naar helmdracht met een voor- en nameting.*

Afbeelding 4.1 laat zien dat het percentage bromfietzers dat de helm draagt is toegenomen. Uit dergelijke resultaten wordt vaak geconcludeerd dat het programma succesvol was. Er is immers een positief effect zichtbaar. Echter, dit positieve effect kan niet zonder meer worden toegeschreven aan het programma. Het kan namelijk zo zijn dat er in dezelfde periode andere gebeurtenissen hebben plaatsgevonden die het verschil in voor- en nameting kunnen verklaren. Denk bijvoorbeeld aan een landelijke campagne over de veiligheidseffecten van het dragen van een bromfietshelm of een

intensivering van de politiecontroles op het dragen van een helm. Om deze alternatieve verklaringen uit te sluiten, is een controlegroep nodig.

4.2.2. De controlegroep

Een controlegroep bestaat uit deelnemers die in alle relevante opzichten lijkt op de onderzoeksgroep (hierna 'experimentele groep' genoemd). Het enige verschil is dat deze groep het educatieve programma niet heeft gevolgd. In *Afbeelding 4.2* wordt bij wijze van illustratie weergegeven wat de mogelijke resultaten zijn wanneer de nameting wordt uitgevoerd bij zowel de experimentele groep als bij de controlegroep.

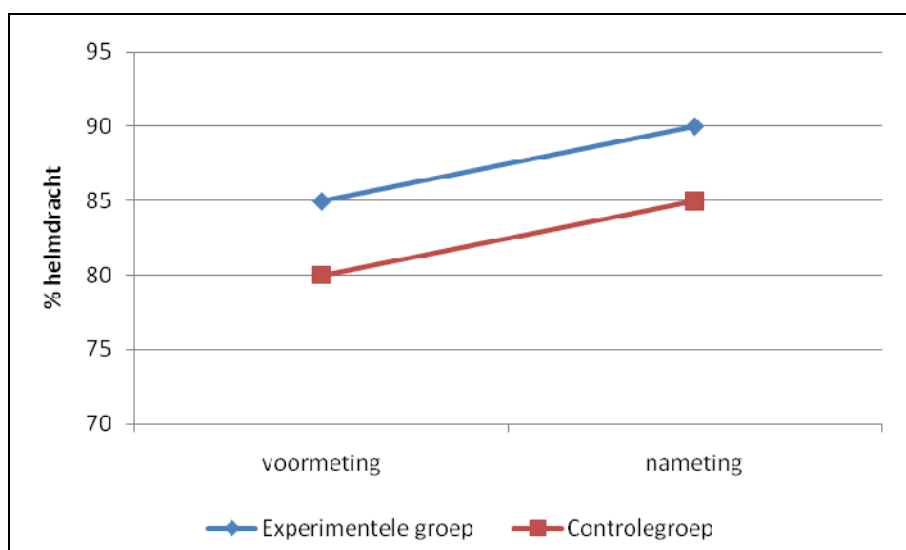


Afbeelding 4.2. Illustratie van een mogelijk resultaat bij een onderzoek naar helmdracht met een voor- en nameting en een controlegroep bij de nameting.

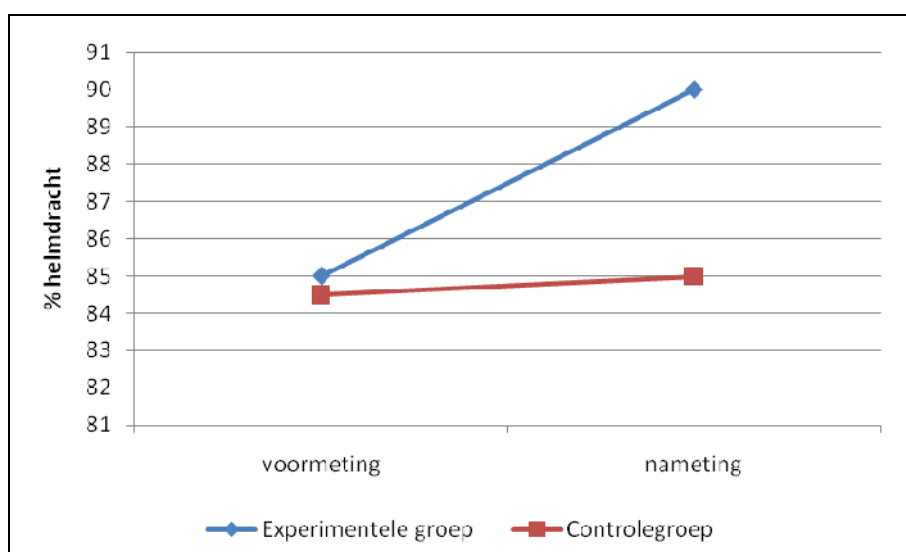
Nu blijkt dat er niet alleen een verschil is tussen voormeting en nameting, maar ook dat de resultaten van de experimentele groep en de controlegroep verschillen. Kunnen we nu dan met zekerheid stellen dat het programma effectief is? Nee, dat kan niet, omdat het best mogelijk is dat ook in de controlegroep het percentage helmdragers is toegenomen. Dat kan vanwege eerder genoemde alternatieve verklaringen, bijvoorbeeld blootstelling aan een massamediale campagne of geïntensiverde handhaving. Het kan ook zo zijn dat de controlegroep toch onbedoeld is blootgesteld aan het programma, doordat deze groep bijvoorbeeld op dezelfde school is geworven en via via toch iets heeft meegekregen van het programma.

Afbeelding 4.3 en Afbeelding 4.4 laten zien dat alleen het uitvoeren van een voor- en nameting onder zowel een experimentele groep als een controlegroep met zekerheid iets kan zeggen over de effectiviteit van het programma. In de eerste afbeelding is te zien dat het programma geen effect had: de controlegroep is in dezelfde mate verbeterd als de experimentele groep. Het verschil in de nameting tussen de twee groepen is volledig toe te schrijven aan aanvangsverschillen in de voormeting. In *Afbeelding 4.4* wordt een situatie geïllustreerd waarbij wel met een grote

mate van zekerheid te stellen is dat de verbetering is toe te schrijven aan het volgen van het programma.



Afbeelding 4.3. Illustratie van een mogelijk resultaat bij een onderzoek naar helmdracht met een voor- en nameting en een controlegroep bij de voor- en nameting: geen effect.



Afbeelding 4.4. Illustratie van een mogelijk resultaat bij een onderzoek naar helmdracht met een voor- en nameting en een controlegroep bij de voor- en nameting: wel een effect.

4.2.3. Eisen aan de controlegroep

Idealiter vinden voor- en nameting onder experimentele groep en controlegroep op hetzelfde moment plaats, omdat dan de kans op andere toevallige verschillen zo klein mogelijk is. De controlegroep moet verder op zo veel mogelijk relevante kenmerken gelijk zijn aan de testgroep, bijvoorbeeld op kenmerken als leeftijd en schooltype. Ook zou er ongeveer evenveel tijd tussen de twee metingen moeten zitten als bij de experimentele groep.

Verder moeten deelnemers idealiter *at random* worden toegewezen aan één van de twee groepen (de experimentele of de controlegroep). 'At random' betekent 'puur op grond van toeval': elke persoon moet evenveel kans hebben om in de ene of in de andere groep terecht te komen. Als er geen random toewijzing plaatsvindt, kan het zo zijn dat de controlegroep op relevante aspecten afwijkt van de experimentele groep. Bij de evaluatie van verkeerseducatieprojecten is random toewijzing echter zelden mogelijk. De meeste programma's worden uitgevoerd op scholen en die scholen worden niet at random toegewezen aan het programma, maar melden zich veelal zelf aan. Leerlingen hebben daardoor niet meer evenveel kans om in de experimentele groep terecht te komen; dit is afhankelijk van de school.

Wanneer individuele random toewijzing niet mogelijk is, zou men er in ieder geval naar moeten streven om de klassen of de scholen die meedoen aan de evaluatie, random toe te wijzen. Dit kan gebeuren door eerst een aantal scholen te werven en dan door toeval te bepalen welke school het programma krijgt en welke school niet. Ook random toewijzing op dit niveau gebeurt vaak echter niet. De scholen die geïnteresseerd zijn in een educatieprogramma zijn vaak scholen die tóch al geïnteresseerd zijn in verkeersveiligheid. Zij bieden zich aan voor het programma en komen tijdens het evaluatieonderzoek automatisch in de experimentele groep terecht. Als er nu een controlegroep wordt gezocht, zal deze heel vaak bij aanvang al minder goed scoren op allerlei relevante aspecten dan de school of scholen die al uit zichzelf geïnteresseerd waren in het programma. Een deel van dit probleem kan worden ondervangen door het uitvoeren van een voormeting. Met een covariantieanalyse kan dan in ieder geval nog statistisch gecontroleerd worden of er aanvangsverschillen bestaan tussen de twee groepen, middels een covariantieanalyse. Zie voor meer informatie over deze procedure Twisk, Vlakveld & Commandeur (2007).

4.2.4. *Bedreigingen van de validiteit*

Als een evaluatie van een verkeerseducatieprogramma bepaalde resultaten oplevert, is het altijd nodig om te vragen of die resultaten wel echt veroorzaakt worden door het programma. Zoals we al eerder zagen zijn er meestal ook andere verklaringen mogelijk. De mate waarin resultaten van een onderzoek mogen worden toegeschreven aan de behandeling (het volgen van het educatieve programma) noemen we '(interne) validiteit'. De eventuele andere verklaringen, die dus niets te maken hebben met het volgen van het programma, noemen we 'bedreigingen van de validiteit'. Bij het doen van evaluatieonderzoek is het zaak zo veel mogelijk bedreigingen van validiteit uit te sluiten, zodat (steeds) aannemelijk(er) wordt dat de resultaten verklaard kunnen worden door het programma. Een voorbeeld van een bedreiging van de validiteit is 'history': het optreden van een toevallige gebeurtenis (bijvoorbeeld een ernstig ongeval) tussen de voor- en nameting. Het gedrag van de participanten wordt zodanig beïnvloed door deze gebeurtenis, dat het gevolgen heeft voor de uitkomst van het experiment. Een andere bedreiging van de validiteit is 'mortality'. Dit vindt plaats als deelnemers na de voormeting niet meer meedoen aan het onderzoek, waardoor de groepen niet meer vergelijkbaar zijn. In de *Bijlage* staat een overzicht van mogelijke bedreigingen van de validiteit.

4.3. Analyse van de resultaten

Wanneer de gegevens zijn verzameld, zullen deze geanalyseerd moeten worden. Dit rapport gaat hier niet al te diep op in, omdat voor data-analyse een gespecialiseerde kennis over statistische analysemethoden nodig is. Het zal daarom in de meeste gevallen noodzakelijk zijn om een expert in te schakelen van buitenaf of van binnen de eigen organisatie. Wel is het nuttig om een aantal stappen te schetsen die over het algemeen doorlopen worden bij data-analyse.

Als eerste vindt meestal een globale inspectie van de data plaats. Hierbij wordt gekeken of er op het oog geen vreemde dingen in zitten (bijvoorbeeld variabelen waarop alle respondenten dezelfde waarde hebben of respondenten die dezelfde waarde hebben ingevuld bij alle variabelen). Deze variabelen of respondenten moeten uit de dataset worden verwijderd. Het zijn zogenaamde 'outliers'.

Vervolgens kijkt men naar de verzamelingen vragen die geacht worden een bepaald aspect te meten. Van deze set met vragen (bijvoorbeeld vragen om de houding ten opzichte van het dragen van een bromfietshelm te meten) wordt bekeken of ze samenhangen. Dat kan door correlaties tussen alle variabelen uit te rekenen of door een maat van interne samenhang te gebruiken: de betrouwbaarheidscoëfficiënt alfa. In de sociale wetenschappen wordt een alfa van boven de 0,70 over het algemeen als acceptabel beschouwd. We kunnen er dan van uitgaan dat de vragen allemaal ongeveer hetzelfde meten en mogen een samengestelde score berekenen: een gemiddelde of een somscore. Nu hebben we één variabele die het betreffende concept meet. Elke respondent heeft een eigen score op deze variabele. Vervolgens kan onderzocht worden of er verschillen zijn tussen voormeting en nameting en tussen experimentele groep en controlegroep. Dit gebeurt meestal met een variantieanalyse. Met deze analyse wordt de gemiddelde score op de voor- en nameting van de experimentele groep vergeleken met die van de controlegroep. Hierbij wordt niet alleen gekeken naar het gemiddelde, maar wordt ook rekening gehouden met de spreiding tussen de proefpersonen.

We concluderen dat het programma effectief is geweest als het verschil tussen voormeting en nameting in de experimentele groep groter is dan in de controlegroep.

Verder zijn er nog twee concepten die belangrijk zijn in de analyse: significantie en relevantie. Significantie wordt over het algemeen uitgedrukt in een p -waarde. Als een p -waarde kleiner is dan 0,05, betekent dit dat de kans dat het gevonden verschil toe te schrijven is aan toeval, kleiner is dan 5%. In de sociale wetenschappen is deze zogenaamde foutenmarge van 5% gangbaar. Een gevonden verschil kan significant zijn, maar toch weinig zeggen. Dat gebeurt bijvoorbeeld als men grote groepen proefpersonen heeft of heel weinig variantie in de data. Dat maakt de analyse erg gevoelig, waardoor zelfs heel kleine verschillen significant worden. De vraag is of het gevonden verschil dan ook echt relevant is. Wanneer de gemiddelde score op een vraag naar de houding ten opzichte van helmdracht verschuift van 3,5 naar 3,6 (op een vijfpuntsschaal), vinden we dit dan een succesvol resultaat? Om dit te bepalen kunnen we kijken naar een maat voor de

grootte van het effect. Statistische analyseprogramma's bieden verschillende mogelijkheden om de grootte van het effect te bepalen.

5. Het bijstellen van het programma

Wanneer de resultaten van de evaluatie binnen zijn, kan bekeken worden of het programma effectief is. Er zijn verschillende scenario's mogelijk: de evaluatie laat een positief resultaat zien, de evaluatie laat een negatief resultaat laten zien of de evaluatie laat geen resultaat zien. Dit hoofdstuk bespreekt deze drie scenario's. Zoals in de inleiding al gesteld, wordt in dit rapport alleen gekeken naar de effectiviteit van het programma op de geformuleerde (leer)doelen.

5.1. De evaluatie laat een positief resultaat zien

Een positief resultaat is uiteraard de best denkbare uitkomst van de evaluatie. Wel dient bekeken te worden of er alternatieve oorzaken te vinden zijn voor het effect (zie ook *Paragraaf 4.2.4*). Men moet dus, ook in het geval van een succesvol project, kritisch blijven op de kwaliteit. Ook zou het programma na verloop van tijd nogmaals geëvalueerd moeten worden.

5.2. De evaluatie laat geen resultaat zien

Het kan zijn dat de evaluatie geen resultaat laat zien. Mits de evaluatie optimaal is uitgevoerd (voldoende deelnemers, geen bedreigingen van de validiteit), is het uitblijven van resultaat aanleiding om te kijken naar de uitvoering van het programma. Hierbij kan teruggegrepen worden op de stappen die beschreven staan in *Hoofdstuk 2*; men moet opnieuw de vragen die daarbij horen stellen. Richt het programma zich wel op de juiste doelgroep? Vertoont de doelgroep het probleemgedrag wel? Is het gedrag waar het programma zich op richt wel gerelateerd aan verkeersveiligheid? Hoe is de didactische aanpak? Ook kan gekeken worden naar de uitvoering van het programma. Een eventuele procesevaluatie kan bijvoorbeeld boven tafel krijgen of het programma aanslaat bij de doelgroep.

5.3. De evaluatie laat een negatief resultaat zien

Een negatief resultaat kan, net als geen resultaat, te wijten zijn aan een suboptimale onderzoeksopzet. Het kan bijvoorbeeld zo zijn dat de houding ten opzichte van bepaald risicovol gedrag in de loop van de tijd negatiever is geworden, maar dat dit geldt voor de betreffende leeftijdsgroep in het algemeen. Wanneer een controlegroep ontbreekt, is een dergelijk effect niet uit te sluiten. Het is ook mogelijk dat het gehele onderzoek is opgezet volgens het meest ideale onderzoeksdesign: met een voor- en nameting voor zowel een experimentele als een controlegroep. Dan moet wel geconcludeerd worden dat dit programma averechts werkt en zal opnieuw moeten worden gekeken naar doelgroep, didactische aanpak enzovoort (zie *Paragraaf 5.2*).

6. Naar een verkort meetinstrument

Voor ontwikkelaars en uitvoerders van verkeerseducatieprogramma's is het uitzetten van een goede evaluatie onder de doelgroep vaak moeilijk. Kennis, attitude, gedrag enzovoort worden meestal gemeten aan de hand van vragenlijsten, maar deze vragenlijsten zijn vaak lang en soms moeilijk te begrijpen door de doelgroep. Andere meetinstrumenten, zoals observaties, zijn tijdrovend en logistiek ingewikkeld. Dit kan ertoe leiden dat een evaluatie niet wordt uitgevoerd of dat een uitgevoerde evaluatie onbruikbaar is doordat de vragen niet serieus zijn beantwoord.

Er is nog een andere reden waarom programma's soms niet of niet goed worden geëvalueerd. Soms lijkt de effectiviteit van een programma zo overduidelijk te zijn en zijn de betrokkenen zo enthousiast, dat evaluatie overbodig wordt geacht. Echter, resultaten uit het verleden laten zien dat zelfs de meest veelbelovende educatieprogramma's geen effect of zelfs een averechts effect op gedrag of verkeersveiligheid kunnen hebben.

Dit hoofdstuk bevat een voorstel voor een laagdrempelige methode die ontwikkelaars en uitvoerders kan helpen en motiveren om hun programma's (beter) te evalueren: een verkort meetinstrument.

6.1. De constructie van een itembank

De basisgedachte van het voorstel is om een database met vragen te ontwikkelen. Idealiter zijn veel van deze vragen in eerder onderzoek al toegepast en gevalideerd. Denk aan vragen die binnen het Effecten van Verkeerseducatie Onderzoek (EVEO) zijn gebruikt. Een dergelijke database met vragen wordt in de testleer ook wel een 'itembank genoemd'. Binnen de itembank zouden de items in twee dimensies moeten worden ingedeeld: *onderwerp* en *competentie*. Verkeerseducatieprojecten kunnen betrekking hebben op verschillende onderwerpen, zoals veilig fietsen, alcohol in het verkeer, verkeersregels, veilig oversteken en de dode hoek. Onder de term 'competentie' vallen zaken als kennis, attitude, gedrag, risicoperceptie en zelfreflectie.

Onder het onderwerp 'veilig oversteken' kunnen bijvoorbeeld vragen worden opgenomen die betrekking hebben op kennis (bijvoorbeeld "Is het veiliger om bij het oversteken zo snel mogelijk te lopen of juist heel rustig?") of juist op attitude ("Vind je het belangrijk om altijd eerst goed te kijken of er ander verkeer aankomt voor je oversteekt?"). Op deze manier ontstaat een soort matrixstructuur, waarbij in elke cel van de matrix slechts een beperkt aantal vragen staat (zie voor een voorbeeld *Tabel 6.1*).

	Kennis	Houding	Gedrag	Risicoperceptie	Enz.
Oversteken	<ul style="list-style-type: none"> - Hoe snel kun je het beste lopen als je moet oversteken? (kan ook met een filmpje!) - Loop je het beste in een rechte of een schuine lijn? - Schrijf op wat je zou doen voor je oversteekt 	<ul style="list-style-type: none"> - Vind je het belangrijk om altijd goed te kijken voor je oversteekt? - Als je haast hebt en snel oversteekt zonder goed te kijken, denk je dan wel eens: de auto's wachten vast wel voor mij? 	Hoe vaak heb je afgelopen dagen naar links en naar rechts gekeken voor je overstak?	Hoe gevaarlijk vind je het om: <ul style="list-style-type: none"> - zonder te kijken over te steken? - hard te rennen tijdens het oversteken? - langzaam te lopen tijdens het oversteken? 	
Fietsen	<ul style="list-style-type: none"> - Wat moet je doen om veilig linksaf te slaan? Som de handelingen op. - Als je een auto uit een zijstraat ziet komen, wanneer moet je dan voor deze auto stoppen? 	Hoe belangrijk vind je het om: <ul style="list-style-type: none"> - je hand uit te steken voor je afslaat? - op te letten wat een auto gaat doen? - voor een verkeerslicht dat net op rood is gesprongen te stoppen? 	Hoe vaak heb je de afgelopen dagen: <ul style="list-style-type: none"> - op groen gewacht? - je hand uitgestoken? - gewacht voor een voetganger op een zebraapad? 		
Brommer					
Alcohol en verkeer					
Enzovoort					

Tabel 6.1. Voorbeeld van een itembank.

6.2. Het construeren van een meetinstrument

Dankzij de itembank kan elke uitvoerder meteen aan de slag met de evaluatie. Belangrijk hierbij is dat vooraf duidelijke keuzes worden gemaakt over de onderwerpen en competenties die het belangrijkste zijn. Het zal namelijk vaak voorkomen dat een educatief programma zich richt op een veelheid aan competenties en soms ook op meerdere onderwerpen. Voor de evaluatie zal men dan een keuze moeten maken, omdat anders de vragenlijst veel te lang en omslachtig wordt. Dit betekent dat men moet accepteren dat over de effecten van het programma op sommige leerdoelen geen uitspraken kunnen worden gedaan.

Voorbeeld

'Je raakt zoek in de dode hoek' is een programma dat is ontwikkeld door het Regionaal Orgaan Verkeersveiligheid Gelderland. Het programma beoogt de kennis, houding en het gedrag van kinderen te verbeteren als het gaat om de dodehoekproblematiek. Om het effect van het programma op een laagdrempelige manier te evalueren, moet de organisatie die de evaluatie wil uitvoeren, bepalen welke leerdoelen het meest relevant zijn. Dodehoekprogramma's zijn vaak sterk kennisgericht, dus het ligt hier voor de hand om vast te stellen of de kennis van dit onderwerp na het volgen van het programma is toegenomen. Maar ook het herkennen van gevaarlijke situaties is binnen dit project belangrijk. Een evaluator kan ervoor kiezen om het project op deze twee onderwerpen mee te evalueren. Zaken als houding en gedrag komen dan in de eerste evaluatie niet aan bod. Die kunnen eventueel wel in een vervolgonderzoek worden geëvalueerd.

Voorbeeldvragen dode hoek: kennis

- Zet op deze tekening kruisjes waar de chauffeur van deze vrachtauto je zeker niet kan zien.
- Zet kruisjes waar de chauffeur je zeker wel kan zien.
- Deze vrachtauto wil rechts afslaan bij dit verkeerslicht. Het verkeerslicht staat op rood. Jij komt aanrijden op de fiets en wilt rechtdoor. Wat is de veiligste plaats om te wachten?

Voorbeeldvragen dode hoek, herkennen van gevaarlijke situaties

- Klik de foto's aan van situaties die je gevaarlijk vindt.
- Hoe gevaarlijk vind je:
 - o vlak achter een vrachtauto langslopen?
 - o vlak voor een vrachtauto langslopen?
 - o rechts naast een vrachtauto fietsen?
 - o op de stoep wachten voor een vrachtauto?
- Je ziet straks een serie foto's van een fietser en een vrachtauto. Klik aan welke foto lijkt op wat je zelf zou doen.

6.3. Karakterisering en selectie van vragen

Wanneer de itembank gereed is, kunnen de vragen worden gekarakteriseerd aan de hand van verschillende kenmerken. De twee belangrijkste kenmerken zijn die waarop de vragen zijn geconstrueerd, namelijk onderwerp en competentie. Maar er kunnen meer kenmerken worden genoteerd. Zo kan van elke vraag worden vastgesteld voor welke leeftijd de vraag geschikt is (bijvoorbeeld: geschikt voor kinderen van vier tot acht jaar), welk opleidingsniveau vereist is (bijvoorbeeld: geschikt voor vmbo, havo of vwo) en wat het type vraag is (bijvoorbeeld: tekstvraag, fotovraag, videovraag).

Op deze manier kan een afgewogen mix van vragen worden samengesteld die niet te talig is en die geschikt is voor de juiste leeftijd en het juiste kennisniveau.

7. Conclusies

Dit rapport bevat adviezen die ontwikkelaars en uitvoerders van verkeers-educatieprogramma's kunnen helpen om een evaluatie uit te (laten) voeren. Het rapport beoogt de evaluatie van dergelijke programma's te vergemakkelijken en heeft daarmee ook tot doel dat programma's vaker en zorgvuldiger worden geëvalueerd dan nu het geval is. Dit slothoofdstuk zet de belangrijkste zaken uit het rapport op een rijtje en blikt vooruit op de toekomst van het verkorte meetinstrument.

7.1. De kwaliteit van educatieve programma's

Om de kwaliteit van educatieve programma's te optimaliseren zou kwaliteitszorg een continu punt van aandacht moeten zijn. Op die manier blijft men steeds alert op mogelijke verbeteringen die doorgevoerd kunnen worden. De manier waarop kwaliteitszorg wordt ingevuld, is grotendeels een vrije keus van de uitvoerder of ontwikkelaar. In dit rapport wordt de PDCA-cyclus beschreven, een bekend principe uit de kwaliteitszorg dat uitgaat van kwaliteitszorg als cyclisch proces. In de PLAN-fase wordt het programma ontwikkeld en denkt men na over doelgroep, leerdoelen, didactische aanpak enzovoort. In de DO-fase wordt het programma daadwerkelijk uitgevoerd. In de CHECK-fase wordt bekeken of het programma voldoet aan de uitgangspunten zoals eerder geformuleerd. Deze evaluatie kan betrekking hebben op het proces en op de inhoud. Een inhoudelijke effectmeting vindt in deze fase plaats. De educatieve checklist (Vissers, 2010) is een instrument waarmee ook andere aspecten van verkeerseducatieprogramma's kunnen worden geëvalueerd. In de ACT-fase wordt bekeken of het programma moet worden bijgesteld.

7.2. Het belang van een goede evaluatie

Er zijn verschillende redenen om educatieve programma's te evalueren. Ten eerste is het belangrijk om te kunnen vaststellen of het programma inderdaad een bijdrage levert aan het verbeteren van de verkeersveiligheid. Het is lastig om dit direct af te meten aan aantallen ongevallen. Wel kan men onderzoeken of een programma leidt tot een verandering in gedrag waarvan bekend is dat het samenhangt met verkeersveiligheid. Ten tweede kunnen subsidieverstrekkingen met een goede evaluatie beter afwegen aan welk programma subsidie moet worden verstrekt. En ten derde kan evaluatie aan de oppervlakte brengen welke eventuele zwakten er in een programma aanwezig zijn en hoe deze verbeterd kunnen worden.

7.3. Voorwaarden voor een goede evaluatie

Voor een evaluatie is het belangrijk dat alternatieve verklaringen voor een gevonden effect zo veel mogelijk worden uitgesloten. Dat kan door de effectmeting op te zetten volgens een experimenteel design met een voor-en nameting en met een controlegroep. Idealiter worden deelnemers willekeurig toegewezen aan experimentele of controlegroep. De controlegroep dient zo veel mogelijk vergelijkbaar te zijn met de experimentele groep.

Wanneer geen effect of een averechts effect van het programma wordt gevonden, terwijl de evaluatie zorgvuldig is opgezet, dan kan dat verschillende oorzaken hebben. Het kan zijn dat het programma de beoogde doelgroep niet bereikt, dat het anders wordt uitgevoerd dan aanvankelijk bedacht of dat de deelnemers niet gemotiveerd zijn. Een tegenvallende evaluatie zou daarom aanleiding moeten zijn om alle ontwikkelstappen uit de CHECK-fase opnieuw te doorlopen.

7.4. Een verkort meetinstrument

Het uitvoeren van een evaluatie kan voor uitvoerders of ontwikkelaars een behoorlijke opgave zijn. Vragenlijsten zijn vaak lang en moeten meerdere keren worden ingevuld, waardoor het moeilijk is om de doelgroep gemotiveerd te houden. In dit rapport is een voorstel gedaan om op een laagdrempelige manier toch een evaluatie te kunnen uitvoeren. Door de constructie van een itembank met een beperkt aantal vragen per competentie en per onderwerp kan een evaluator vragen kiezen die het meest belangrijke onderwerp of de meest belangrijke competentie meten. Dat betekent wel dat er keuzes moeten worden gemaakt. Dat is soms lastig, zeker bij een programma dat een scala aan leerdoelen heeft, maar het levert wel een evaluatie op die op het betreffende aspect een betrouwbare uitspraak kan doen over het effect van het programma. Bovendien kunnen met een dergelijke itembank verschillende programma's met elkaar worden vergeleken; deze vergelijking is gemakkelijker te maken als steeds dezelfde (soort) vragen worden gebruikt.

Literatuur

- Aarts, L. & Schagen, I. van (2006). *Driving speed and the risk of road crashes: A review*. In: *Accident Analysis and Prevention*, vol. 38, nr. 2, p. 215-224.
- Blomberg, R.D., Peck, R.C., Moskowitz, H., Burns, M., & Fiorentino, D. (2005). *Crash risk of alcohol involved driving: A case-control study*. Dunlap and Associates, Inc., Stamford.
- Caird, J.K., Willness, C.R., Steel, P. & Scialfa, C. (2008). *A meta-analysis of the effects of cell phones on driver performance*. In: *Accident Analysis and Prevention*, vol. 40, nr. 4, p. 1285-1604.
- Clarke, D.D., Ward, P. & Truman, W. (2005). *Voluntary risk taking and skill deficits in young driver accidents in the UK*. In: *Accident Analysis and Prevention*, vol. 37, nr. 3, p. 523-529.
- Cochran, W.G. (1963). *Sampling techniques*. 2nd ed. John Wiley & Sons Inc., New York.
- Dragutinovic, N. & Twisk, D. (2006). *The effectiveness of road safety education: a literature review*. R-2006-6. SWOV Institute for Road Safety Research, Leidschendam.
- Evans, L. (1996). *Safety-belt effectiveness: the influence of crash severity and selective recruitment*. In: *Accident Analysis and Prevention*, vol. 28, nr. 4, p. 423-433.
- Goldenbeld, C., Houtenbos, M. & Ehlers, E. (2010). *Gebruik van draagbare media-apparatuur en mobiele telefoons tijdens het fietsen; Resultaten van een grootschalige internetenquête*. R-2010-5. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Leidschendam.
- Huijbers, J.J.W. & Kampen, L.T.B. van (1985). *Schatting van het effect van letselpreventiemaatregelen voor voetgangers, fietsers en bromfietzers bij botsingen met personenauto's*. R-85-36. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid SWOV, Leidschendam.
- Liu, B.C., Ivers, R., Norton, R., Boufous, S., Blows, S. & Lo, S.K. (2007). *Helmets for preventing injury in motorcycle riders*. In: *The Cochrane Database of Systematic Reviews* 2007, nr. 4.
- Mesken, J., Lajunen, T. & Summala, H. (2002). *Interpersonal violations, speeding violations and their relation to accident involvement in Finland*. In: *Ergonomics*, vol. 45, nr. 7, p. 469 - 483.
- Mynttinen, S., Sundström, A., Koivukoski, M., Hakuli, K., Keskinen, E., & Henriksson, W. (2009). *Are novice drivers overconfident? A comparison of self-assessed and examiner-assessed driver competences in a Finnish and*

Swedish sample. In: Transportation Research Part F: Traffic Psychology and Behaviour, vol. 12, nr. 2, p. 120-130.

Mynttinen, S., Sundström, A., Vissers, J., Koivukoski, M., Hakuli, K., & Keskinen, E. (2009). *Self-assessed driver competence among novice drivers - a comparison of driving test candidate assessments and examiner assessments in a Dutch and Finnish sample*. In: Journal of Safety Research, vol. 40, nr. 4, p. 301-309.

Parker, D., Reason, J.T., Manstead, A.S.R. & Stradling, S.G. (1995). *Driving errors, driving violations and accident involvement*. In: Ergonomics, vol. 38, nr. 5, p. 1036-1048.

Reason, J., Manstead, A., Stradling, S., Baxter, J. & Campbell, K. (1990). *Errors and violations on the road: a real distinction?* In: Ergonomics, vol. 33, nr. 10/11, p. 1315-1332.

Ruiter, R.A.C., Abraham, C. & Kok, G. (2001). *Scary warnings and rational precautions: a review of the psychology of fear appeals*. In: Psychology and Health, vol. 16, nr. 6, p. 613-630.

SWOV (2010). *Autogordels en kindersitjes*. SWOV-Factsheet december 2010. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid, Leidschendam.

Thompson, D.C., Rivara, F.P., & Thompson, R. (2004). *Helmets for preventing head and facial injuries in bicyclists (Cochrane Review)*. In: The Cochrane Library, nr 3. John Wiley & Sons Ltd., Chichester, UK.

Tronsmoen, T. (2008). *Associations between self-assessment of driving ability, driver training and crash involvement among young drivers*. In: Transportation Research Part F: Traffic Psychology and Behaviour, vol. 11, nr. 5, p. 334-346.

Twisk, D.A.M., Vlakveld, W.P. & Commandeur, J.J.F. (2007). *Wanneer is educatie effectief? Systematische evaluatie van educatieprojecten*. R-2006-28. Stichting Wetenschappelijk Onderzoek Verkeersveiligheid, Leidschendam.

Vissers, J., Betuw, A. van, Nägele, R., Kooistra, A. & Harteveld, M. (2005). *Leerdoelendocument Permanente Verkeerseducatie*. TT-04-056. Traffic Test, Veenendaal.

Vissers, J.A.M.M. (2010). *Checklist verkeerseducatie: Kwaliteitsindicatoren voor het beoordelen van verkeerseducatieprogramma's*. DHV, Amersfoort.

In deze bijlage worden factoren besproken die de validiteit of geldigheid van een evaluatie van een verkeerseducatieproject kunnen bedreigen.

History

Toevallige gebeurtenissen kunnen de uitkomst van een evaluatie beïnvloeden. Dit kan bijvoorbeeld gebeuren doordat er ook landelijk aandacht wordt besteed aan het onderwerp van het educatieprogramma. Ook kan het voorkomen dat op een school een dodehoekprogramma wordt gegeven en dat in dezelfde periode een van de leerlingen van de school betrokken raakt bij een dodehoekongeval. De impact van zo'n gebeurtenis op een school kan zo groot zijn dat alleen daardoor al de houding ten opzichte van verkeersveiligheid verandert.

Maturation

'Maturation' betekent simpelweg 'groei'. Als participanten tijdens een onderzoek (flink) veranderen door ontwikkelings- of groeiprocessen, kan dit de uitkomst van het onderzoek beïnvloeden.

Testing

Wanneer een participant dezelfde test voor de tweede keer invult, kan dit een gevolg hebben voor het resultaat. Bijvoorbeeld doordat de participant nu bekend is met de procedure of doordat er leereffecten optreden.

Instrumentation

Er is sprake van instrumentatie wanneer een gevonden effect veroorzaakt wordt door de manier waarop is gemeten, in plaats van door het volgen van het programma. Instrumentatie kan plaatsvinden doordat er in de nameting bijvoorbeeld een andere schaal wordt gebruikt of doordat er in de nameting andere beoordelaars zijn dan in de voormeting.

Regression

Regressie naar het gemiddelde is een statistisch fenomeen. Het betekent dat wanneer een participant binnen een steekproef extreem hoog of laag scoort, de kans groot is dat hij de volgende keer dichterbij het gemiddelde zit. Binnen een onderzoek naar educatie kan dit gebeuren wanneer participanten van een onderzoek niet random zijn geselecteerd, maar bij de eerste test juist lager of hoger scoren dan het gemiddelde. Bij een tweede test zullen zij dan altijd dichterbij het gemiddelde scoren, zonder dat dit te maken heeft met het educatieprogramma.

Mortality

'Mortality' betekent het verlies van participanten in het onderzoek. Bij een onderzoek gebruikt men meestal een voor- en een nameting. Als tijdens de manipulatie participanten afhaken, heeft dit invloed op de nameting, zeker als vooral de minder gemotiveerde deelnemers afhaken.

Selection

Het is belangrijk dat participanten random worden geselecteerd, waardoor er geen systematische verschillen ontstaan tussen de groepen. Door aselechte

toewijzing is de kans op systematische verschillen het kleinst. Wanneer er systematische verschillen zijn tussen individuen in de experimentele groep en de controlegroep, kunnen deze verschillen de uitkomst van de evaluatie beïnvloeden.