

ONGERUBRICEERD

**Behavioural and Societal
Sciences**Kampweg 5
3769 DE Soesterberg
Postbus 23
3769 ZG Soesterberg

www.tno.nl

T +31 88 866 15 00
F +31 34 635 39 77**TNO-rapport****TNO 2014 R10243****Evaluatie van leerinterventies en
teamfunctioneren in dynamische teams**

Datum	februari 2014
Auteur(s)	Dr. J.E. Korteling Dr. E.A.P.B. Oprins Ir. W. Venrooij
Exemplaarnummer	
Oplage	16
Aantal pagina's	52 (incl. bijlagen, excl. RDP en Distributielijst)
Aantal bijlagen	1
Opdrachtgever	Ministerie van Defensie
Projectnaam	V1212-P7-Meten: Evaluatie
Projectnummer	053.01961

Alle rechten voorbehouden.

Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt door middel van druk, foto-kopie, 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 belang-hebbenden is toegestaan.

© 2014 TNO

ONGERUBRICEERD

Samenvatting

De moderne militair moet zich kunnen handhaven in een dynamische en complexe (internationale) omgeving waarin veelal in coalitie-verband moet worden opgetreden (joint, interagency) en waarbij de teamsamenstelling niet vooraf vaststaat.

De voorbereiding van de militair op het werken in dynamische teams als onderdeel van geïntegreerde eenheden is daardoor niet eenvoudig. Daarnaast is voor het trainen van teams meer nodig dan het opleiden van de individuele teamleden.

Daarom dienen nieuwe leerinterventies te worden ontwikkeld waarvan de effectiviteit ook moet worden geëvalueerd. Er is daarbij vooral behoefte aan instrumenten waarmee continu, snel en ad hoc het momentane teamfunctioneren kan worden vastgesteld.

De doelstelling van dit project is daarom het ontwikkelen van methoden voor het bepalen van de leeropbrengst van leerinterventies en voor het evalueren van samengestelde dynamische teams in het bijzonder. Dit rapport levert als eerste een generiek *evaluatieraamwerk* voor het meten van de effectiviteit van leerinterventies. Dit raamwerk kan worden gebruikt voor het evalueren van leerinterventies en het meten van transfer of training en voor het doen van evaluatie- en effectonderzoek. Het raamwerk is niet alleen bedoeld voor het meten van interventie-effecten, maar vooral ook op het krijgen van inzicht in de onderliggende processen van die effecten. Daarnaast levert het evaluatieraamwerk informatie over constructen die gemeten kunnen worden en over niveaus waarop effecten kunnen worden gemeten, typen prestatie-maten, methodieken, onderzoeks-designs, typen meetinstrumenten en kwaliteitscriteria.

Vervolgens wordt dieper ingegaan op methode-ontwikkeling voor het daadwerkelijk evalueren van dynamische teams. Daartoe wordt een gedeelte van het generieke evaluatieraamwerk voor teamtraining verder uitgewerkt. Op basis van literatuur-onderzoek, onder andere naar modellen voor team-functioneren, wordt hiervoor een prototype meetinstrument, getiteld *TeamScan*, ontwikkeld en beproefd. Dit meetinstrument is te gebruiken voor zowel het evalueren van het resultaat van team-trainingsinterventies als van het evalueren van teamfunctioneren on-the job. *TeamScan* is bedoeld als een soort “peilstok” waarmee de meest relevante aspecten van teamfunctioneren snel en herhaaldelijk kunnen worden gemeten. Door de relatief eenvoudige wijze waarop longitudinale, herhaalde metingen kunnen worden uitgevoerd, kunnen juist de dynamische aspecten van teams door de tijd heen goed in kaart worden gebracht.

Tot slot werd *TeamScan* als prototype d.m.v. een pilot beproefd. De resultaten gaven aan dat *TeamScan* een bruikbaar instrument is dat prima in staat lijkt om veranderingen in de teamsituatie waar te nemen. Daarnaast lijkt het instrument goed bruikbaar, niet alleen voor het groepsbeeld, maar ook voor individuele beelden, en discrepanties daartussen.

De kennis van dit rapport en het bijbehorende meetinstrument kan worden gebruikt voor het doen van een breed scala aan (militaire) effectstudies. Dat kan betrekking hebben op effecten van leerinterventies, maar ook onderzoek naar de ontwikkeling van teamfunctioneren of effecten van allerlei andere factoren die van invloed kunnen zijn op teameffectiviteit. Hierbij kan worden gedacht aan effecten van verandering van taken, teamsamenstelling, nieuwe (geavanceerde) opleidings-leermiddelen, omgevingsfactoren, structuur van de trainingsorganisatie, nieuwe inhoudelijke inrichting van trainingstrajecten en combinaties hiervan.

Inhoudsopgave

	Samenvatting	2
1	Inleiding	4
1.1	Aanleiding	4
1.2	Dynamische teams	4
2	Inleiding	6
2.1	Onderzoek naar de effectiviteit van training	6
2.2	Bestaande evaluatiemodellen	7
2.3	Generiek evaluatieraamwerk voor training	11
2.4	Indicatoren in het evaluatieraamwerk.....	13
3	Onderzoeksmethoden (design).....	18
3.1	Experimenteel onderzoeksdesign	18
3.2	Operationaliseren van indicatoren.....	21
3.3	Meetinstrumenten	22
3.4	Keuze van meetinstrumenten.....	26
3.5	Realiseerbaarheid, bruikbaarheid, betrouwbaarheid en validiteit	26
4	Metten van de effectiviteit van dynamische teams	29
4.1	Inleiding dynamische teams	29
4.2	Effectiviteit van teamtraining en teamprestaties.....	29
4.3	Teamwork model	32
4.4	Meetinstrument	35
5	Pilot studie	39
5.1	Inleiding en vraagstelling	39
5.2	Teamtaak.....	39
5.3	Procedure	40
5.4	Resultaten.....	40
5.5	Voorlopige conclusies met betrekking tot het meetinstrument.....	43
6	Conclusies en toepassing	44
6.1	Conclusies	44
6.2	Toepassing	46
7	Referenties	48
	Bijlage(n)	
	A Tabel Correlaties	

1 Inleiding

1.1 Aanleiding

De krijgsmacht treedt op in verschillende joint, combined en/of interagency samenwerkingsverbanden. Voor de militair betekent dit dat hij zich moet kunnen handhaven in een dynamische en complexe (internationale) omgeving waarin actoren verschillende belangen nastreven en waar de teamsamenstelling niet vooraf vaststaat, maar varieert afhankelijk van de opdracht. Het werken van militairen in zogenaamde dynamische teams is niet eenvoudig. Voor het trainen van teams is dan ook meer nodig dan het opleiden van de individuele teamleden. Training van deze militairen/teams zal zich moeten richten op die competenties die flexibiliteit in taakuitvoering binnen teams bevorderden. Daarnaast zullen de trainingsmethoden, -technieken en leermiddelen flexibel moeten kunnen inspelen op veranderingen.

De mogelijkheden voor 'live' training worden steeds meer beperkt, enerzijds door het vaak 'niet beschikbaar zijn' van alle betrokkenen tijdens de trainingsmomenten en anderzijds doordat de trainingen op zich onder druk staan door allerlei beperkingen, ook financieel. Daarom wordt steeds meer gebruik gemaakt van innovatieve alternatieven, zoals synthetische en embedded leeromgevingen, al of niet gekoppeld. Daarbij moet het rendement van de schaarse trainingen verder omhoog. Daartoe is het niet alleen nodig dat middelen, zoals simulatoren, op technisch vlak onderling en met operationele systemen kunnen samenwerken, maar ook dat de trainingsmethoden en organisatie hier adequaat op is ingericht. In dit verband dient tevens het ontstaan van nieuwe leerparadigma's te worden genoemd. Hierbij staat de lerende meer centraal en is deze meer verantwoordelijk voor zijn eigen leerproces. Daarbij worden steeds meer nieuwe media gebruikt om leeromgevingen te creëren waarin de lerenden doelgericht kunnen oefenen of ervaringen kunnen uitwisselen.

De ultieme vraag is in hoeverre al deze veranderingen ook daadwerkelijk *effectief* en *rendabel* zijn? Dit is van belang, enerzijds om er achter te komen wat wel en wat niet werkt en waarom dit zo is teneinde het proces zoveel mogelijk te optimaliseren. Anderzijds is dit van belang om de juiste beleidskeuzes te maken ten aanzien van toekomstige alternatieven en investeringen. Dit vereist een goede en breed toepasbare (generieke) toolbox opgehangen aan een raamwerk om leereffecten en rendement van allerlei O&T processen in kaart te brengen. Het gaat hier om een overkoepelend evaluatieraamwerk bestaande uit modellen, methoden, meet-instrumenten en indicatoren die nodig zijn bij het bepalen van de effectiviteit en efficiency van leerinterventies en – innovaties. Hiermee kunnen de effectiviteit en het rendement van opleiding, training en oefening, bijvoorbeeld met betrekking tot dynamische teams, beter worden onderbouwd en gemeten.

1.2 Dynamische teams

Een (conventioneel) team kan worden gedefinieerd als twee of meer personen die een specifieke rol hebben, onderling afhankelijke taken uitvoeren, zich aanpassen, en een gemeenschappelijk doel delen (Krokos, Baker, Alonso, & Day, 2009, p. 384). Een dynamisch team kan worden beschouwd als een team met de volgende aanvullende kenmerken:

- De teams bestaan uit mensen die zelden of nooit met elkaar hebben gewerkt.
- Het team is in principe samengesteld om een korte-termijn doel te bereiken. De teamsamenstelling varieert daartoe afhankelijk van de specifieke (korte-termijn) opdracht en gaat weer uiteen zodra het doel is bereikt.
- Het team opereert in snel veranderende omstandigheden en moet hiermee dus flexibel kunnen omgaan, c.q. hier op inspelen.

Dynamische teams, zoals hierboven omschreven, hebben veel overeenkomst met ad hoc teams (Mills et al., 1999.) met dit verschil dat in een dynamisch team meer nadruk ligt op verandering (in de omgeving, dan wel binnen het team) en daarmee op adaptiviteit en flexibiliteit.

Naar het meten en evalueren van de effectiviteit van dynamische teams is nog weinig onderzoek gedaan. Er is vooral behoefte aan methoden en tools geschikt voor meting en evaluatie van het leerproces over de hele (longitudinale) keten van Opleiding, Training, Oefening, Operationeel Gereed, Inzet Gereed en Inzet. Daarnaast dienen, gegeven het voortdurend veranderende spectrum aan teamsamenstellingen en omstandigheden waarin teams moeten kunnen functioneren, instrumenten te worden ontwikkeld die continu, snel en ad hoc teamprocessen kunnen monitoren. Dit rapport geeft hiertoe een aanzet.

De doelstelling van het project is het ontwikkelen van kennis voor het evalueren van verschillende soorten leerinterventies (opleiding, training, oefening) en met een speciale focus op dynamische militaire teams. Het rapport levert daarom op basis van de wetenschappelijke literatuur als eerste een generiek evaluatieraamwerk voor het onderzoeken van de effectiviteit van leerinterventies. Dit raamwerk bevat indicatoren (cf. constructen die gemeten kunnen of moeten worden), informatie over niveaus waarop effecten kunnen worden gemeten, methodieken, typen prestatie-maten en meetinstrumenten, onderzoeks-designs en kwaliteitscriteria. Het levert daarmee een wetenschappelijk gefundeerde basis voor de methodologische opzet van onderzoek. Vervolgens gaat het rapport meer specifiek in op de evaluatie van het functioneren van dynamische teams. Als eerste wordt ingezoomd op het evalueren van leerinterventies bij teams en op het meten en evalueren van prestaties en adaptiviteit van dynamische teams. Tot slot wordt, op basis van een voor dit doel geschikt teamprestatie-model, een meetinstrument ontwikkeld en beproefd.

2 Inleiding

Dit hoofdstuk beschrijft de resultaten van het literatuuronderzoek dat heeft geleid tot een generiek evaluatieraamwerk waarmee onderzoek kan worden gedaan naar de effectiviteit van leerinterventies. Dit evaluatieraamwerk is in principe generiek maar kan specifiek worden ingevuld voor de evaluatie van teamtraining.

2.1 Onderzoek naar de effectiviteit van training

Onderzoek naar de effectiviteit van leerinterventies in de praktijk, zoals bijv. een (beroeps)opleiding, simulator training, serious game, teamtraining of een andere interventie wordt zelden op een wetenschappelijk gefundeerde manier uitgevoerd. (Salas et al., 2003; Akl et al., 2010; Korteling & Oprins, 2013; Sitzmann, 2011). In het algemeen blijkt het lastig om een effectiviteitsonderzoek methodologisch goed op te zetten zoals bijvoorbeeld het inzetten van voldoende proefpersonen, het regelen van een controlegroep en het betrouwbaar uitvoeren van geschikte voor- en nametingen (Korteling & Oprins, 2013). Daarnaast blijkt het moeilijk om eenduidige uitspraken te doen over een bepaalde leerinnovatie omdat evaluatiestudies op een uiteenlopende manier worden uitgevoerd. Dit blijkt bijvoorbeeld uit reviewstudies die zijn gedaan op het gebied van serious gaming ten behoeve van leren (bijvoorbeeld Akl et al., 2010; Hartevelde, 2012; Hays, 2005; Mayer, 2012; Sitzmann, 2011).

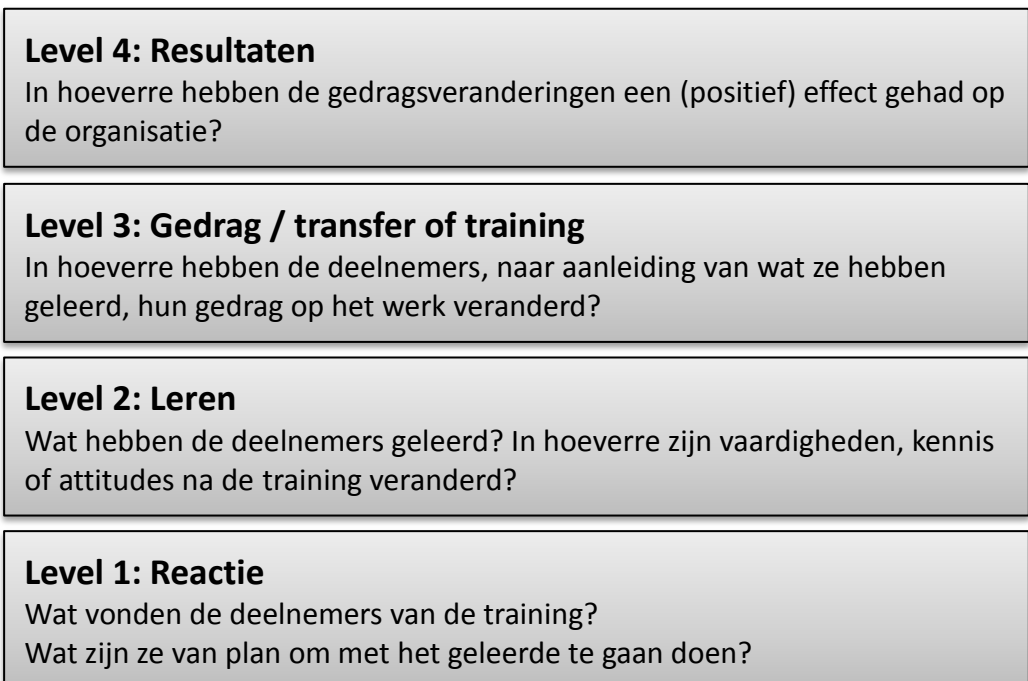
Er is relatief veel effectiviteitsonderzoek uitgevoerd in het medische domein, van waaruit de term 'evidence-based learning' afkomstig is (Schilling et al., 2005). Onder 'evidence-based learning' verstaan we het ontwerpen van leerinterventies gebaseerd op empirische bewijzen dat de interventie ook *daadwerkelijk* effect heeft. We leven in een tijd waarin kosten een belangrijkere rol gaan spelen, mede omdat leerinnovaties technologisch steeds geavanceerder worden en daardoor een hogere investering vragen. Daarom is het van toenemend belang aan te tonen dat de leerinterventie effectief is. Dit geldt des te meer voor domeinen waarin de veiligheid centraal staat zoals het militaire domein of de luchtvaart (Salas et al., 2003; Oprins, 2008).

Onderzoek naar de effectiviteit van leerinterventies vereist gevalideerde methoden en meetinstrumenten. De ontwikkeling van een generiek evaluatieraamwerk moet hieraan bijdragen. Zo'n evaluatieraamwerk bestaat uit een methodiek en een set constructen of indicatoren die geoperationaliseerd zijn tot een (gevalideerd) meetinstrument, in het onderhavige geval voor teamfunctioneren. Het streven is om een evaluatieraamwerk te ontwikkelen dat tot op zekere hoogte generiek is voor het evalueren en valideren van een breed scala aan typen leeroplossingen, aangevuld met trainings- en domein-specifieke metingen. Dit generieke karakter maakt het in de eerste plaats mogelijk om sneller te komen tot betrouwbare en valide meetinstrumenten. Bovendien kunnen uitspraken worden gedaan over meerdere leeroplossingen indien de effectiviteit ervan gemeten is met behulp van dezelfde meetinstrumenten. Op deze manier kan inzicht verkregen worden in welke leeroplossingen effectiever zijn dan andere. Dit evaluatieraamwerk moet vooral praktisch ingezet kunnen worden bij trainingen die buiten het laboratorium worden uitgevoerd, bijvoorbeeld bij teamtraining.

Het onderzoeksprogramma, waar het onderhavige project deel van uit maakt, is gericht op dynamische teams in een dynamische omgeving. Hoewel het evaluatieraamwerk zelf generiek is, wordt dit raamwerk in dit project specifiek uitgewerkt voor teamtraining. Daarbij is het van belang dat het evaluatieraamwerk en de hieruit volgende meetinstrumenten zich er goed voor lenen een aantal keer achtereenvolgend bij hetzelfde team gebruikt te kunnen worden teneinde veranderingen ten gevolge van de dynamiek te kunnen meten.

2.2 Bestaande evaluatiemodellen

In het verleden zijn er diverse pogingen gedaan om een generiek evaluatiemodel te ontwerpen voor de evaluatie van leeroplossingen. Het meest populaire model dat nog steeds veelvuldig in de praktijk wordt gebruikt is dat van Kirkpatrick (1976, 2006). Volgens dit model zijn er vier niveaus waarop een opleiding, training of oefening geëvalueerd kan worden: 1) reactie of beleving van de leerling over de training, doorgaans gemeten met behulp van vragenlijsten of interviews; 2) directe leereffecten van de training zelf (behaalde kennis, vaardigheden en houdingen), meestal gemeten in tests en assessments die in principe los staan van gedragseffecten in de praktijk; 3) gedrag in de praktijk na afloop van de training; 4) effecten van de training voor de organisatie als geheel, waar de trainee deel van uitmaken. Het meeste evaluatieonderzoek beperkt zich tot niveau 1 en 2 (Cohn et al., 2009). Later is hier een vijfde niveau aan toegevoegd dat betrekking heeft op return-on-investment (ROI) dat aangeeft wat de training heeft opgeleverd in termen van kosten en baten, zie Figuur 1.



Figuur 1 Evaluatiemodel van Kirkpatrick (1976, 2006).

De meerwaarde van dit model van Kirkpatrick is dat de effectiviteit van leeroplossingen gemeten moet worden op diverse niveaus, dus niet alleen de uitkomst direct na afloop van het leren. De populariteit van dit model in praktisch gebruik komt bovendien door de systematische wijze waarin leeroplossingen geëvalueerd zouden moeten worden en de vereenvoudigde weergave van het complexe proces van evaluatie (Alliger & Janak, 1989; Bates, 2004). Het model heeft echter enkele beperkingen.

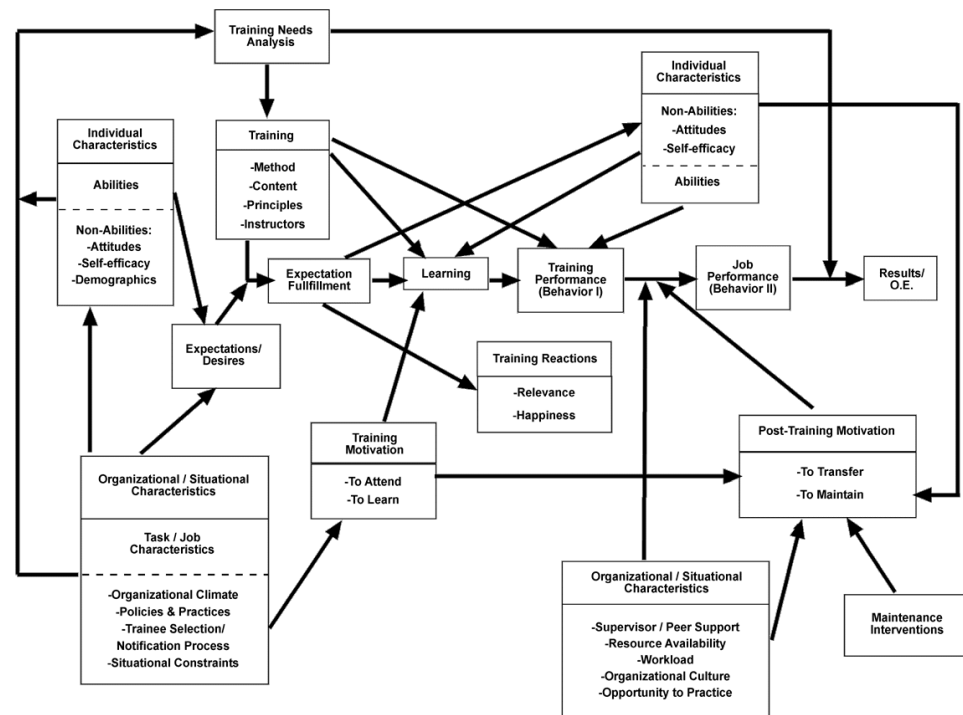
Het model zou een zekere causaliteit kunnen suggereren, namelijk dat de reacties van de leerlingen (niveau 1) het leren stimuleren (niveau 2) met als gevolg transfer naar de praktijk (niveau 3) en uiteindelijk resultaten op organisatieniveau (niveau 4). Van een dergelijk sterke causale relatie tussen de verschillende niveaus hoeft lang niet altijd sprake te zijn en de niveaus hoeven niet perse met elkaar samen te hangen (Alliger & Janak, 1989; Bates, 2004).

Belangrijker nog is de kritiek dat het model van Kirkpatrick alleen gericht is op de leeruitkomsten. Er kan hierbij onderscheid gemaakt worden tussen *evaluatie*- en *effectiviteitsonderzoek* (Alvarez, Salas & Garofano, 2004; Kraiger, Ford & Salas, 1993; Salas & Cannon-Bowers, 2001). Evaluatieonderzoek meet de mate waarin de leerlingen bepaalde leeruitkomsten hebben bereikt, terwijl effectiviteitsonderzoek probeert deze leeruitkomsten te verklaren. Effectiviteitsonderzoek beschouwt hiermee het leerproces als geheel en onderzoekt dus ook *waarom* bepaalde leeruitkomsten wel of niet worden behaald. Alvarez, Salas & Garofano (2004) maken dan ook onderscheid tussen evaluatiemodellen en effectiviteitsmodellen.

Het model van Kirkpatrick is uitsluitend een evaluatiemodel. Later is dit model verder uitgewerkt door bijvoorbeeld Tannenbaum et al., (1993) en Cannon-Bowers et al., (1995), Holton (1996), Kraiger (2002), en Alvarez, Salas & Garofano (2004). Deze modellen hebben gemeen dat ze de type uitkomstvariabelen verder uitwerken. Belangrijk is het onderscheid tussen verschillende type uitkomsten, bijvoorbeeld de indeling in cognitieve, skill-based en affectieve leeruitkomsten (Kraiger, Ford & Salas, 1993) dat niet weergegeven is in het model van Kirkpatrick. Een dergelijke onderverdeling moet worden gemaakt omdat ze verschillende type metingen vereisen in evaluatieonderzoek. Affectief leren bijvoorbeeld wordt doorgaans gemeten in (subjectieve) zelfevaluatie terwijl cognitieve uitkomsten meer objectief gemeten kunnen worden in diverse testvormen (zie ook par 3.3). Daarnaast zijn verschillende typen uitkomsten op organisatie- en gedragsniveau van belang (Salas, Milham & Bowers, 2003).

In tegenstelling tot evaluatieonderzoek is effectiviteitsonderzoek gericht op de *invloeden* op de leeruitkomsten. Daarom bevatten effectiviteitsmodellen de kenmerken van individuele leerlingen, de leerinterventie en de organisatie of omgeving. Ten opzichte van bovengenoemde *uitkomstvariabelen* worden deze kenmerken aangeduid als *procesvariabelen*. Deze variabelen beïnvloeden het leerproces voor, tijdens en na de leerinterventie (Alvarez, Salas & Garofano, 2004). De genoemde modellen van Holton (1996), Tannenbaum et al., (1993) en Cannon-Bowers et al., (1995), en Alvarez, Salas & Garofano (2004) bevatten dan ook deze drie typen procesvariabelen die van invloed zijn op leren. Ook het model van Baldwin & Ford (1988), dat gericht is op transfer, omvat deze typen procesvariabelen. Het aantal en type procesvariabelen dat in de diverse modellen voorkomt verschilt van elkaar. We presenteren hier in Figuur 2 het Evaluatiemodel

van Tannenbaum et al., (1993) en Cannon-Bowers et al., (1995) omdat dit model veel van de procesvariabelen weergeeft die relevant zijn voor leren. Het model geeft weer hoe complex trainingseffectiviteit is en welke factoren van invloed kunnen zijn op leren en op leeruitkomsten op korte en langere termijn. Zo zijn belangrijke factoren op individueel niveau vermeld die het leerproces beïnvloeden waaronder *self-efficacy* en motivatie. Op organisatieniveau worden bijvoorbeeld ondersteuning (support) en cultuur genoemd.

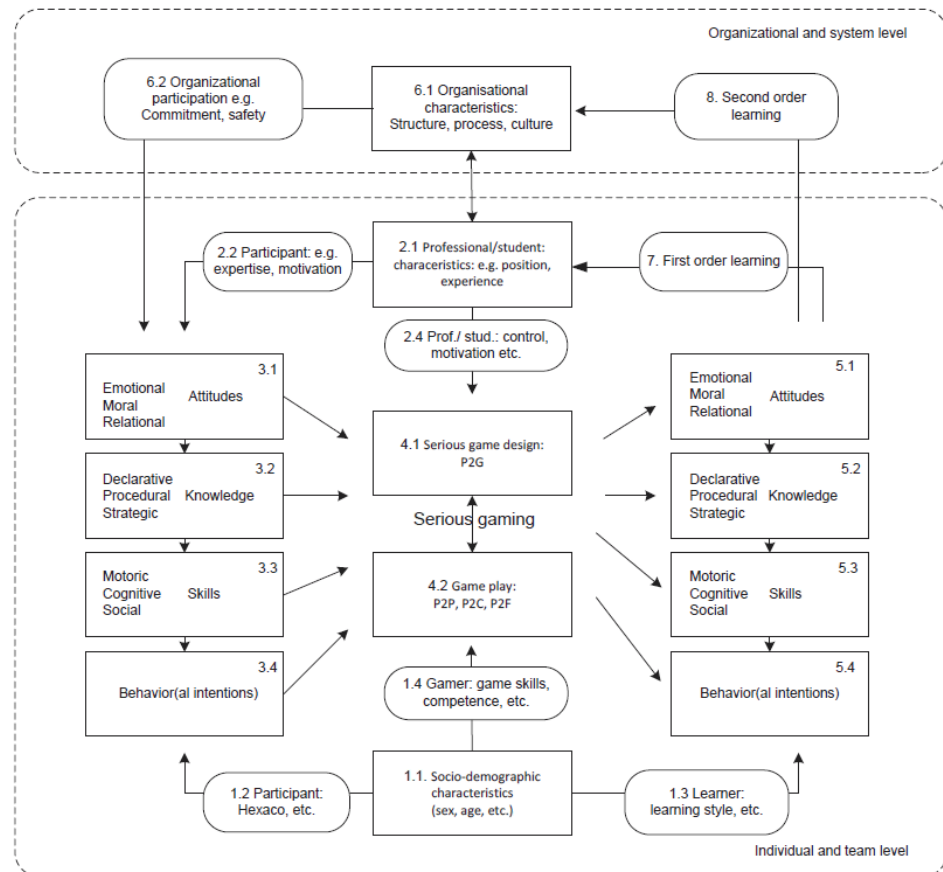


Figuur 2 Evaluatiemodel van Tannenbaum et al., (1993), Cannon-Bowers et al., (1995).

Kraiger, Ford & Salas (1993) benadrukken de *constructbenadering* in evaluatie-onderzoek. Zij splitsen verschillende typen variabelen op in generieke constructen. Zo behoort motivatie tot de affectieve variabelen die de constructen 'motivational disposition', 'self-efficacy' en 'goal setting' bevatten. Deze constructbenadering heeft twee voordelen. Ten eerste kunnen de mechanismen die bijdragen aan domeinspecifieke leeruitkomsten worden geëvalueerd. Bovendien heeft de constructbenadering potentie voor de validatie van de metingen over meerdere studies heen. Dit houdt in dat steeds dezelfde meetinstrumenten (bijvoorbeeld vragenlijsten) kunnen worden gebruikt die dezelfde constructen meten. Hierdoor kunnen uitspraken worden gedaan over de invloed van de diverse variabelen op leren, zoals motivatie en self-efficacy (Kraiger, Ford & Salas, 1993).

Behalve deze generieke evaluatiemodellen zijn er ook leerspecifieke evaluatiemodellen ontwikkeld voor bepaalde leeroplossingen. Connolly, Stansfield en Hailey (2009) beschrijven welke modellen er bestaan voor serious gaming, maar deze zijn op een algemeen niveau. Mayer (2012) daarentegen heeft een model ontwikkeld voor serious gaming dat lijkt op bovengenoemde effectiviteitsmodellen. Dit model bevat diverse typen variabelen op individueel, team, organisatie- en op systeemniveau die ook zijn weergegeven in de meer generieke evaluatiemodellen (bijvoorbeeld Kraiger, Ford & Salas, 1993; Tannenbaum et al., 1993; Cannon-Bowers et al., 1995), zie Figuur 3. Net als in het model van Kirkpatrick worden de

uitkomstvariabelen gemeten op meerdere niveaus, bijv. op organisatieniveau. Het model van Mayer (2012) bevat de condities voor, tijdens en na de interventie op basis waarvan (quasi) experimentele onderzoeks-designs kunnen worden afgeleid. Mayer (2012) benadrukt tevens dat er enerzijds generieke meetinstrumenten nodig zijn om vergelijkend onderzoek te kunnen doen over verschillende serious games (generaliseerbaarheid en standaardisatie), en anderzijds dat specifieke instrumenten nodig zijn om de leeruitkomst te meten die voortdurend anders is, afhankelijk van wat er in de training wordt geleerd (specificiteit en flexibiliteit).



Figuur 3 Evaluatiemodel (Mayer, 2012).

Samengevat, een goed model voor evaluatie alsmede het meten van de effectiviteit van leeroplossingen zou zowel uitkomst- als procesvariabelen moeten bevatten (Alvarez, Salas & Garofano, 2004; Kraiger, Ford & Salas 1993; Mayer, 2012; Tannenbaum et al., 1993; Cannon-Bowers et al., 1995). De effecten moeten op diverse niveaus worden gemeten zoals Kirkpatrick (1976, 2006) reeds benadrukte: niet alleen de directe reacties en leeruitkomsten moeten worden gemeten maar ook transfer naar de praktijk en effecten op organisatieniveau. De leeruitkomsten moeten echter verder worden uitgesplitst door bijvoorbeeld onderscheid te maken in cognitieve- vaardigheid- en affectieve leeruitkomsten (Kraiger, Ford & Salas, 1993; Salas, Milham & Bowers, 2003). De constructbenadering van Kraiger, Ford and Salas (1993) houdt in dat niet alleen domeinspecifieke leeruitkomsten gemeten moeten worden zoals wordt gesuggereerd in het model van Kirkpatrick. Het meten van generieke constructen is nodig om een breder inzicht in leren te verkrijgen en vergelijkend onderzoek beter mogelijk te maken (Mayer, 2012). Het model zou

tevens invloedsfactoren moeten bevatten die betrekking hebben op voor, tijdens en na de leerinterventie (Alvarez, Salas & Garofano, 2004; Mayer, 2012).

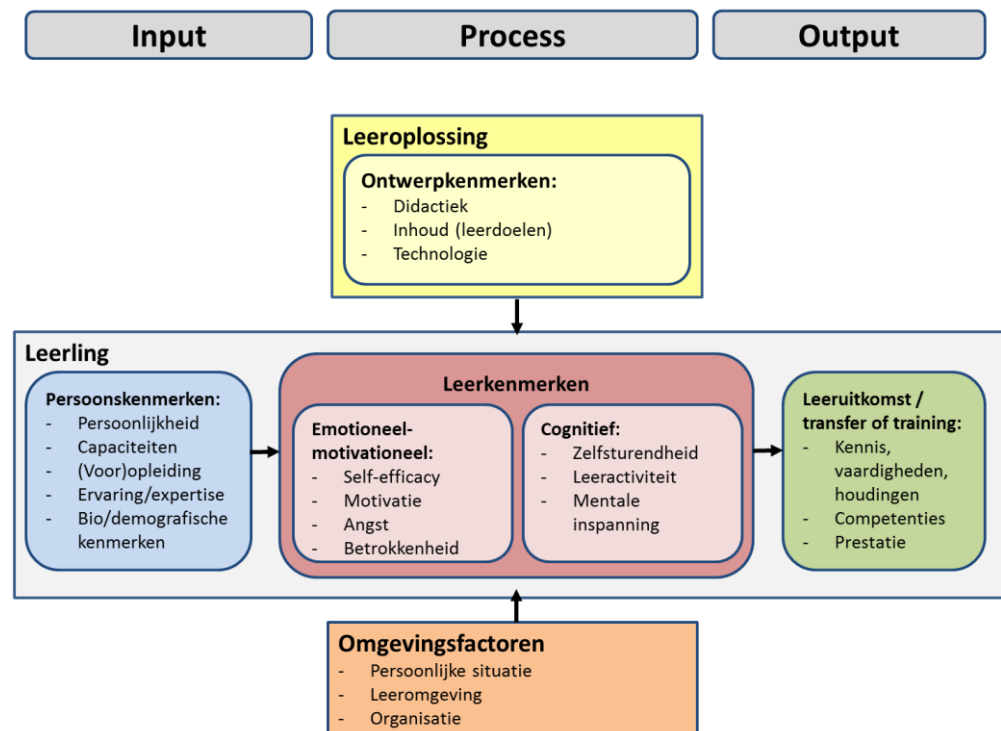
2.3 Generiek evaluatieraamwerk voor training

Het generiek evaluatieraamwerk dat wij op basis van bovenstaande argumentatie voorstellen, bouwt voort op de bestaande evaluatiemodellen uit de literatuur. Het kan dienen als leidraad voor het opzetten van studies naar de effectiviteit van trainingen. De basisgedachte hierbij is dat we bij effectiviteitsstudies inzicht willen krijgen in de *onderliggende* factoren van het leerproces. Dat betekent dat we niet uitsluitend naar de leer-input en leer-output moeten kijken, de zogenaamde *Black-Box* benadering, maar tevens naar het *leerproces* zelf. De diverse *beïnvloedende factoren* worden dan meegenomen zodat we inzicht krijgen in *hoe* er wordt geleerd.

Hoe het raamwerk voor het evalueren van leerinterventies moet worden gebruikt is van vele factoren afhankelijk, zoals: de vraagstelling, doelgroep, de trainings-situatie, het type interventie zelf, beschikbare hulpmiddelen en infrastructuur en de omgeving waarin dit gebeurt. Daarom is het moeilijk een algemene gebruiksaanwijzing te geven voor de wijze waarop het raamwerk wordt toegepast of welke aspecten ervan wordt meegenomen. Het belangrijkste is dat het raamwerk altijd als denkkader wordt gebruikt die structuur biedt bij het opzetten van evaluatie-onderzoek. Daarbij kan het raamwerk tevens dienen als “checklist” die ervoor zorgt dat er geen belangrijke aspecten of dimensies van het onderzoek over het hoofd hoeven te worden gezien. De stappen die daarbij worden doorlopen worden in het hoofdstuk 3 behandeld.

Hoewel dit Evaluatieraamwerk zich vooralsnog beperkt tot het *individuele* niveau (micro), kan dit tevens worden gebruikt op naar groeps- of teamniveau (meso). De omgeving (macro) wordt in dit model beschouwd als beïnvloedende factor vanuit bijvoorbeeld de organisatie, persoonlijke omstandigheden of relevante ervaringen buiten de leerinterventie om.

Het evaluatieraamwerk is weergegeven in Figuur 4:



Figuur 4 Generiek evaluatieraamwerk effectiviteit training.

Het voorgestelde evaluatieraamwerk heeft bepaalde basisprincipes. In lijn met Alvarez, Salas & Garofano (2004) en Mayer (2012) wordt dit raamwerk gepresenteerd als een input – proces – output model. Het is van belang te weten wat er voorafgaand, tijdens en na afloop van de leerinterventie gemeten moet worden. Zoals in vele andere modellen (bijvoorbeeld Tannenbaum et al., 1993; Mayer, 2012) wordt de input bepaald door bepaalde persoonskenmerken van de lerende. Een ander basisprincipe is dat het model zowel proces- als uitkomstvariabelen bevat zoals in andere evaluatiemodellen (Alvarez, Salas & Garofano, 2004; Cannon-Bowers et al., 1995; Kraiger, Ford & Salas, 1993; Tannenbaum et al., 1993, Mayer, 2012). Dit is nodig omdat dit model zich niet beperkt tot evaluatieonderzoek maar met name ook van toepassing is op effectiviteitsonderzoek.

De procesvariabelen zijn generiek zodat diverse leerinterventies met elkaar vergeleken kunnen worden (Mayer, 2012).

De uitkomstvariabelen zijn domeinspecifiek, afhankelijk van de leerdoelen. De uitkomstvariabelen moeten op diverse niveaus worden gemeten zoals reeds voorgesteld door Kirkpatrick (1976, 2006) waarbij ook transfer moet worden gemeten (Baldwin & Ford, 1988). Ze moeten worden geoperationaliseerd tot meetinstrumenten die direct gerelateerd zijn aan de leerdoelen van de leeroplossing.

Er zijn diverse typen procesvariabelen zoals in andere evaluatiemodellen (Kraiger, Ford & Salas, 1993; Salas, Milham & Bowers, 2003). In het onderhavige rapport wordt de *constructbenadering* gevolgd. Dit betekent dat gewerkt wordt met generieke, domeinonafhankelijke indicatoren die van toepassing zijn op de leerkenmerken van allerlei soorten leerinterventies (Kraiger, Ford and Salas, 1993; Mayer, 2012). In het evaluatieraamwerk worden deze indicatoren opgesplitst in emotioneel-motivationele leerkenmerken en cognitieve leerkenmerken. Deze zijn

sterk aan elkaar gerelateerd. Emotioneel-motivationale indicatoren hebben invloed op de informatieverwerking (ofwel cognitieve factoren) en leren (bijvoorbeeld Hockey, 1997; Richardsson et al., 2012), en dit leidt tot bepaalde leeruitkomsten. Naast de emotionele-motivationale en cognitieve *leer*kenmerken worden de kenmerken van de te evalueren leeroplossing *ontwerp*kenmerken genoemd.

Ook zijn er bepaalde factoren vanuit de omgeving of organisatie die het leerproces en de leeruitkomsten beïnvloeden. Daarnaast zijn taakkenmerken uiteraard relevant (zie ook bv. het model van Tannenbaum et al., 1993). Deze zijn echter niet meegenomen in het evaluatieraamwerk omdat ze in effectiviteitsonderzoek vaak niet apart gemeten worden. In een experiment zijn de taakkenmerken ook vaak gerelateerd aan de onafhankelijke variabele(n). Daarnaast hebben ze te maken met de leerinhoud (datgene wat geleerd moet worden) en komen ze als zodanig tot uitdrukking in de domein- en taakspecifieke leeruitkomsten (hoe de taak wordt uitgevoerd, welke competenties hiervoor nodig zijn). Hiervoor moeten, afhankelijk van de taak, veelal specifieke maten ontwikkeld worden (bijvoorbeeld toets, competentiebeoordeling) zoals hierboven uitgelegd.

2.4 Indicatoren in het evaluatieraamwerk

2.4.1 *Persoonskenmerken*

De persoonskenmerken zijn het minst veranderbaar en beïnvloeden het leerproces zelf. Ze worden meegenomen naar de leersituatie als basis waarop wordt verder gebouwd en kunnen hiermee ook worden beschouwd als voorspellers voor leren. In de literatuur wordt veelal voor de (stabele) dispositionele persoonskenmerken (Roe, 2002, 2005) het onderscheid gemaakt tussen persoonlijkheid (cf. *Big Five*), cognitieve capaciteiten (vgl. intelligentie) en andere kenmerken als biografische (bijvoorbeeld leeftijd, sexe) of demografische eigenschappen (bijvoorbeeld afkomst). Hier vallen tevens opleiding en ervaring onder. Ook dit zijn stabiele en weinig veranderbare gegevens en moeten daarmee worden beschouwd als input. Omdat persoonskenmerken de individuele basis vormen voor iedere leersituatie, wordt het succes van leerinterventies in belangrijke mate bepaald door de effectiviteit waarmee deze kenmerken in leerinterventies worden meegenomen.

2.4.2 *Leerkenmerken*

Bij het evalueren van leeroplossingen is het van belang om te focussen op die leerkenmerken die middels het ontwerp naar verwachting worden beïnvloed. De keuze in dit evaluatiemodel is gebaseerd op de leerkenmerken die in de literatuur vaak worden genoemd en die tevens in combinatie met elkaar worden onderzocht. Ook maken ze onderdeel uit van veel van bovengenoemde effectiviteitsmodellen. Pintrich en The Groot (1990) bijvoorbeeld hebben de Motivated Strategies for Learning Questionnaire (MSLQ) ontwikkeld die de indicatoren bevat zoals weergegeven in het evaluatiemodel: self-efficacy, motivatie, en zelfsturendheid. De MSLQ wordt vaak gebruikt in effectiviteitsonderzoek omdat deze gevalideerde vragenlijst belangrijke leerkenmerken in samenhang met elkaar op dezelfde manier meet. Ryan & Deci (2000) hebben deze indicatoren ook opgenomen in een model voor evaluatie van leren. Cordova & Lepper (1996) hebben self-efficacy, motivatie en anxiety toegepast in een gecombineerd meetinstrument.

Dit laat zien in welke mate deze leerkenmerken met elkaar verweven zijn. Hieronder worden de vermelde leerkenmerken motivatie, self-efficacy, zelfsturendheid en anxiety gedefinieerd waarbij tevens duidelijk wordt hoezeer ze samenhangen. Het doel van deze beschrijving is niet om de resultaten van een uitputtende literatuurstudie te presenteren maar om de relevantie van deze vaak genoemde leerkenmerken naar voren te halen. Afhankelijk van het doel van de uit te voeren effectiviteitsstudie kan een selectie worden gemaakt of leerkenmerken worden toegevoegd die niet in het model staan indien deze relevant zijn voor het domein of de specifieke onderzoeksvraag.

2.4.2.1 *Emotioneel-motivationale kenmerken*

Motivatie

Motivatie is een breed begrip en bevat veel elementen waarvan is bewezen dat ze een sterke invloed hebben op het leergedrag. Pintrich & De Groot (1990) onderscheiden in de genoemde MSLQ drie motivationele componenten in leren: (1): verwachtingen van de leerlingen, namelijk of ze in staat zijn om de taak uit te voeren.; (2) de waarde die de leerlingen hechten aan het belang van de taak en hun interesse in het leren zelf; (3) het gevoel en emotionele reacties ten opzichte van de taak. Zo kan 'anxiety' (angst, bezorgdheid, stress) een negatieve invloed hebben op het leerproces. Dit bevestigt de nauwe relatie tussen motivatie en emoties.

Self-efficacy

Self-efficacy is gedefinieerd als "individuals' beliefs about their performance capabilities in a particular domain, or the power of believing you can" (Bandura, 1986; Schunk, 1985), in het Nederlands samengevat als 'geloof in eigen kunnen'. Self-efficacy bevat het gevoel van de individu in staat te zijn om bepaalde doelen of taken te behalen door eigen handelen in bepaalde situaties (Schunk, 1985). Het betreft het zelfvertrouwen in eigen cognitieve vaardigheden om te leren of om een taak te vervullen. Self-efficacy wordt sterk bepaald door eerdere, positieve dan wel negatieve, ervaringen ten aanzien van leren en/of presteren in een bepaald domein. Daarmee kan de self-efficacy van een persoon sterk verschillen over verschillende taken of domeinen. Motivatie en self-efficacy worden vaak in samenhang met elkaar beschouwd als belangrijke drijfveren voor het leerproces (bijvoorbeeld Pintrich & De Groot, 1990; Tams, 2008).

Betrokkenheid

Betrokkenheid (cf. 'engagement') ligt heel dicht tegen motivatie aan maar wordt als apart leerkenmerk vermeld omdat engagement bij bepaalde innovatieve leeromgevingen relevant is om te meten (zie bijvoorbeeld Oprins et al., 2013). Betrokkenheid beschrijft de relatie of band tussen persoon en activiteit (Russel, Ainley en Frydenberg, 2005) of 'a person's active involvement in a task or activity' (Reeve, Jang Carrell, Jeon en Barch, 2004). Betrokkenheid is een belangrijk 'meta construct' voor onderwijskundig onderzoek (Fredericks, Blumenfeld en Paris, 2004). Appleton, Christenson, Kim en Reschly (2006) geven een overzicht van de rol van betrokkenheid in opleidingen. Volgens Appleton et al., (2006) valt betrokkenheid uiteen in de volgende constructen: leerling – instructeur relatie (psychologisch), relevantie van materiaal (cognitief), peer support (psychologisch), doelen en toekomstverwachtingen (cognitief), family support (psychologisch), en extrinsieke motivatie (cognitief).

Angst ('anxiety')

"Anxiety" kan worden aangeduid als angsten die optreden bij leerlingen gedurende hun studieproces (Pecoraro, 2006). Het ligt heel dicht tegen self-efficacy aan: bij een lage self-efficacy treedt angst vaker op. Dit kan versterkt worden door eerdere (negatieve) ervaringen in leren of presteren.

Uit onderzoek is gebleken dat er een significante correlatie bestaat tussen een hoog niveau van "anxiety" en een lage (academische) prestatie (Vitasari, Wahab, Othman, Herawan & Sinnadurai, 2010c). Er zijn verschillende vormen van angst gerelateerd aan leren. Een hele bekende is "exam anxiety" (Sansgiry, Monali & Kavita, 2006). Andere bronnen van angsten zijn bijvoorbeeld "language anxiety", "mathematics anxiety", "social anxiety", "family anxiety", "presentation anxiety" en "library anxiety" (Vitasari, Wahab, Herawan, Othman & Sinnadurai, 2011).

Voor "mathematics anxiety" is in dit verband aangetoond dat een hoge mate van "anxiety" leidt tot vermijding van de taak en dat een hoge mate van "mathematics anxiety" een hoge correlatie heeft met slechte prestaties op de wiskundige taken (Erden & Akgul, 2010).

2.4.2.2 *Cognitieve kenmerken*

De cognitieve kenmerken hebben te maken met hogere, en meer bewust controleerbare ("reflectieve") vormen van, informatieverwerking. Ze hebben betrekking op de mate van en het type cognitieve activiteit dat wordt verricht tijdens het leren. De hier vermelde indicatoren zijn grotendeels afgeleid van een metastudie uitgevoerd voor individuele kenmerken en voorspellers voor leren door Richardson et al., (2012).

Zelfsturendheid

Zelfsturend leren ('self-directed learning', gerelateerd aan 'self-regulated learning'; Pintrich & De Groot, 1990) is de kern van veel nieuwe onderwijskundige concepten. Zelfsturend leren is de mate van verantwoordelijkheid en regie die leerlingen nemen in hun leerproces (Fisher & King, 2009; Stubbé & Theunissen, 2008). Het is een breed construct dat diverse facetten bevat, zoals regie nemen, reflectie en samenwerken (Barnard, Lan, To, Paton & Lai, 2009; Fisher & King, 2009; Stubbé & Theunissen, 2008).

Leeractiviteit

Leeractiviteit betreft de overkoepelende term voor de cognitieve activiteiten die de leerling onderneemt om te leren. Dit betreft, bijvoorbeeld de toepassing van diverse leerstrategieën (bijvoorbeeld Barnard, Lan, To, Paton & Lai, 2009) zoals 'rehearsal', 'organization' en 'elaboration', en ook samenwerkende (leer)activiteiten als 'help seeking' and 'peer learning' (Richardson et al., 2012). Dit construct is sterk gerelateerd aan zelfsturend leren onder de aanname dat zelfsturende leerlingen in staat zijn om zelf de meest passende leeractiviteiten te kiezen en deze ook actief toe te passen (Barnard, Lan, To, Paton & Lai, 2009; Stubbé & Theunissen, 2008).

Mentale inspanning

Terwijl leeractiviteit betrekking heeft op de activiteiten die de leerling onderneemt, gaat het bij mentale inspanning om de mate waarin dit moeite kost. De aanname is dat mentale inspanning ('mental effort') de informatieverwerking sterk beïnvloedt (bijvoorbeeld Hockey, 2003) en hiermee leren. De mate van aandacht en concentratie bepaalt hoeveel mentale inspanning in het leren wordt gestopt. Zijlstra (1993) heeft de Rating Scale Mental Effort (RMSE) ontwikkeld die vaak wordt

gebruikt om mentale inspanning tijdens leren en/of presteren te meten. Er bestaat een optimum aan mentale inspanning die geleverd moet worden om te leren of presteren. De mentale inspanning mag niet te hoog worden om nog voldoende (leer)prestaties te kunnen leveren. De complexiteit van de leertaak kan hierop worden aangepast (Salden et al., 2004; Van Merriënboer & Kirschner, 2007).

Beïnvloedende factoren

Er kunnen diverse factoren van invloed zijn op het leerproces die in deze studie niet allemaal zijn uitgewerkt. Zo is er veel onderzoek gedaan naar de invloed van psychosociale dimensies van leeromgevingen op leerprocessen (Dorman, 2003; Seng & Fraser, 2008; Zandvliet & Straker, 2001). Het is aangetoond dat leerlingen beter leren als ze het leerklimaat als positief ervaren. De vragenlijst *What Is Happening In this Class (WIHIC)* is een gevalideerde vragenlijst die de volgende constructen bevat (Dorman, 2003; Seng & Fraser, 2008): 'leerling cohesiveness', 'teacher support', 'involvement', 'investigation', 'task orientation', 'cooperation', en 'equity'. Een leerklimaat wordt beschouwd als veilig als leerlingen alle kansen krijgen en alles durven uitspreken ten aanzien van het leren. De rol van de begeleider of instructeur is hierbij van groot belang. Daarnaast moeten een optimale keuze van taken en middelen dergelijke kansen bieden aan de leerlingen.

2.4.3 *Leeruitkomst*

De uitkomstmaten kunnen op diverse niveaus worden gemeten (zie ook het Kirkpatrick model in Fig. 1). Allereerst kan domeinspecifieke kennis, vaardigheden en houdingen worden gemeten op basis van de beoogde leerdoelen van de leeroplossing. Competenties worden gedefinieerd als de toepassing van het geheel aan kennis, vaardigheden en houdingen in een bepaalde echte (werk)situatie (Roe, 2002; Oprins, 2008). Dit vraagt doorgaans om het meten van transfer of training naar een bepaalde (werk)context. Een noodzakelijke voorwaarde hiervoor is dat er een overeenkomst bestaat tussen elementen van de leersituatie en praktijk (Korteling, Helsdingen & Theunissen, 2013; Korteling & Oprins, 2013). De prestatie wordt beschouwd als het resultaat van competent handelen. Prestaties kunnen worden uitgedrukt in termen van bijv. accuratesse, aantal gemaakte fouten, aantal goede handelingen of de kwaliteit daarvan, snelheid etc. Doorgaans is prestatie ook domeinspecifiek hoewel er pogingen zijn gedaan om taakprestatie ook generiek te meten door te scoren op de dimensies van een hiërarchisch taakmodel, zoals dat van Rasmussen (1986) en Michon (1985). Prestaties worden grotendeels bepaald door competenties. Daarnaast zijn individuele-, of teamprestaties het gevolg van omgevingsfactoren of persoonskenmerken (Roe, 2002). Ten behoeve van een maximaal valide design verdient het daarom de voorkeur om ze in combinatie met elkaar te meten.

De mate waarin een bepaalde leeruitkomst echter wordt bereikt, hangt ook af van het beginniveau van een lerende dat daarom altijd moet worden meegenomen in effectiviteitsonderzoek. Wat voor de één effectief is kan bij de ander niet aanslaan. Dat komt doordat leren noodzakelijkerwijs altijd voortbouwt op *bestaande* eigenschappen, i.e. kennis, vaardigheden, attitudes, competenties (Korteling et al., 2012). Piaget (1950) noemde dit proces *assimilation* (fitting into mental framework) en *accommodation* (reframing). Op basis van bestaande kennis en vaardigheden bouwen individuen aan nieuwe kennis en vaardigheden vanuit nieuwe ervaringen. Dit betekent dat leren alleen effectief kan plaatsvinden als de nieuwe informatie past bij de bestaande kenmerken van de lerende (adaptief, op maat) zodat de

lerende de stof makkelijk in zijn framework kan assimileren. Als dat niet goed mogelijk is, bijvoorbeeld omdat de stof zodanig afwijkt van het bestaande kennis en vaardigheden, dan is accommodatie nodig. Dat is moeilijk en vereist inspanning omdat de lerende de nieuwe stof actief moet proberen eigen te maken door zijn bestaande raamwerk van kennis en vaardigheden aan te passen (Büchel, Coull & Friston, 1999).

3 Onderzoeksmethoden (design)

Het generiek evaluatieraamwerk (zie hoofdstuk 2) kan als basis worden gebruikt om (experimentele) onderzoeken op te zetten voor het meten van de effectiviteit van opleiding, training of oefening. Het biedt een denkkader dat structuur biedt bij het opzetten en uitvoeren van evaluatieonderzoek. Het kan er voor zorgen dat er geen belangrijke aspecten van het onderzoek worden gemist. De te volgen stappen zijn als volgt:



Figuur 5 Algemeen stappenplan onderzoeksdesign.

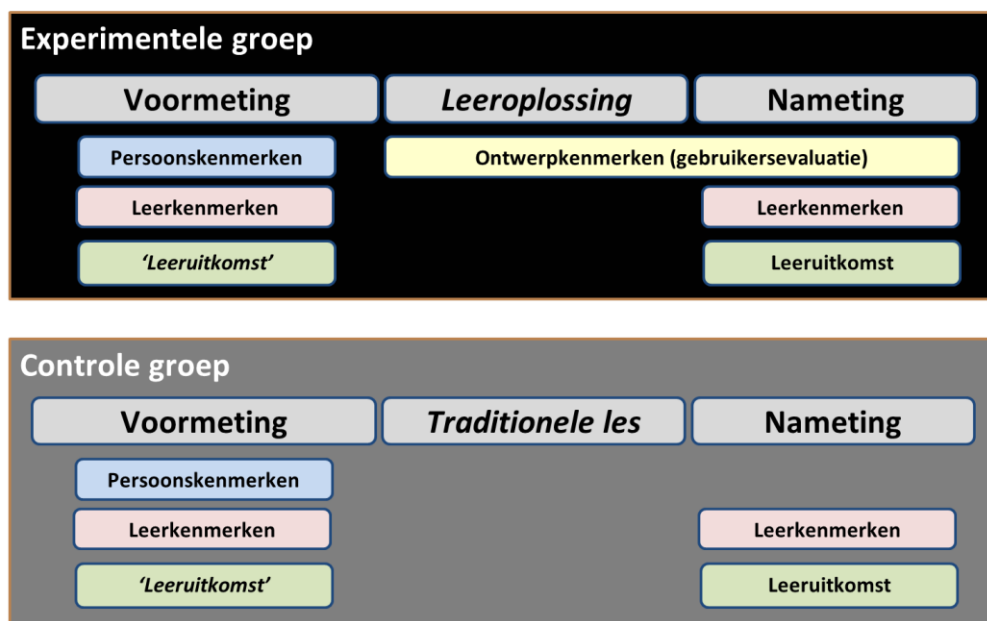
Het onderhavige hoofdstuk werkt dit algemene stappenplan verder uit. Het beperkt zich tot de generieke principes die hier gevolgd dienen te worden zonder dit reeds toe te passen op een specifiek domein of training.

3.1 Experimenteel onderzoeksdesign

Om een goede effectiviteitsstudie van opleiding en training of oefening te kunnen uitvoeren, moet een adequaat experimenteel design gehanteerd worden. In de praktijk bij het evalueren van trainingen is hier over het algemeen te weinig sprake van (Caro, 1977; Salas et al., 2003; Akl et al., 2010; Sitzman, 2011).

3.1.1 *Experimental-versus-control-group method*

In de ideale situatie is er een experimentele groep met de nieuwe leeroplossing (bijvoorbeeld een trainingssimulator) en een controlegroep met de conventionele of standaard opleiding, training of oefening. Dit wordt ook wel de "*Experimental-versus-control-group method*" genoemd (Campbell & Stanley, 1963; Korteling et al., 2013). Hierbij wordt bij beide groepen een voormeting en een nameting uitgevoerd. Een controlegroep die helemaal geen training volgt is niet geschikt omdat er dan geen valide conclusies kunnen worden getrokken over de effectiviteit van de specifieke leerinterventie ten opzichte van de vergelijkingssituatie (Ackerman et al., 2010). In lijn met het evaluatieraamwerk weergegeven in Figuur 4 zou de ideale experimentele onderzoeksopzet er dan als volgt uit zien:



Figuur 6 Experimentele opzet op basis van evaluatieraamwerk.

Figuur 6 geeft aan dat de indicatoren zoals weergegeven in het evaluatieraamwerk zowel bij de experimentele als bij de controlegroep worden gemeten en (behalve bij grote aselekt samengestelde proefgroepen) zowel in de voormeting als de nameting. De indicatoren die worden gemeten moeten idealiter zoveel mogelijk hetzelfde blijven. De ontwerpkenmerken worden doorgaans uitsluitend bij de experimentele groep bevraagd.

Echter, in de praktijk kan niet altijd gewerkt worden met een controlegroep bijvoorbeeld omdat de training geheel nieuw is of dat dit risicovolle situaties oplevert (Campbell & Stanley, 1963; Salas et al., 2003). Hetzelfde geldt voor de voormeting. Bovendien hoeft niet perse alles gemeten te worden; dit hangt van de situatie af en van de onderzoeksvraag. De volgende quasi-experimentele methoden worden onderscheiden:

3.1.2 *Self-control- transfer-method*

Bij deze methode is de experimentele groep ook gelijk de controlegroep. Een groep leerlingen die op conventionele wijze (regulier) wordt getraind, krijgt op een gegeven moment voor een bepaalde tijd een leerinterventie (bijvoorbeeld simulatortraining of een trainingssgame). Het gaat hierbij om het prestatieverschil vlak voor en vlak na de leerinterventie. Dit verschil kan worden toegeschreven aan de simulatortraining. Op basis van dit soort data is het echter niet mogelijk om harde conclusies te trekken over transfer of training of over de efficiëntie ervan omdat er geen vergelijking wordt gemaakt met een controlegroep die op de conventionele manier (en bijvoorbeeld met real-life middelen) wordt getraind.

3.1.3 *Pre-existing-control-transfer method*

Er zijn studies mogelijk waarin een controlegroep die op conventionele wijze wordt getraind niet nodig zou hoeven te zijn. Bijvoorbeeld: als een bepaalde leerinterventie in een bestaand trainingsprogramma wordt geïntroduceerd. Prestaties van leerlingen van een eerder programma op een eerder bepaalde criteriumtaak kan dan worden vergeleken met prestaties van de experimentele groep waarop de nieuwe leerinterventie werd toegepast.

3.1.4 *Uncontrolled-transfer method*

Er zijn ook gevallen mogelijk waarin geen controlegroep mogelijk is. Dat kan zo zijn als veiligheid en risicobeheersing doorslaggevend zijn, bijvoorbeeld een noodlanding maken met een vliegtuig. In dat geval kan de trainingseffectiviteit van de interventie (in dit geval meestal met gebruikmaking van simulatiemiddelen) worden vastgesteld door de prestatie te meten de eerste keer dat de leerlingen de geleerde taak in de praktijk uitvoeren (bv een echte landing maken in een vliegtuig). Dit wordt *first shot performance* genoemd. Hierbij kunnen geen definitieve conclusies worden getrokken over de leereffecten van de leerinterventie omdat die bij gebrek aan voormeting en controlegroep niet te bepalen zijn (Caro, 1977).

3.1.5 *Quasi-transfer-of-training method*

De the Quasi-Transfer-of-Training method (QToT) wordt veelal gebruikt om trainingssimulatoren op een (kosten) efficiënte te valideren. Het belangrijke verschil tussen de experimental-versus-control-group methode and the QToT method is that bij de eerste door de controlegroep in de praktijksituatie wordt getraind, terwijl dat bij de QToT methode niet het geval is. Bij de QToT method krijgt de experimentele groep een leerinterventie, bijvoorbeeld een educatieve game, waarvan de effectiviteit wordt geëvalueerd. Ter vergelijking wordt de controle groep getraind op een high fidelity simulator, met hoge (lieft bewezen of aantoonbare) trainingseffectiviteit. Uiteindelijk worden beide groepen geëvalueerd op criteriumtaken in deze high-fidelity simulator. Het prestatie verschil toont de leer resultaten van het te beoordelen system ten opzichte van de high fidelity simulator. De waarde van deze informatie neemt toe naarmate er meer bekend is over de effectiviteit van het high-fidelity systeem.

3.1.6 *Backward-transfer method*

Ook deze methode is toegesneden op het evalueren van simulatiemiddelen die in trainingen worden gebruikt. Bij de backward transfer method moet een functionaris of team die goed zijn getraind en al een goede (operationele) prestatie hebben laten zien, een prestatie leveren in bijvoorbeeld de simulator or serious game. Als dat ongeveer net zo goed gaat als in de praktijk dan spreekt men van "backward transfer". De aanname hier is dat er bij leerlingen die met het simulatiemiddel worden getraind ook transfer of training in de andere richting ("forward transfer") zal plaatsvinden. Op basis van dit soort data is het niet mogelijk om harde conclusies te trekken over de transfer of effectiviteit van het middel omdat er geen vergelijking wordt gemaakt met een controlegroep die op de conventionele manier (en bijvoorbeeld met real-life middelen) wordt getraind.

3.1.7 *Performance-improvement method*

Bij de Performance-improvement method wordt iedere trainingssessie de prestatie van leerlingen bemeten. Een essentieel uitgangspunt van een effectief trainingsprogramma is prestatieverbetering door leerlingen over de successievelijke trainingssessies. Als dit niet, of weinig, plaatsvindt, zou je ook weinig leerresultaten mogen verwachten voor de werkelijke taak. Echter, bij de behandeling van het pyramidemodel van Kirkpatrick (1976, 2006) lieten we al zien dat het optreden van leereffecten tijdens een training niet noodzakelijk dat wat geleerd is ook relevant is voor de praktijk. Deze aanname is alleen plausibel als de trainingsomgeving een sterke overeenkomst (fysisch, functioneel en psychologisch) met de werkelijke taakomgeving (Korteling, Helsdingen & Theunissen, 2013).

3.2 **Operationaliseren van indicatoren**

In het evaluatieraamwerk zijn de indicatoren opgenomen die bemeten moeten worden. Deze indicatoren, ook wel constructen genoemd, moeten geoperationaliseerd worden tot meetinstrumenten waarvan vervolgens de bruikbaarheid, betrouwbaarheid en validiteit wordt onderzocht.

Een construct is een 'onstoffelijk voortbrengsel van de menselijke geest' (Van Dale Woordenboek), en dat zijn dingen als 'motivatie', 'tevredenheid' of 'capabel zijn'. Vaak gaat het om complexe psychosociale fenomenen waarvan iedereen een beeld heeft, maar die in de praktijk nuances en individuele interpretaties kennen. Constructen zijn per definitie niet direct te meten omdat ze abstract onstoffelijk zijn. Daarom kun je er (indirect) naar vragen, maar omdat iedereen er een eigen interpretatie van heeft, werkt dat niet optimaal. De gangbare methodologische benadering binnen sociaalwetenschappelijk onderzoek is daarom zo'n construct eerst zo precies mogelijk te definiëren en vervolgens te analyseren en operationaliseren naar een set van meer eenduidig interpreteerbare deel-aspecten (zie bijvoorbeeld Baarda & De Goede, 1995; Swanborn, 1987), zoals een set vragen of stellingen in een vragenlijst.

Zo volstaat een rechtstreekse vraag of iemand gelukkig is meestal niet omdat zo'n vraag teveel sociaal wenselijke antwoorden oplevert en omdat verschillende mensen zeer verschillende opvattingen hebben over wat geluk voor hen betekent. De constructaanpak impliceert in dit voorbeeld dat het thema 'geluk' met verschillende meer concrete en eenduidig te interpreteren vragen benaderd wordt om uiteindelijk uit al die facetten een totaalbeeld te verkrijgen hoe gelukkig iemand is. Dat kunnen bijvoorbeeld vragen zijn als 'lacht u veel', 'wordt u blij wakker', 'bent u gelukkig met uw leven', 'vinden uw vrienden u gelukkig' etc. Eén construct kan uitwaaiëren naar verschillende aan elkaar gerelateerde (sub) constructen. Dat kan zowel bij het definiëren van de constructen gebeuren als meer iteratief bij het opstellen van de vragenlijsten. Zo zouden vragen over 'geluk' twee sub-constructen kunnen vormen, bijvoorbeeld 'gelukkig volgens anderen', 'gelukkig volgens zelf'. Gebruikelijke methoden bij het bepalen van de te meten constructen voor specifieke onderzoeksdoeleinden zijn: literatuuronderzoek naar reeds bestaande constructen, en het inschakelen van subject matter experts.

3.3 Meetinstrumenten

3.3.1 *Het meten van constructen*

Generieke constructen kunnen worden geoperationaliseerd tot generieke meetinstrumenten die in een breed scala aan (effectiviteits)studies kunnen worden gebruikt. Generieke meetinstrumenten, die in diverse projecten kunnen worden afgenomen, maken grotere aantallen proefpersonen mogelijk, waardoor validatie ervan op langere termijn haalbaar wordt. Daarnaast bestaat er behoefte aan standaardisatie van de gehanteerde meetinstrumenten. Voor het meten van bepaalde constructen zijn veelal diverse vragenlijsten in omloop, waarvan de kwaliteit verschillend is. Het streven is dan om deze diversiteit terug te brengen tot een set standaard vragenlijsten. Daarnaast moeten deze gericht zijn op het meten van generieke constructen die relevant zijn voor validatie van leerinnovaties in diverse domeinen. In elke studie dient een keuze worden gemaakt welke constructen op welke manier gemeten moeten worden. Ten behoeve van de eenduidigheid en kwaliteit van het meetinstrument is het ontwerp van meetinstrumenten, zoals vragenlijsten (bijvoorbeeld schaalverdeling, type vragen) zoveel mogelijk hetzelfde. Een hoge score op generieke vragenlijsten, hoe betrouwbaar en construct-valide deze ook mogen zijn, betekent nog niet dat het uiteindelijke doel (uitkomst, resultaat) van een training wordt behaald, namelijk transfer op langere termijn van competenties naar de praktijk on-the-job. Hoge scores op generieke vragenlijsten geven wel aan dat de kans op positieve trainingsresultaten wordt vergroot en geven vooral ook inzicht in de sterke en zwakkere kanten van een training.

Veel constructen of indicatoren kunnen ook op een andere manier dan door middel van een vragenlijst worden gemeten. Grofweg bestaan er drie soorten meetinstrumenten die kunnen worden gebruikt voor het meten van leereffecten (Freeman, Stacy & Olivares, 2009). Dit zijn: 1) vragenlijsten gericht op (subjectieve) vaak introspectieve evaluatie door trainees, instructeurs of training experts, 2) methoden gebaseerd op objectieve (geïncorporeerde, geautomatiseerde) meting van (leer)prestaties en andere factoren (bijvoorbeeld leergedrag, leerstrategieën op basis van data logging), en 3) observatie op basis van gestructureerde registratie van data (semi-objectief of intersubjectief). Afhankelijk van het doel van het onderzoek (bijvoorbeeld objectieve prestatie meting vs. meningen van gebruikers achterhalen) en de praktische haalbaarheid (bv. beschikbaarheid over digitale data, tijdsinvestering) wordt doorgaans een keuze gemaakt voor bepaalde type meetinstrumenten. Bij voorkeur is dit een combinatie (triangulatie) omdat meerdere soorten metingen de onderzoeksresultaten betrouwbaarder maakt. In het onderstaande zullen we deze drie hoofdcategorieën kort bespreken,

3.3.2 *Subjectieve metingen (vragenlijsten, zelf-evaluatie en introspectie)*

Subjectieve metingen, bijvoorbeeld zelf-observaties of metingen met behulp van zelf-evaluatie vragenlijsten of andere methoden waarbij mensen (eigen) gedrag of (leer)ervaringen monitoren, interpreteren of daarover moeten reflecteren hebben belangrijke praktische voordelen. Ze zijn snel te realiseren en goed bij grote aantallen proefpersonen te gebruiken. Daarnaast zijn ze in principe eenvoudig en, indien gewenst, flexibel aanpasbaar aan, of specifiek te richten op, vraagstellingen, taken en doelgroepen.

De algemene aanname bij zelf evaluatie instrumenten, zoals vragenlijsten, is dat mensen kennis en inzicht hebben in zichzelf. Dat betekent dat ze zich bewust zijn van hun opvattingen, mentale gesteldheden en (daarmee) de oorzaken van hun gedrag, zoals daar ook van wordt uitgegaan in bijvoorbeeld theorieën over rationele besluitvorming (e.g. Ajzen, 1991). Mensen worden dan bevraagd over attitudes, verwachtingen en motieven. Verwacht wordt dat mensen bewust, reflectief en introspectief gedachten, gevoelens en gedrag achterhalen of beschrijven en met elkaar in verband brengen. Echter, en nog steeds in navolging van het bekende gedachtengoed van Sigmund Freud, zijn de meeste (recente) psychologische modellen over menselijk gedrag (bijvoorbeeld De Houwer, 2006; Greenwald & Banaji, 1995; Stacy, Ames, Sussman, & Dent, 1996; Strack and Deutsch, 2004; Wiers et al., 2007) het er over eens dat gedrag het resultaat is van een complex samenspel van niet alleen reflectieve (bewuste, gecontroleerde) processen, maar vooral ook (veel meer nog) impulsieve (onbewuste, automatische) processen.

Op basis van de diverse genoemde voordelen van (zelf-)evaluatie vragenlijsten zijn de meeste competentiemetingen gebaseerd op zelf-rapportage (of observatie rapporten). Nosek, Hawkins, & Frazier (2011) hebben met betrekking tot dit type, op introspectie of zelf-evaluatie gebaseerde, instrumenten verschillende factoren geïdentificeerd die de waarde ervan beperken. Kort samengevat zijn de belangrijkste daarvan: een lage motivatie om persoonlijke ideeën, gedachten of gevoelens te rapporteren of een beperkt bewustzijn (awareness) daarvan, dan wel beperkingen in het vermogen om dergelijke (bewuste) mentale gesteldheden te verbaliseren.

- Reacties of associaties ten opzichte van bijvoorbeeld een leerinterventie kunnen zo bijvoorbeeld foutief, of helemaal niet, worden gerapporteerd. Soms komt dit simpelweg omdat het individu daarin geen zin in heeft of omdat dit strijdig is met zijn eigen belang. Dit laatste is in het bijzonder relevant als er persoonlijke of politieke belangen spelen, bijvoorbeeld als leeruitkomsten relevant zijn voor de loopbaan of toekomstmogelijkheden van individuen.
- Ten tweede zit er een belangrijke impliciete of associatieve component in veel cognities, gevoelens of attitudes, waar men zich niet bewust van is, waardoor het simpelweg onmogelijk is om hierover expliciet te rapporteren.
- Tot slot, zelfs al zouden mensen zich bewust kunnen zijn van, en voldoende inzicht hebben in, de vele aspecten van hun mentale gesteldheid en al zouden zij gemotiveerd zijn om hier eerlijk en oprecht verslag van te doen, dan *nog* is het vaak heel moeilijk om dit goed in verbale concepten te vatten, of daarnaar te vertalen, dan wel kiezen uit een serie alternatieve beschrijvingen. Dit komt omdat het vocabulaire of het conceptuele kader dat hiervoor nodig onvoldoende is of ontbreekt (Korteling, Pieters & Jetten, 2014).

Onbewuste, impliciete cognities, aandachts-biases etc. hebben een zeer sterke invloed op het gedrag (Greenwald & Banaji, 1995; Pauls & Crost, 2005) en zouden daarom wel gemeten moeten kunnen worden. De moeite om aan de beperkingen van zelf-observatie maten en vragenlijsten tegemoet te komen hebben geleid tot de veelgehoorde opvatting, vooral onder research psychologen, dat het verzamelen van dergelijke subjectieve data “zacht” is en weinig wetenschappelijk gehalte heeft (e.g., Annett, 2002). Het hangt natuurlijk van de aard en doelstelling van het

onderzoek af of dat zo is. Als die doelstelling gericht is op de beleving, ervaring of interpretatie van interventies door proefpersonen, dan is gebruik van vragenlijsten goed verantwoord. Voorbeelden hiervan zijn het meten van bewegingsziekte of het vaststellen van meningen over de training. Als het gaat om het zoeken naar (theoretische) verklaringen en mechanismen, dan kunnen vragenlijsten bij die zoektocht helpen, maar dienen ze doorgaans te worden aangevuld met metingen (gedrag, prestatie, fysiologisch) gericht op onderliggende, onbewuste en impliciete mechanismen.

3.3.3 *Objectieve (geïnstrumenteerde) metingen*

Op basis van technologische vooruitgang wordt het objectief, geïnstrumenteerd meten van moeilijk (introspectief) observeerbare emotionele, mentale of cognitieve gesteldheden en attitudes geleidelijk aan steeds beter mogelijk (Freeman, Stacy & Olivares, 2009). Hierbij kan worden gedacht aan het (direct) meten van angst, concentratie, inzicht of verslavingsproblemen met behulp van psychofysiologische technologieën en meetmethoden, zoals eye tracking, EEG, fMRI of impliciete tests (e.g. Bowden, Jung-Beeman, Fleck & Kounios, 2005; Brouwer & Korteling, 2012; Korteling et al., 2013, Korteling, Jetten & Pieters, 2014). Ook is er op basis van data logging technieken steeds meer mogelijk op het gebied van geautomatiseerd objectief meten van gedrags-indicatoren die informatie en inzicht kunnen geven over het “leergedrag” van leerlingen. Het valt buiten de scope van de huidige studie om specifiek op de technische mogelijkheden en beperkingen van deze nieuwe meetmethoden dieper in te gaan.

Objectieve meting van trainingseffecten met behulp van geautomatiseerde meetinstrumenten kent echter ook diverse beperkingen. Dit soort metingen stellen in principe altijd hoge eisen aan de controle van de condities waaronder wordt gemeten. Daarnaast moeten de prestatiematen ecologisch valide zijn. Dat wil zeggen dat ze representatief moeten zijn voor de taken en onderliggende processen en/of competenties waarover het gaat teneinde werkelijk relevante informatie te geven over het leren en presteren ten aanzien van het werkelijke (on the job) gedrag. Het is daarbij dan goed als deze prestatiematen duidelijke prestatiecriteria hebben die zijn afgeleid van de leerdoelen, zodat ze direct inzicht verschaffen in de mate waarin leerdoelen al of niet worden gehaald. Metingen moeten netjes, en op gestandaardiseerde wijze worden uitgevoerd waarbij de omstandigheden zoveel mogelijk gelijk of vergelijkbaar worden gehouden. Het zuiver en gestandaardiseerd meten is soms vrij lastig omdat ze veelal in praktische (dynamische) trainingsomstandigheden op scholen en trainingslocaties moeten worden uitgevoerd. Volgens Korteling & Oprins (2013) kan dit tot grofweg een drietal problemen leiden:

- Controlegebrek: zuivere en gestandaardiseerde metingen worden verstoord door rigide trainingsschema's en procedures, gebrek aan controle over gebeurtenissen, te weinig beschikbare leerlingen of te weinig toegang tot benodigde praktijksystemen of materiaal waarmee wordt getraind (Cohn et al., 2009; Boldovici, Bessemer & Bolton, 2002). Dit kan zeer nadelig uitpakken voor de betrouwbaarheid (consistentie) van de metingen (zie par. 3.5).
- Meetproblemen: het is vaak lastig te achterhalen wat en hoeveel er nu precies is geleerd met betrekking tot de werkelijke taak of baan waarvoor de training is bedoeld omdat normale taakomstandigheden waarin de (gestandaardiseerde, geautomatiseerde) metingen dan moeten plaatsvinden daarvoor ongeschikt zijn.

- Beperkte beschikbaarheid van controle condities: effectmetingen worden veelal meestal gedaan nadat er een nieuwe interventie ontwikkeld, en vaak ook in het curriculum geïmplementeerd, is. Het is dan lastig om de vergelijking met de oude (controle) conditie te maken (Cohn, et al., 2009). In veel andere gevallen zijn middelen (oefenmaterieel, operationele systemen) gewoon niet beschikbaar om empirische evaluatie van trainingseffecten mogelijk te maken.

Door dit soort factoren zijn evaluatiestudies, waarin processen en effecten objectief en geautomatiseerd worden gemeten, relatief moeilijk realiseerbaar met veelal een wat beperktere scope of reikwijdte. Alleen door specifiek voor een evaluatie-experiment een specifiek trainingsprogramma met representatieve proefpersonen en realistisch taakomstandigheden te creëren kan aan bovengenoemde eisen en bezwaren goed tegemoet worden gekomen. (Korteling & Oprins, 2013). Dit is met name voor objectieve geïnstrumenteerde metingen doorgaans extra van belang omdat ze technologisch moeilijker te realiseren zijn en (mede daardoor) minder flexibel zijn dan in designs gebaseerd op subjectieve metingen.. Daarnaast houden leerlingen, docenten of andere proefpersonen bij subjectieve evaluaties doorgaans wel rekening met (variërende) omstandigheden en contextuele factoren. Geautomatiseerde meetinstrumenten leveren daarentegen in principe absolute data zonder daarin contextuele factoren daarin te verdisconteren. Die moeten dan apart worden meegenomen. Dat is extra lastig omdat deze factoren vooraf vaak nog niet goed bekend zijn. Met dit soort onzekerheden moet juist bij dynamische teams goed rekening worden gehouden.

3.3.4 *Semi-objectieve, intersubjectieve metingen, observatie checklists*

Niettemin het bovenstaande is het belangrijk om werkelijk inzicht te krijgen in de kwaliteit en voor en nadelen van nieuwe leerinterventies. Daarom kunnen de bovengenoemde subjectieve en objectieve metingen worden aangevuld met bijvoorbeeld gestructureerde metingen (bijvoorbeeld met observatie checklists die worden ingevuld door experts, instructeurs of studenten). Bij dit soort beoordelingen wordt getracht de subjectiviteit van de individuele beoordelaar zoveel mogelijk uit te schakelen. Het wordt daarom ook wel semi-objectief of intersubjectief genoemd (bijvoorbeeld Freeman, Stacy & Olivares, 2009). Voorbeelden hiervan zijn gestructureerde checklists of beoordelingsschalen. Deze bevatten welomschreven definities van te evalueren/scoren aspecten van een interventie of leermiddel of van concepten. Hierbij wordt tevens zo precies mogelijk gedefinieerd op welke wijze observaties moeten worden uitgevoerd, gecategoriseerd, gescoord en verwerkt. De beoordelingen moeten vanzelfsprekend gerelateerd en geïnterpreteerd worden met betrekking tot de context van het de leerdoelen, het trainingsprogramma en de competenties die moeten worden getraind.

Specifiek ten aanzien van de ontwerpeigenschappen en gebruik van een bepaalde leeroplossing kan een (onafhankelijke, onpartijdige) expertevaluatie worden uitgevoerd. Dit houdt in dat experts op basis van een gestructureerde checklist (om subjectiviteit zoveel mogelijk uit te sluiten) een leeroplossing systematisch op diverse aspecten, ontwerpeigenschappen en wijze van gebruik. De beoordeling van deze aspecten van de leerinterventie vindt dan zoveel mogelijk plaats op basis van objectief vastgestelde (didactische) criteria. Een voorbeeld hiervan is de TNO CEES 3.0 methode (*TNO Checklist for Evaluation of Educational Simulations*) die door TNO is ontwikkeld om op een snelle en efficiënte manier zoveel mogelijk informatie te verkrijgen over de kwaliteit van (de inzet van) trainingssimulatoren

(Emmerik & Korteling, 2013). Een dergelijke methode staat in principe los van experimenteel onderzoek en kan als extra of afzonderlijke methode worden toegepast om meer inzicht te krijgen in onderliggende factoren die de effectiviteit van leeroplossingen bepalen.

3.4 Keuze van meetinstrumenten

Trainingsuitkomsten worden bepaald door persoons- en leerkenmerken van de leerling en de ontwerp kenmerken van de leerinterventie (inclusief de gebruikte leermiddelen). De voorgaande paragraaf beargumenteerde dat de diverse typen meetinstrumenten allemaal hun voordelen en beperkingen hebben. Om hier vat op te krijgen en om inzicht te krijgen in onderliggende processen kan het, indien praktisch haalbaar, raadzaam zijn om gebruik te maken van een *combinatie* van verschillende soorten meetinstrumenten. Zo kan bijvoorbeeld in een evaluatiestudie van een leerinterventie een vergelijking worden gemaakt van een aantal harde en objectieve prestatie indicatoren voorafgaand en na de leerinterventie. Daarnaast kan in dezelfde studie een beeld worden gekregen van mogelijke onderliggende factoren door vragenlijsten af te nemen bij leerlingen gericht op leerprocessen en hun ervaringen. Tot slot kunnen de resultaten hiervan nader worden geanalyseerd of verklaard op basis van gegevens verkregen met semi-objectieve gestructureerde checklists of scoringslijsten. Deze geven inzicht in (evidence-based) kwaliteitskenmerken van gebruikte leermiddelen of leerinterventies.

3.5 Realiseerbaarheid, bruikbaarheid, betrouwbaarheid en validiteit

Freeman, Stacy & Olivares, 2009) formuleren op de eerste plaats twee criteria voor de kwaliteit van maten en meetinstrumenten voor menselijk gedrag, namelijk *realiseerbaarheid of haalbaarheid* (uitvoerbaarheid, “feasibility”) en *bruikbaarheid/-toepasbaarheid of functionaliteit* (“utility”, “functionality”). De klassieke testtheorie stelt daarnaast bepaalde noodzakelijke eisen aan de psychometrische kwaliteit van een meetinstrument (Drenth & Sijtsma, 1990; Guion, 1998; Roe & Daniëls, 1994). Psychometrische kwaliteit bestaat uit *betrouwbaarheid* en *validiteit*. Een betrouwbaar meetinstrument zal in een gegeven situatie steeds dezelfde uitkomsten geven. Een valide meetinstrument meet daadwerkelijk datgene wat je wilt weten (Drenth & Sijtsma, 1990). Betrouwbaarheid en validiteit zijn niet slechts “nice to have”, maar, evenals realiseerbaarheid en bruikbaarheid, noodzakelijke voorwaarden voor een goed meetinstrument. Bij de ontwikkeling, introductie en feitelijke toepassing van meetinstrumenten dient dan ook met alle vier aspecten rekening te worden gehouden.

3.5.1 Realiseerbaarheid

De Realiseerbaarheid of uitvoerbaarheid (haalbaarheid, feasibility) van een meetinstrument betreft of het überhaupt gemaakt kan worden en wel op een efficiënte manier (Freeman et al., 2009). Sommige maten zijn bijvoorbeeld onderworpen aan regels die er zijn om de privacy of gezondheid van individuen te beschermen. De realiseerbaarheid wordt natuurlijk ook bepaald door de kosten van het ontwikkelen van maten, afname, administratie en het gebruik ervan. De prijs van metingen waarbij observatoren of analisten nodig zijn kan snel oplopen vanwege de hoge kosten die gemoeid zijn met het inhuren en trainen van experts teneinde betrouwbare metingen te bewerkstelligen. Automatisering of computer-ondersteuning kan hierbij van dienst zijn.

3.5.2 *Bruikbaarheid*

Naast realiseerbaarheid wordt in dit verband vaak ook praktische bruikbaarheid genoemd als onderdeel van de psychometrische kwaliteit (Roe & Daniëls, 1994). De bruikbaarheid van een maat wordt bepaald door of het zijn doel kan dienen. Dit betreft o.a. de mate waarin de gebruikers met het meetinstrument kunnen en willen werken. Maten die voor formatieve evaluaties in leertrajecten worden gebruikt moeten op zijn minst betekenisvol en begrijpelijk zijn voor leerlingen of docenten om effectief te kunnen zijn, bijvoorbeeld als middel voor feedback. Om dit te bereiken moeten ontwikkelaars van maten prestaties in het betreffende domein bestuderen door middel van observatie, interviews, en document analyse. Ook voor stafmedewerkers is het noodzakelijk om het instrument goed te begrijpen om er mee te kunnen werken en om er consequenties uit te trekken, i.e. om het te kunnen laten functioneren (Freeman et al., 2009). Messnick (1995) noemde dit aspect "consequential validity". Voor de ontwikkeling van een vragenlijst die breed toegepast moet kunnen worden, is dit ook een belangrijke eis die mede het ontwerp van het meetinstrument bepaalt.

3.5.3 *Betrouwbaarheid*

Een meetinstrument kan alléén valide zijn (zie laatste aspect) als het ook betrouwbaar, of consistent, is. Wanneer bij meting van hetzelfde onder gelijke omstandigheden de uitkomst immers steeds verschillend blijkt te zijn, dan meet het instrument blijkbaar andere dingen dan het specifiek zou *moeten* meten. Doorgaans wordt dan ook in een relatief vroeg stadium (voor de validiteit) de betrouwbaarheid van een meetinstrument bepaald. Dit kan op diverse manieren. Heel gebruikelijk is het bepalen van de interne consistentie ofwel de homogeniteit: de mate waarin onderliggende items hetzelfde construct meten, weergegeven door Cronbach's Alpha (Drenth & Sijtsma, 1990). Een mogelijk nadeel van deze methode om de betrouwbaarheid van bijvoorbeeld een vragenlijst te verhogen is dat, kan met name voor praktische toepassingen ontstaan. Door de "versmalling" die voortvloeit uit het focussen van de meting op iets enkelvoudigs kun je aspecten missen die, bij praktische vragen mogelijk relevant kunnen zijn maar de "essentie" van een bepaald construct niet raken (Gipps, 1994). Andere veelgebruikte methoden zijn test-hertest betrouwbaarheid (herhaalde meting van dezelfde test), split-half betrouwbaarheid (test splitsen in twee delen), of de interbeoordelaars-betrouwbaarheid (overeenkomst tussen twee beoordelaars).

3.5.4 *Validiteit*

Ook voor validiteit zijn er diverse methodes beschikbaar. Criteriumvaliditeit heeft betrekking op de voorspellende waarde van een test of meetinstrument (Drenth & Sijtsma, 1990; Swanborn, 1987). Bijzonder relevant voor vragenlijstontwikkeling is construct validiteit: de mate waarin het beoogde construct ook daadwerkelijk wordt gemeten (Swanborn, 1987; Drenth & Sijtsma, 1990; Roe & Daniëls, 1994). Omdat constructen onstoffelijk zijn is het vrij lastig om hier achter te komen; veel onderzoek naar construct validiteit heeft daarom paradoxale en tautologische trekjes. Voor het bepalen van construct validiteit wordt doorgaans de relatie gelegd met een ander (extern) criterium, bijvoorbeeld een vergelijkbare test of instrument dat hetzelfde beoogt te meten. De voor de hand liggende vraag is dan natuurlijk hoe groot de constructvaliditeit van dit externe criterium is? Ook factoranalyses kunnen inzicht bieden in de betekenis van de constructen (Drenth & Sijtsma, 1990; Roe & Daniëls, 1994). Deze geven inzicht in de structuur van de vragenlijst, namelijk in hoeverre bepaalde items die behoren tot een bepaald construct

voldoende met elkaar samenhangen en onafhankelijk zijn van andere items. Een belangrijk resultaat van factoranalyses is dat het aantal variabelen teruggebracht kan worden (datareductie). De face validiteit heeft ook als doel het bepalen van de mate waarin het meetinstrument een bepaald construct weerspiegelt. Dit is echter meer gebaseerd op persoonlijke indruk en gevoel en kan daarom ook worden gezien als de "geloofwaardigheid" van het meetinstrument. Dit kan in de praktijk belangrijk zijn om docenten en leerlingen te overtuigen van de kwaliteit en het nut van een instrument zodat er adequaat mee wordt omgegaan en het effectief wordt gebruikt.

4 Meten van de effectiviteit van dynamische teams

4.1 Inleiding dynamische teams

Militaire teams moeten zich kunnen handhaven in een dynamische en complexe (internationale) omgeving. Een vaste teamsamenstelling is daarbij, zoals eerder beargumenteerd, geen vanzelfsprekendheid meer. We zien dan ook een sterke toename van voor specifieke gelegenheden samengestelde (ad hoc) teams. Deze zijn *tailor-made* voor de opdracht en veelal geïncorporeerd in een genetwerkte omgeving. Hierbij varieert de teamsamenstelling afhankelijk van de opdracht, die veelal een concrete en korte-termijn doel heeft. De voorbereiding van de militair op het flexibel werken in dynamische teams is niet eenvoudig. De training van deze militairen/teams zal zich moeten richten op die aspecten die flexibiliteit en aanpassingsvermogen in de taakuitvoering binnen teams bevorderden.

In tegenstelling tot de vorige hoofdstukken, die waren gericht op het evalueren van leeroplossingen in het algemeen richt dit hoofdstuk zich daarom op dynamische teams. Daarbij gaat het niet alleen om het evalueren van leeroplossingen en -interventies, maar ook op het meten van effectiviteitsfactoren van (dynamische) teams, los van gerichte interventies. Het onderhavige hoofdstuk richt zich, met andere woorden, op het *meten van de teameffectiviteit* in generieke zin. Dit kan bijvoorbeeld de uitkomst zijn van een leerinterventie gericht op het teamfunctioneren maar ook het monitoren van de ontwikkeling van een team tijdens een operatie. Om onderzoek naar teameffectiviteit te ondersteunen en om effecten van teamtraining te evalueren in dynamische team-settings hebben we in dit verband een (generiek bruikbaar) meetinstrument ontwikkeld dat is beproefd in een pilot.

4.2 Effectiviteit van teamtraining en teamprestaties

4.2.1 *Teameffectiviteit*

Om als dynamisch team goed te kunnen presteren zijn er twee typen activiteiten noodzakelijk: *Taskwork* en *Teamwork* (Mills Mills, Pascual, Blendell, and Verral, 1999; Essens, Vogelaar, Mylle, Blendell, Parijs, Halpin et al., 2005). *Taskwork* wordt hier door ons gedefinieerd als: de manier waarop een teamlid gebruik maakt van inhoudelijke kennis en vaardigheden bij het uitvoeren taken. *Teamwork* wordt gedefinieerd als "de manier waarop een teamlid in wisselwerking met anderen taken coördineert en zo teamdoelen realiseert" (zie ook Mills et al., 1999, p. 11). *Teamwork* zorgt ervoor dat het *taskwork* van de individuele teamleden dusdanig wordt verenigd zodat het team zijn gemeenschappelijke doel kan bereiken.

In Tabel 1 wordt het onderscheid tussen de eerder behandelde *proces-* en *uitkomstmaten* en de zojuist beschreven *teamwork-taskwork* dimensie schematisch weergegeven. Dit overzicht laat zien dat een meetinstrument generieker is, en dus breder toepasbaar, naarmate het meer gericht is op *teamwork* activiteiten en de onderliggende *processen* die het uiteindelijk behaalde resultaat (uitkomst, output) bepalen.

Tabel 1 Onderscheid termen proces-uitkomst maten en team- taskwork activiteiten.

	Generiek	Specifiek
Meetinstrument	Procesmaat	Uitkomstmaat
Activiteit teamleden	Teamwork	Taskwork

Binnen de militaire context worden dynamische teams samengesteld uit teamleden met elk hun eigen specifieke inhoudelijke deskundigheid (expertise, domeinkennis). Deze specifieke deskundigheid is meestal de belangrijkste reden waarom iemand aan een dynamisch team wordt toegevoegd. Gezien de grote hoeveelheid specifieke taken en bijbehorende deskundigheden die binnen militaire dynamische teams kunnen bestaan, is het niet goed mogelijk om daarvoor generieke, breed toepasbare prestatie-maten te ontwikkelen. Daarom is besloten om het taskwork gedeelte buiten de scope van het onderzoek te houden. Wij gaan er in het vervolg steeds van uit dat de individuele taak expertise voor het behalen van de (dynamische) teamdoelstellingen binnen het team voorhanden is. Anders dan taskwork komen teamwork, en de processen die daar de grondslag van zijn, voor allerlei teams sterk overeen. Vanwege dit generieke karakter valt hier dus het meeste winst te behalen. Daarom is ervoor gekozen om een instrument te ontwikkelen gericht op *teamwork* en de meting en evaluatie van onderliggende generieke *teamprocessen* die de basis vormen van de uiteindelijke dynamische team prestatie.

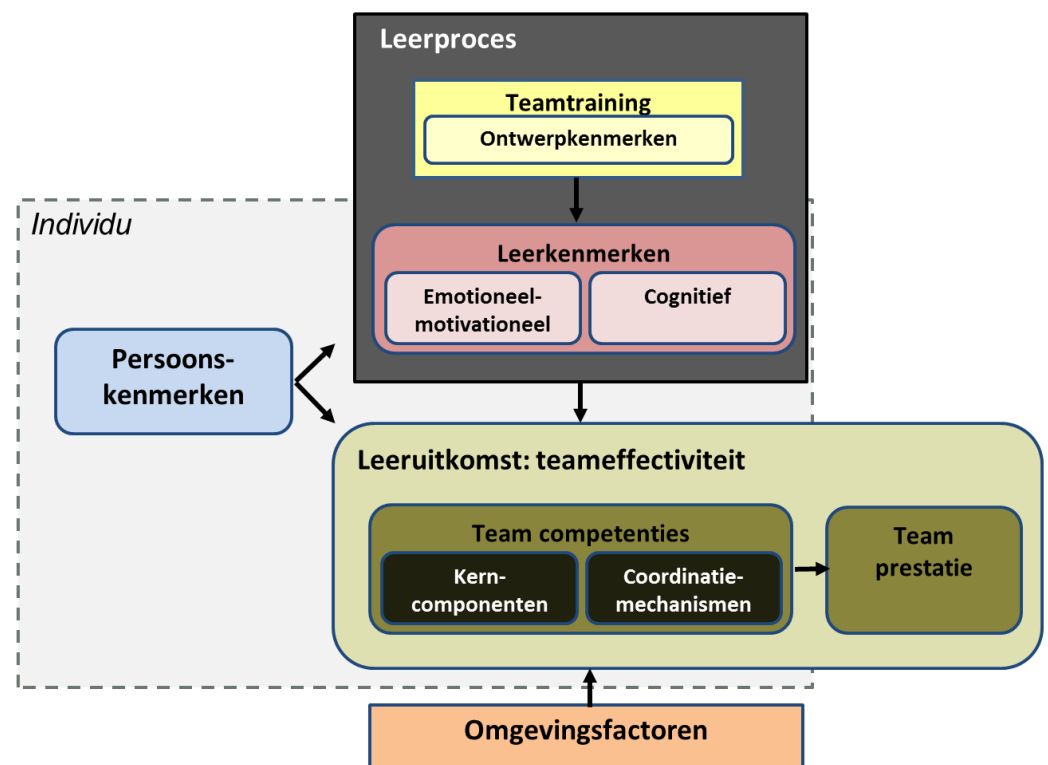
Het in hoofdstuk 2 gepresenteerde raamwerk voor het evalueren van leerinterventies (leeroplossingen, trainingen) kan zowel voor individuele- als voor teamtraining worden toegepast. Bij (dynamische) teams kan een interventie bijvoorbeeld een teambuilding zijn, een teamoefening in een simulator, coaching van het team op de werkplek, etc. Het doel daarvan is de teamwork processen, en daarmee de *teameffectiviteit* te verbeteren. In dat geval is het effect op teameffectiviteit de uitkomst van de interventie. Uiteraard kunnen de teamwork processen die ten grondslag liggen aan teameffectiviteit ook losstaand van een training of interventie gemeten worden.

4.2.2 *Het meten van teameffectiviteit*

De teameffectiviteit, al of niet in het kader van een leerinterventie, kan op diverse manieren worden gemeten. Indien er wordt gemeten of een training of interventie op individueel niveau effect heeft gehad op het functioneren in een team, kunnen teamcompetenties van de individuele teamleden worden gemeten (zie bijvoorbeeld Venrooij, 2009). De aanname hierbij is dat deze competenties op individueel niveau gezamenlijk van invloed zijn op de uiteindelijke teamprestatie. Indien er wordt gemeten hoe het team als geheel functioneert, kan beoordeeld worden of het team als geheel bepaalde taken goed uitvoert (uitkomst, resultaat) en op welke manier het team tot het eindresultaat (proces) komt.

Bij een evaluatie van een teamtraining of –interventie kunnen zowel de generieke aspecten specifiek op het leerproces gericht als de team- of taakspecifieke uitkomsten gemeten moeten worden. Dit werd reeds uitgelegd in hoofdstuk 2 over

het generieke evaluatieraamwerk. Afhankelijk van het type training of interventie moeten ook bepaalde ontwerpkenmerken (kenmerken van de leerinterventie) worden geëvalueerd; bijv. een training met behulp van simulatie is verschillend van een teambuilding oefening op het veld. Onderstaande figuur 7 geeft een schematische weergave van het evaluatieraamwerk toegepast op teamtraining. Deze figuur geeft dus een gedeelte van het complete evaluatieraamwerk weer, uitgewerkt voor teams. Hierbij is de leeruitkomst (Teameffectiviteit) vergroot weergegeven. Deze bestaat uit teamcompetenties die nodig zijn om een bepaalde teamprestatie te kunnen leveren. De teamcompetenties bestaan uit zogenaamde *kerncomponenten* en *coördinatiemechanismen* gebaseerd op het model van Salas et al., (2005); dit model wordt verder gepresenteerd en uitgewerkt in de volgende paragraaf. Bovendien is gevisualiseerd wat bij het evalueren van een leeroplossing op individueel niveau gemeten kan worden, bijvoorbeeld door de lerende een vragenlijst af te nemen die bestaat uit persoonskenmerken, leerkenmerken en teamcompetenties. Een keus uit de persoonskenmerken en leerkenmerken moet worden gemaakt die past bij de specifieke doelstelling van het evaluatieonderzoek; niet alle indicatoren hoeven en kunnen (praktisch gezien) te worden bevraagd (zie vorige paragraaf voor een set aan mogelijke indicatoren waaruit gekozen moet worden). De teamprestatie als algeheel *eindresultaat* kan bijvoorbeeld worden gemeten door het team een bepaalde taak te laten uitvoeren en deze (voor zover mogelijk) objectief te meten (bijvoorbeeld aantal fouten, tempo/tijd).



Figuur 7 Evaluatieraamwerk toegepast op teamtraining.

Teameffectiviteit kan ook gemeten worden zonder dat er sprake is van een leerinterventie bijvoorbeeld om de algehele ontwikkeling van een team gedurende een bepaalde periode in kaart te brengen. In dat geval beperkt de evaluatie van een team zich tot het meten van teameffectiviteit onder invloed van andere factoren dan een leerinterventie. In dat geval worden ook de kenmerken van het leerproces,

zoals die in de black box zijn gepresenteerd, niet meegenomen. Daarbij kunnen tevens een aantal omgevingsfactoren worden meegenomen die op teameffectiviteit van invloed zijn, zoals omstandigheden waarin het team moet opereren (bijvoorbeeld type missiegebied en dreiging vijand) maar ook team-specifieke factoren zoals personeelwisselingen en een de mix van competenties en persoonlijkheden.

In het vervolg van dit hoofdstuk zal worden toegelicht wat voor meetinstrument is ontwikkeld om teameffectiviteit te bemeten. Eerst zal de keuze van het gebruikte teammodel worden onderbouwd en toegelicht. Daarna wordt het ontwikkelde meetinstrument beschreven.

4.3 Teamwork model

4.3.1 *Keuze van teamwork model*

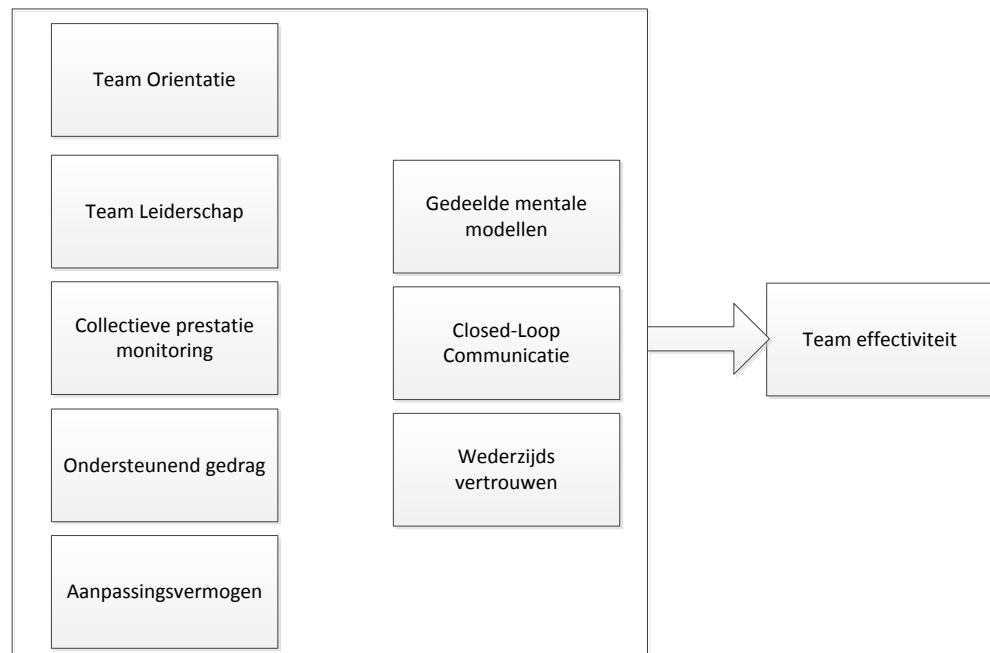
Het raamwerk voor het meten van dynamische team-effectiviteit is gebaseerd op een combinatie van het eerder in dit rapport beschreven algemene evaluatie raamwerk en een bestaand wetenschappelijk teammodel. In deze paragraaf zal de keuze voor dit teammodel verder worden toegelicht.

In de afgelopen jaren is er in de literatuur een groot aantal verschillende modellen voor teamwork beschreven. Diverse onderzoekers hebben pogingen ondernomen om deze verschillende modellen te integreren tot een model dat de meest essentiële processen voor teamwork beschrijft (Krokos et al., 2009, Essens et al., 2005, Cannon-Bowers, Tannenbaum, Salas en Volpe, 1995). Bij het vergelijken van teamwork modellen valt het op dat deze op een groot aantal punten variëren, zoals abstractieniveau, doelgroep, generaliseerbaarheid, gebruikte modelstructuur (bijvoorbeeld Input Process Output) en complexiteit.

Gezien de aard van het onderhavige project hebben we de modellen op basis van de volgende criteria met elkaar vergeleken:

- 1 Toepasbaar op een verscheidenheid van dynamische teams;
- 2 Fundering op wetenschappelijk onderzoek;
- 3 Begrijpelijkheid en plaatsbaar binnen militaire context;
- 4 Behapbaar abstractie- en detailniveau (beperkt aantal concepten) .

Na vergelijking van de teamwork modellen hebben we gekozen voor het model van Salas, et al., (2005). Zij hebben op basis van een uitgebreide studie van de literatuur van de afgelopen twintig jaar een model opgesteld waarin de zogenaamde 'Big Five' van teamwork processen wordt gepresenteerd. Het model is toepasbaar op verschillende verschijningsvormen van dynamische teams. Tevens bevat het een beknopt aantal constructen die vrij eenvoudig te begrijpen zijn en goed in de militaire context te plaatsen zijn. Het model omvat vijf essentiële teamwork *kerncomponenten* en drie *coördinatiemechanismen* die worden beschouwd als de meest essentiële processen die nodig zijn voor elk soort team (Salas et al., 2005). De kern van het model van teamwork voorgesteld door Salas et al., (2005) is weergegeven in Figuur 8.



Figuur 8 Model van teamwork (Salas et al., 2005, p. 571).

In deze figuur staan aan de linkerkant de vijf kerncomponenten en aan de rechterkant de drie coördinatie mechanismen. In het resterende deel van deze paragraaf worden de genoemde vijf kerncomponenten en de drie coördinatie mechanismen verder toegelicht.

4.3.2 Kerncomponenten

1. Team oriëntatie

Dit onderdeel van teamwork heeft niet alleen betrekking op gedrag, maar ook op houding (Salas et al., 2005). Team oriëntatie is gedefinieerd als het tijdens de groepsinteractie rekening houden met het gedrag van anderen en de overtuiging dat teamdoelen boven die van het individu staan (Salas et al., 2005, p. 560). Team oriëntatie is dus breder dan alleen de individuele voorkeur om in een team te werken (Driskell & Salas, 1992).

2. Team leiderschap

Er is in de afgelopen decennia veel onderzoek gedaan naar de rol van leiderschap bij het behalen van de effectiviteit van teams (Steward & Manz, 1995, Zaccaro, Marks, O'Connor-Boes & Constanza, 1995). Team leiderschap wordt gedefinieerd als het vermogen om de activiteiten van de andere teamleden te leiden en coördineren, teamprestaties te beoordelen, taken toe te wijzen, teamkennis en vaardigheden te ontwikkelen, teamleden te motiveren, te plannen en organiseren, en te zorgen voor een positieve sfeer (Salas et al., 2005, p.560). Teamleiderschap heeft geen directe invloed op de effectiviteit van teams, maar draagt eraan bij via de positieve invloed op de andere kerncomponenten en coördinatie mechanismen (Salas et al., 2005).

3. *Collectieve prestatie monitoring*

Collectieve prestatie monitoring wordt door Salas et al., (2005) gedefinieerd als het vermogen om gemeenschappelijke inzicht van de teamomgeving te ontwikkelen en passende strategieën te hanteren voor het in de gaten houden van de prestatie van teamgenoten. Als de voortgang van taken van teamleden op een collegiale manier kan worden gecontroleerd, stelt dit het team in staat om tijdig te anticiperen door ondersteuning te bieden.

4. *Ondersteunend gedrag*

Ondersteunend gedrag wordt omschreven als "het vermogen om op de behoeften van andere teamleden te anticiperen door middel van nauwkeurige kennis over hun verantwoordelijkheden. Bijvoorbeeld door de werklast te schuiven tussen teamleden om zo het evenwicht binnen het team te bewaren tijdens periodes van hoge werkdruk of druk" (Salas et al., 2005, p.560). Deze kerncomponent beïnvloedt de effectiviteit van teams rechtstreeks doordat goed ondersteunend gedrag ervoor zal zorgen dat alle onderdelen van de team taak worden voltooid (Salas et al., 2005).

5. *Aanpassingsvermogen*

Dit kernelement van teamwork verwijst naar het vermogen om processen binnen het team aan te passen aan de (veranderende) omgeving of situatie zodat de gestelde teamdoelen worden gerealiseerd (Salas et al., 2009). Met name binnen teams die opereren in dynamische omstandigheden is deze kerncomponent essentieel voor het behalen van resultaat. Deze veranderende omstandigheden kunnen zowel binnen het team (andere samenstelling) als daarbuiten (actoren, taak of omgeving) optreden.

4.3.3 *Coördinatiemechanismen*

Als aanvulling op deze kerncomponenten van teamwork, staan aan de rechterkant van het model van Fig. 8 ook drie *coördinatiemechanismen*. Deze mechanismen hebben een positief effect op de vijf belangrijkste processen van teamwork (Salas, et al., 2005). Deze drie coördinerende mechanismen zijn:

1. Closed-loop communicatie

Dit coördinatiemechanisme heeft betrekking op een specifieke manier van communiceren die effectief teamwork faciliteert (Salas et al., 2009). Deze manier van communiceren bestaat uit overbrengen van een boodschap van de zender naar de ontvanger waarbij de zender controleert of de boodschap juist is ontvangen en is geïnterpreteerd (Salas et al., 2009.)

2. Gedeelte mentale modellen

Gedeelte mentale modellen zijn gedefinieerd als georganiseerde kennis van de relatie structuur binnen het team omtrent de taakuitvoering en op welke wijze de teamleden met elkaar zullen communiceren (Salas et al., 2005, p.560). Door het hebben van een gedeeld mentaal model kunnen teamleden nieuwe informatie op eenzelfde manier interpreteren. Hierdoor zal een gedeeld mentaal model de volgende teamwork kerncomponenten positief beïnvloeden: coördinatie door bijvoorbeeld de teamleider (Marks, Sabella, Burke, Zaccaro, 2002; Salas et al., 2009), ondersteunend gedrag (Ilgen et al., 2005) collectieve prestatie monitoring, en aanpassingsvermogen (Salas et al., 2005).

3. Vertrouwen

Wederzijds vertrouwen is het gedeelde geloof dat individuen in een team bepaalde activiteiten zullen uitvoeren die belangrijk zijn voor de anderen binnen het team waarbij zij de belangen en rechten van andere teamleden erkennen en beschermen (Salas et al., 2009, p. 46). Het hebben van onderling vertrouwen stelt het team in staat om bijvoorbeeld zonder achterdocht de voortgang op elkaars taken in de gaten te kunnen houden.

4.4 Meetinstrument

TNO heeft reeds eerder een gevalideerde vragenlijst ontwikkeld (Venrooij, 2009) die, op basis van het hierboven beschreven model van Salas et al., (2005), de kernelementen en coördinatiemechanismen kan meten. Deze gevalideerde vragenlijst bestaat uit bijna 40 vragen en kan in principe goed worden toegepast in het in paragraaf 4.2 beschreven evaluatieraamwerk voor dynamische team effectiviteit. Dit is echter een competentievragenlijst op individueel niveau die als gevolg van de lengte niet geschikt om snel herhaald (bijvoorbeeld dagelijks of wekelijks) de teamsituatie te meten. Gezien de snel veranderende aard van dynamische teams is hier vanuit toegepast wetenschappelijk oogpunt echter wel vraag naar. Daarnaast bestaat er een behoefte vanuit de praktijk om snel en herhaald aan dynamische teams te kunnen meten. Eerder in dit rapport is reeds aandacht besteed aan het feit dat de toenemende dynamiek van Defensie operaties specifieke eisen stelt aan teams om effectief hun taken uit te voeren. Het effectief omgaan met veranderingen die zich binnen en buiten het team afspelen is een uitdaging. Mentale modellen moeten bijvoorbeeld continu worden aangepast op een nieuwe werkelijkheid, de opbouw van wederzijds vertrouwen wordt bemoeilijkt door snelle personele wisselingen en communicatie binnen het team wordt complexer zodra de diversiteit in specialismen groter wordt. Dynamiek heeft dus een potentieel verstrendend effect op de coördinatie mechanismen van een team. Zodra deze coördinatiemechanismen worden verstoord zullen ook de kerncomponenten van teamwork minder goed kunnen functioneren en uiteindelijk mogelijk zelfs wegvallen. Operationele dynamiek kan op die manier de effectiviteit van een geheel team in gevaar brengen.

Het is de taak van de teamleider om de effectiviteit van het dynamische team te waarborgen en daarom verstoringen die voortkomen uit de dynamiek tijdig op te merken en hierop te acteren. Voor teamleiders van dynamische teams is daarom het hebben van inzicht in de huidige stand van zaken betreffende de kernelementen en coördinatie mechanismen van groot belang. Door het tijdig inspelen op verstoringen hierin kan de teamleider voorkomen dat deze het teamwork negatief beïnvloeden. Een vragenlijst die snel een beeld geeft van het teamfunctioneren en tevens inzicht verschaft in die aspecten van teamwork die niet optimaal verlopen kan de teamleider ondersteunen bij zijn taak. Op basis van de uitkomsten kan de teamleider interventies plegen gericht op tijdige aanpassing en herstel.

Het ontwikkelde instrument is anders van aard dan reeds bestaande meetinstrumenten om aspecten van teameffectiviteit te meten, zoals bijvoorbeeld CTEF (Essens et al., 2005), CTCQ (Venrooij, 2009). Ons instrument, getiteld *TeamScan*, betreft een snel uit te voeren scan waarmee inzicht kan worden verkregen in hoe het team er qua kernelementen en coördinatie mechanismen voor staat. Omdat het instrument snel en veelvuldig in de praktijk moet worden afgenomen is de lengte

ervan beperkt. Daartoe zijn in een workshop op basis van de beschikbare beschrijvingen van de teamwork elementen van Salas et al., (2005) de elementen tot twee, maximaal drie goed te begrijpen items teruggebracht.

TeamScan is nadrukkelijk bedoeld om, als een soort peilstok, frequent (bijvoorbeeld wekelijks) de momentane situatie op te nemen. Hierbij wordt gebruik gemaakt van een continue schaal, zodat ook subtiele veranderingen in de situatie over de tijd (visueel) te kunnen weergeven. Juist voor dynamische teams, die snel wisselen in samenstelling, en waarbij de veranderingen over de tijd centraal staan, is een dergelijk “snel” meetinstrument van belang. Een (beperkte) ratingschaal biedt deze mogelijkheid over het algemeen veel minder, ook al heeft een ratingschaal met heldere ankers voordelen omdat dan voor de beoordelaars duidelijker is wat een bepaald oordeel betekent (zie bijvoorbeeld Aiken, 1996; Guion, 1998). De vorm van *TeamScan* lijkt hiermee enigszins op meetinstrumenten als de veel gebruikte Rating Scale Mental Effort (RSME, Zijlstra 1993) die ook is bedoeld om herhaald een oordeel te kunnen geven en de verandering over de tijd te kunnen weergeven, in dit geval over mentale inspanning. Er is gekozen voor een range van 1 tot 10 omdat beoordelaars dit, vanwege het verband met schoolcijfers, relatief beter kunnen.

In *TeamScan* zijn voor elk van de vijf kerncomponenten en drie coördinatie-mechanismen op basis van de definities van Salas et al., (2005) één of meerdere items geformuleerd. Hierbij moet worden benadrukt dat de items per construct (zijnde kerncomponent of coördinatiemechanisme) niet per definitie gezamenlijk het construct zelf meten want hiervoor is het aantal items per construct te laag. Elk item meet een apart onderdeel van het kerncomponent of coördinatie mechanisme. Analyses hebben betrekking op veranderingen over de tijd voor het gemeten onderdeel. In een eventuele toekomstige validatie van dit meetinstrument moet in kaart worden gebracht wat de bijdrage is van de diverse items aan de teameffectiviteit als geheel alsmede de onderlinge samenhang. Vervolgens kan dan een definitieve selectie van de meest relevante items worden gemaakt. Hieronder staat de lijst van items in de volgorde van afname.

Team oriëntatie:

1. Betrokkenheid: De teamleden voelen zich betrokken bij het team.
2. Wil om samen te werken: De teamleden willen graag met elkaar samenwerken.

Wederzijdse prestatie monitoring:

1. Monitoring: De teamleden volgen de voortgang van elkaars taken.
2. Feedback: De teamleden geven elkaar feedback over elkaars taakuitvoering.

Ondersteunend gedrag

1. Herkennen: De teamleden herkennen onevenwichtige werklastverdeling.
2. Ondersteunen: De teamleden helpen elkaar als dat nodig is.

Adaptiviteit

1. Buiten het team: Het team speelt adequaat in op veranderingen/gebeurtenissen in de omgeving (het spel).
2. Binnen het team: Het team speelt adequaat in op veranderingen binnen het team.

Closed Loop communicatie

1. Verzamelen: De teamleden verzamelen voor het team relevante informatie.
2. Communiceren: De teamleden communiceren relevante informatie en doen dit. Zo nodig ongevraagd.
3. Contoleren: De teamleden zorgen ervoor dat de informatie door hun team genoten goed wordt ontvangen en begrepen

Gedeelde mentale modellen

1. Inhoud taken: De teamleden kennen en begrijpen elkaars taken.
2. Relaties taken: De teamleden kennen en begrijpen de relaties tussen elkaars taken.
3. Communicatie: De teamleden gebruiken en "verstaan elkaars taal".

Wederzijds vertrouwen

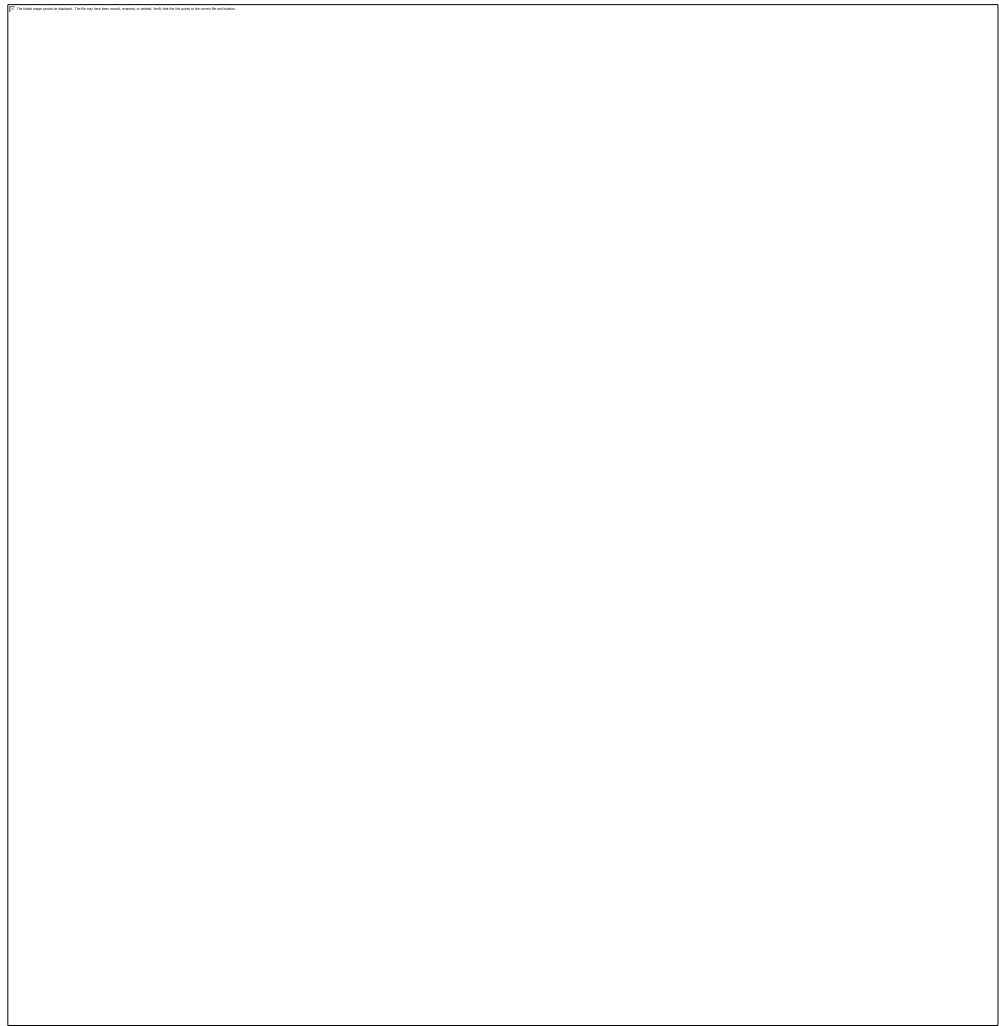
1. Taken: De teamleden vertrouwen erop dat hun teamgenoten hun taken goed zullen uitvoeren.
2. Belangen: De teamleden vertrouwen erop dat de andere teamleden de belangen van hun teamgenoten zullen beschermen.
3. Opstelling: De teamleden durven zich kwetsbaar op te stellen (open en eerlijk zijn en fouten durven toegeven).

Leiderschap

1. Samenwerking: De teamleider bevordert (wanneer nodig) een optimale samenwerking op alle relevante team-aspecten (denk daarbij aan alle hier boven genoemde aspecten).
2. Informeel leiderschap: De teamleden verdelen/verschuiven informeel leiderschap al naargelang de eisen van de (veranderende) taken en omstandigheden.

Totaal

Totaal-inschatting over de kwaliteit waarmee het team werkt aan het gezamenlijke doel.



Figuur 9 Screenshot digitale versie *TeamScan*.

5 Pilot studie

5.1 Inleiding en vraagstelling

In paragraaf 3.4 van het evaluatieraamwerk werden vier aspecten genoemd die bepalend zijn voor de uiteindelijke kwaliteit, of waarde, van een meetinstrument. In een eerste, zeer beperkte pilot studie werd een eerste inventarisatie gedaan van twee kwalitatieve aspecten van *TeamScan*: realiseerbaarheid (uitvoerbaarheid, haalbaarheid), en bruikbaarheid (toepasbaarheid, functionaliteit). De meer kwantitatieve kwaliteitsaspecten van meetinstrumenten validiteit en betrouwbaarheid ontbreken hier omdat dit, gegeven de scope van het huidige project, niet haalbaar was. Voor deze pilot werd gebruik gemaakt van een teamspel, getiteld Swift Team Effectiveness And Maturity (STEAM). STEAM is in een eerder project door TNO in samenwerking met het trainingsbureau Schouten & Nelissen ontwikkeld en geëvalueerd. STEAM werd gebruikt door Marine, Marechaussee en Buitenlandse Zaken als onderdeel van de voorbereiding op uitzending in Kunduz, april van dit jaar. De game maakt onderdeel uit van hun opwerktraject. Trainingsbureau Schouten & Nelissen heeft STEAM inmiddels in haar trainingsaanbod opgenomen. STEAM is een spelconcept ontwikkeld waarin.

Omdat STEAM gericht is op het in korte tijd (5 tot 8 uur) inzicht verkrijgen in zowel zichzelf als teamlid als in het team waarvan met deel uitmaakt, leek het als zodanig bij uitstek geschikt om een pilot mee uit te voeren gericht op het meten van dynamische teameffectiviteit. Daarmee zou een eerste evaluatie met hands-on ervaring opgedaan kunnen worden met betrekking tot bovengenoemde aspecten van de kwaliteit van *TeamScan*. Meer specifiek ging het om de vraag in hoeverre *TeamScan* door teamleden en gebruikers (bijvoorbeeld trainers of leidinggevenden) handig te scoren zijn alsmede het verkrijgen van meer informatie over de mate waarin (of voorwaarden waaronder) het instrument realiseerbaar en bruikbaar is.

5.2 Teamtaak

In STEAM bestuurt een team van 11 personen een land. Vijf deelnemers besturen als regiomanager ieder een eigen provincie, vijf deelnemers fungeren als minister met een eigen portefeuille en er is een president die leiding geeft. Het spel bevat zowel elementen van samenwerking als van onderlinge competitie. Wat het eerste betreft staat het team gezamenlijk voor de opdracht het land als geheel zo welvarend mogelijk te maken. Van competitie is sprake omdat ministers en president bepaalde voorrechten hebben (en functies kunnen wisselen) en omdat de deelnemers een eigen regio moeten stimuleren, dan wel een regio-overstijgende portefeuille beheren.

Het spel wordt door de spelleider afgestemd op de deelnemers, wat onder andere tijdswinst oplevert, maar ook teamprocessen beter zichtbaar kan maken. Dit houdt in dat deelnemers vooraf drie tests doen. Deze assessment brengt verschillende kenmerken van elke deelnemer in kaart. Dit betreft: persoonlijkheid (*De Big Five*, Digman, 1990; Pervin, 2003), Team-rol (*Belbin*, Robbins & Judge, 2007) en gedragsvoorkeur (*DISC*, Bonnstetter, Suiter, Widrick, 2001). De profielen die hieruit voortkomen worden gebruikt om het spel zo in te richten dat kenmerken van spelers snel zichtbaar worden. Trainers hebben zo vooraf al inzicht in leerpotentieel en

potentiele knelpunten van een team voordat het team is gestart met het spel. Op basis van de assessment krijgen de spelers rollen toebedeeld en selecteren de trainers tijdens het spel gerichte interventies. Elke deelnemer wordt zo uitgedaagd op vlakken die specifiek voor hem/haar van belang zijn als teamspeler. Er wordt gestuurd op het ervaren van de potentiële krachten en knelpunten in het team en op het bespreekbaar maken hiervan. Het daadwerkelijk spelen van het spel vindt plaats in drie spelronden (van drie kwartier) gevolgd door drie reflectiesessies (van een half uur). Tijdens de reflectiemomenten bespreekt men onder leiding van de spelleider de eigen ervaring en elkaars gedrag tijdens de voorafgaande spelronde.

5.3 Procedure

De pilot die ten behoeve van het onderhavige onderzoek werd gedaan, werd geleid door een ervaren medewerker van *Schouten & Nelissen*. Daartoe werd *TeamScan* vier maal in een papieren versie door 11 deelnemers ingevuld, één keer na de introductie en voorafgaand aan de eerste spelronde en drie keer na iedere reflectiesessie over de voorafgaande spelronde. Tot slot werd de lijst ook nog een keer ingevuld door de spelleider en door een TNO medewerker die gespecialiseerd is in teamprocessen en die het proces de hele dag had geobserveerd. Aan het slot, nadat *TeamScan* voor de vierde keer was afgenomen, werden door alle betrokkenen nog vier vragen beantwoord over de begrijpelijkheid van de vragen en over de beschikbaarheid van informatie over het teamproces om de vragen te kunnen beantwoorden:

Heb je de vragen steeds goed genoeg begrepen om ze te kunnen beantwoorden? *Ja/nee*

Zo nee, wat vond je lastig?

Had je meestal wel voldoende gehoord en gezien van het gedrag van je teamgenoten om de vragen te kunnen beantwoorden? *Ja/nee*

Zo nee, over welke vragen was je het meest onzeker?

5.4 Resultaten

Realiseerbaarheid

Tijdens de studie is gebleken dat het meetinstrument goed en praktisch af te nemen was. Dit kon zowel met een papieren versie (zoals in de onderhavige pilot) als met de computer-based versie die in Fig. 9 te zien is.

Bruikbaarheid

Met betrekking tot de *Bruikbaarheid* van de vragenlijst gaven 6 van de 11 deelnemers aan geen een moeite te hebben met het scoren van de stellingen. Van de overige 5 gaven 4 aan dat ze vooral onzeker waren over hun eigen inschatting van de intenties van anderen en gaf 1 deelnemer aan dat hij de vragen niet goed kon begrijpen omdat hij ze niet logisch gestructureerd vond (gebrek aan subkopjes e.d.).

Tevens werd aan het eind van de STEAM sessie aan de spelleider en de observator gevraagd om *TeamScan* in te vullen. Door te vergelijken wat de diverse personen hebben ingevuld, kon een indicatie worden verkregen of de vragenlijst voor de 'experts' goed in te vullen is. Indien dit zou plaatsvinden met grote aantallen deelnemers in een gecontroleerde conditie zou deze methode gebruikt kunnen worden om de interbeoordelaarsbetrouwbaarheid vast te stellen maar dat was in deze beperkte pilotstudie niet aan de orde. Als richtlijn is aangehouden dat een verschil van 2 punten enigszins acceptabel is aangezien er altijd meningen bestaan over wat 'goed', dan wel 'slecht' is, valt op dat de spelleidster en observator op slechts drie stellingen meer dan 2 punten van elkaar afwijken bij het scoren. Dat waren de volgende items:

Wederzijdse prestatie monitoring 2: de teamleden geven elkaar feedback over elkaars taakuitvoering.

Ondersteuning 1: de teamleden herkennen onevenwichtige werklastverdeling.

Adaptiviteit 1: het team speelt adequaat in op veranderingen/ gebeurtenissen in de omgeving (het spel).

De spelleidster oordeelde op deze items overal meer dan twee punten negatiever dan de TNO observator. Dit is mogelijk goed te verklaren doordat de spelleidster geen tijd had om constant te observeren en ervaring had met sessies die veel beter verliepen. Hierdoor kon ze feedback en inzichten met betrekking tot de werklast gemist hebben terwijl de observator deze wel zag. Op het gebied van adaptiviteit biedt de stelling ruimte voor interpretatie over wat adequaat is. Aangezien het spel ook bij topmanagement wordt ingezet is het geen vreemde gedachte dat de eisen over wat adequaat is bij de ervaren spelleidster hoger liggen dan de observator die deze groepen niet in relatie tot STEAM heeft meegemaakt.

Concluderend kan worden gesteld dat er tussen de scores gegeven door de experts bleken geen opvallende verschillen te bestaan. Zij hebben daarbij ook aangegeven met de vragen geen moeite te hebben.

Samenhang tussen items

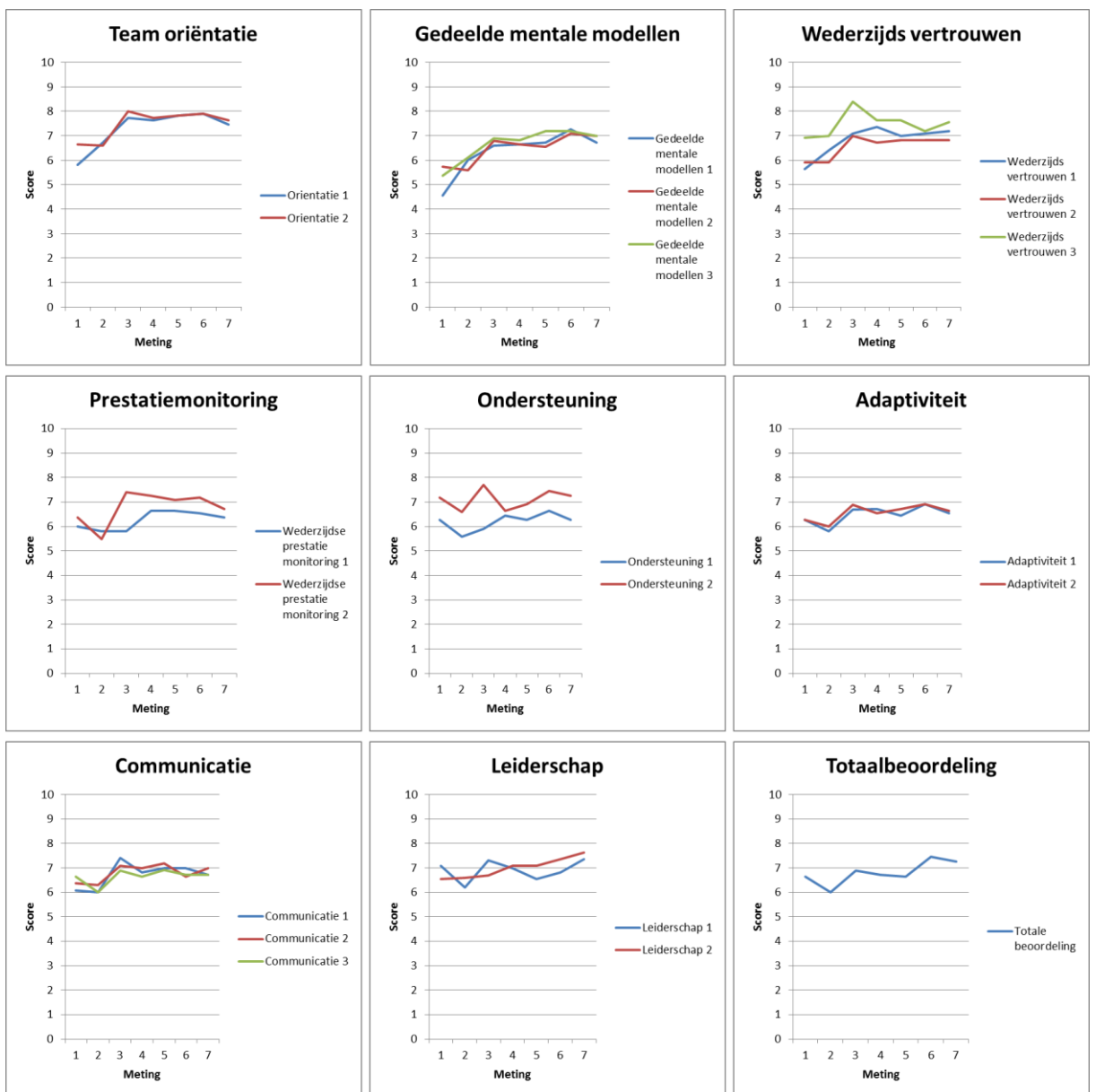
De scores op de verschillende items (zie bijlage A) bleken onderling vrij sterk te correleren, maar weer niet zo hoog (hoogste correlatie: 0,88) dat de verschillende items elkaar zouden kunnen vervangen. Dat wijst erop dat de verschillende items grosso modo betrekking hebben op hetzelfde overall concept, i.e. Teamfunctioneren, maar tegelijkertijd niet overtoollig (redundant) zijn, en dus werkelijk bijdragen aan de totale informatie over Teamfunctioneren.

Herhaalde metingen

De gemiddelde scores per kerncomponent/coördinatiemechanisme (cluster) toonden een groei over tijd, waarbij de sterkste groei bij de eerste sessies plaatsvond (zie Figuur 10). Deze gemiddelden zeggen niet zozeer iets over de individuele ontwikkelingen, maar wel over de gemiddelde beleving en laten zien dat deze over de opeenvolgende sessies heen aan verandering onderhevig was. Te zien is bijvoorbeeld dat de gemiddelde scores bij de meeste variabelen bij sessie 2 daalden of over het algemeen relatief laag waren. Dit was de meting die

afgenomen is na de eerste ronde STEAM. In die eerste ronde werd men doelbewust geconfronteerd met onduidelijkheden en frustraties over het spel, wat er goed toe geleid kan hebben dat de deelnemers een relatief negatieve beoordeling gaven aan hoe zij vonden dat het teamproces verliep. Bij sessie 3 stegen alle gemiddelde scores. Deze meting vond plaats direct na het eerste reflectiemoment na de eerste ronde. In het reflectiemoment werd onder andere besproken dat het enigszins chaotisch verlopen van de eerste ronde normaal is en kwam men tot het inzicht dat een en ander niet zo slecht was als in eerste instantie werd gedacht (en ingevuld).

Later in het proces werden de scores stabiel, mogelijk doordat de mensen elkaar beter leerden kennen en er geleidelijk een beter beeld ontstond over de rolverdeling en kwaliteiten van het team.



Figuur 10 Gemiddelde scores per sessie per cluster.

5.5 Voorlopige conclusies met betrekking tot het meetinstrument

Met de slag om de arm dat het evaluatie-onderzoek beperkt van opzet was en (dus) met een relatief laag aantal proefpersonen is uitgevoerd is *TeamScan* goed realiseerbaar en bruikbaar gebleken. Het instrument lijkt snel en eenvoudig af te nemen en de scoring leverde weinig problemen op. Uit de scores gegeven door de experts kwamen geen opvallende verschillen of interpretatieproblemen naar voren. De scores op de verschillende items lijken ook op een zinvolle manier bij te dragen aan de totale informatie over Teamfunctioneren.

Echter, er is onderzoek nodig met meer proefpersonen waarin daadwerkelijk gekeken wordt naar de psychometrische kwaliteit van het instrument. De hier uitgevoerde pilot was hiertoe te beperkt.

De *TeamScan* scores werden bepaald door het moment van afname. Op basis van het scoreverloop lijkt *TeamScan* dus in staat om veranderingen in de teamsituatie waar te nemen en dus te relateren aan interventies/gebeurtenissen in het teamproces. In combinatie met de hierboven geconstateerde goede bruikbaarheid betekent dit dat het instrument zich goed kan lenen voor het herhaaldelijk monitoren van (dynamische) teams. Op dit moment bestaat er dan ook belangstelling om het instrument in 2014 te gaan gebruiken bij het Operational Headquarters van de EU missie ATALANTA als hulpmiddel voor *Training and Evaluation*.

Het is de vraag welk beeld het best geschetst wordt, het groepsbeeld of de individuele beelden. Beide zouden goed bruikbaar kunnen zijn, afhankelijk van de manier waarop *TeamScan* wordt ingezet. Enerzijds kan een groepscommandant of teamleider de lijst zélf op opeenvolgende momenten invullen om voor zichzelf na te gaan in hoeverre het teamproces (vanuit zijn eigen perspectief bezien) op alle relevante deelaspecten goed verloopt. Dit kan hem bijvoorbeeld helpen om veranderingen door de tijd heen te signaleren en zonder daarbij relevante aspecten over het hoofd te zien. Anderzijds kan hij *TeamScan* op een aantal momenten door de individuele teamleden laten invullen om een goed overzicht te krijgen van het teamfunctioneren door de tijd, zoals dat door de teamleden zélf wordt ervaren. Dat kan tevens inzicht geven in eventuele onevenwichtigheden of grote discrepanties tussen ervaringen van teamleden. Bijvoorbeeld, een hoge score van het merendeel van de teamleden op “wederzijds vertrouwen” met uitzondering van twee andere teamleden die dit heel anders zien kan dan aanleiding zijn om de oorzaak hiervan te achterhalen en weg te nemen. Beide wijzen van gebruik kunnen worden toegepast om op onderdelen van het teamfunctioneren in te kunnen grijpen.

6 Conclusies en toepassing

6.1 Conclusies

Defensie wordt veelvuldig ingezet voor crisisbeheersing- en humanitaire operaties met steeds meer nadruk op *joint* en *combined* optreden en met toenemende samenwerking met andere deelnemers en stakeholders. Die samenwerking tussen instanties is soms zo nauw dat inbedding in elkaars structuren plaatsvindt en gezamenlijke teams worden samengesteld (optreden in een *multi-agency* omgeving). Ook dichterbij huis, namelijk bij crisismanagement en de bestrijding van terrorisme, zien we steeds meer dynamiek in de teams: functionarissen treden toe of treden uit een team, afhankelijk van doel, rol en wijze van inzet. Kenmerkend voor dit optreden is het kort-cyclische karakter en de variatie hiervan. Die variatie kan optreden zowel binnen het team, bijv. in de teamsamenstelling, als in de omgeving daarbuiten. De teams die voor deze wijze van optreden nodig zijn, zijn samengesteld uit leden die een uiteenlopende expertise hebben en die bij elkaar zijn gebracht om een specifiek (korte-termijn) doel te bereiken. Daardoor bestaat dit soort teams bestaat dan ook vaak uit mensen die zelden of nooit met elkaar hebben gewerkt.

De voorbereiding van de militair op het werken in dergelijke dynamische teams, als onderdeel van geïntegreerde eenheden, is daardoor niet eenvoudig. Hiervoor moeten nieuwe leerinterventies worden ontwikkeld. Daarom start het onderhavige rapport met de onderbouwing en presentatie van een generiek *evaluatieraamwerk* voor het meten en evalueren van de effectiviteit van leerinterventies. Volgens dit raamwerk moet een effectstudie ook inzicht geven in de *onderliggende factoren* van het leerproces die de uitkomst ervan bepalen. Daarom dient er volgens het evaluatieraamwerk niet uitsluitend gekeken te worden naar de leer-input en leer-output, maar tevens naar het taak- of leerproces zélf. In tegenstelling tot *output* of uitkomstmaten, die betrekking hebben op het resultaat van een interventie, geven zgn. *procesmaten* inzicht in *hoe* er wordt geleerd en gepresteerd en hoe dit tot stand kwam. Het evaluatieraamwerk beslaat dan ook drie fasen: input, proces en output. De *input* betreft persoonskenmerken van de leerlingen, zoals componenten van de Big Five, biografische kenmerken of cognitieve capaciteiten. Het *proces* omvat de (emotionele en cognitieve) leerkenmerken hiervan, zoals motivatie, angst, self efficacy en zelfsturendheid. Deze aspecten worden sterk beïnvloed door het (didactisch) ontwerp van de leerinterventie en mogelijke beïnvloedende omgevingsfactoren, zoals de organisatie en persoonlijke omstandigheden. Tot slot resulteert dit alles in een output, oftewel een leeruitkomst in termen van relevante kennis, vaardigheden, attitudes (competenties) en prestaties. Omdat er vele soorten leeruitkomsten zijn, op verschillende niveaus, hebben we op basis van de literatuur een taxonomie van typen leeruitkomsten gepresenteerd die geschikt is om toe te passen bij het meten van leereffecten.

Vervolgens worden (quasi-) experimentele designs beschreven voor het meten van de effectiviteit van leerinterventies op basis van het gepresenteerde evaluatieraamwerk. Vaak gaat het daarbij om abstracte (onstoffelijke) constructen die “tastbaar” worden gemaakt, of “geoperationaliseerd” door middel van meet-instrumenten, die speciaal ontworpen worden voor het desbetreffende construct.

Er kunnen wat dit betreft verschillende categorieën meetinstrumenten worden gebruikt (bijvoorbeeld vragenlijsten, observatie-checklists of geautomatiseerde prestatiemeting) die alle hun voor- en nadelen hebben. Afhankelijk van het doel van het onderzoek (bijvoorbeeld objectieve prestatiemeting vs. meningen van gebruikers achterhalen) en de praktische haalbaarheid (bijvoorbeeld beschikbaarheid over digitale data, tijdsinvestering) dient er een keuze te worden gemaakt voor bepaalde type meetinstrumenten. Combinatie van verschillende typen meetinstrumenten, gegeven de aard en randvoorwaarden van de vraagstelling, verdient meestal de voorkeur. Van groot belang bij die keuze is daarnaast de kwaliteit van de meetinstrumenten zelf. Dit wordt bepaald door vier criteria: realiseerbaarheid (feasibility), bruikbaarheid (functionaliteit), betrouwbaarheid en validiteit. Bij de ontwikkeling, introductie en feitelijke toepassing van meetinstrumenten dient altijd met alle vier aspecten rekening te worden gehouden.

In de laatste twee hoofdstukken wordt een aantal onderdelen van het generieke evaluatieraamwerk uitgewerkt naar de evaluatie van teams. Daarbij werd rekening gehouden met het onderscheid tussen teamprocessen en –resultaten en tussen teamwork en taskwork. Een meetinstrument is generieker, en dus breder toepasbaar, naarmate het meer gericht is op het meten van teamprocessen en teamworkactiviteiten. Daarom werd ervoor gekozen om een meetinstrument te ontwikkelen en te evalueren dat is gericht op teamwork en teamprocessen van dynamische teams. Het meetinstrument, getiteld *TeamScan*, kan worden toegepast voor het snel meten en evalueren van teameffectiviteit zowel als gevolg van een trainingsinterventie als tijdens operationele activiteiten.

Het meetinstrument werd ontwikkeld op basis van een wetenschappelijk model voor teamfunctioneren (Salas et al., 2005). Dit model is ontwikkeld op basis van een uitgebreide studie van de literatuur van de afgelopen twintig jaar. Het model bevat acht constructen (zogenaamde “teamcompetenties”) die vrij eenvoudig te begrijpen zijn en goed in de militaire context passen. Met behulp van het meetinstrument kunnen op relatief eenvoudige wijze de meest relevante (en veranderende) *dynamische* aspecten van team functioneren door de tijd heen snel, eenvoudig en herhaald (longitudinaal) worden gemonitord en in kaart gebracht. Dit kan gebeuren tijdens het trainen of samenwerken als team gedurende een bepaalde periode of taakuitvoering.

In het laatste hoofdstuk werd *TeamScan* als prototype door middel van een pilotstudie geëvalueerd. De resultaten gaven aan dat *TeamScan* instrument is dat prima in staat lijkt om veranderingen door de tijd heen in de teamsituatie waar te nemen en dus te relateren aan interventies/gebeurtenissen in het teamproces. Daarnaast lijkt het instrument goed bruikbaar, niet alleen voor het groepsbeeld, maar ook voor individuele beelden, en discrepanties daartussen. *TeamScan* kan dus op verschillende manieren worden gebruikt afhankelijk van de behoefte en omstandigheden in, en rond, (dynamische) teams.

Op basis van de onderhavige pilotstudie met een beperkt aantal proefpersonen kon de betrouwbaarheid van het meetinstrument nog niet worden vastgesteld. Daarnaast heeft er nog geen validatieonderzoek plaatsgevonden waarin onderzocht wordt in hoeverre *TeamScan* ook daadwerkelijk informatie oplevert over en inzicht geeft in de kwaliteit van de teamprocessen. In dergelijk onderzoek zouden bijvoorbeeld

teams die hoog op de vragenlijst scoren een betere taakprestatie moeten laten zien dan gelijkwaardige, maar laag scorende teams (onder dezelfde omstandigheden).

6.2 Toepassing

De toepassing van het door ons ontwikkelde evaluatieraamwerk, betreft het meten van leereffecten op verschillende niveaus, waaronder bijvoorbeeld de *transfer of training* naar de praktijk. Het evaluatieraamwerk kan breed worden ingezet voor het evalueren van leerinterventies op individueel- en op teamniveau. Het kan worden toegepast voor allerlei soorten opleidingen, trainingen met een breed spectrum aan leermiddelen (bijvoorbeeld serious gaming, simulatoren) of interventies (bijvoorbeeld coaching, teambuilding) bij Defensie. Hoewel niet besproken in dit rapport zijn er, naast het nieuw ontwikkelde instrument voor teams, ook andere meetinstrumenten ontwikkeld en gevalideerd door TNO voor het evalueren van leerinterventies (zie bijvoorbeeld Oprins Bakhuis Roozeboom, Visschedijk, Kistemaker, 2013). Deze zijn gericht op emotioneel-motivationale en cognitieve aspecten van leren en zijn ook goed te gebruiken voor O&T bij Defensie.

Belangrijk is op te merken dat het raamwerk niet alleen gericht is op het meten van de uiteindelijke *leeruitkomsten* (output), maar ook op het verkrijgen van inzicht in (veranderende) *onderliggende processen* of factoren die effecten van leerinterventies bepalen. Dit kunnen persoonskenmerken van de doelgroep zijn (input), emotionele of cognitieve leerkenmerken van leerlingen of ontwerpkenmerken van de leerinterventie (leermiddelen, leerstof, werkvormen etc.). Deze factoren bepalen grotendeels eind-effecten van leerinterventies op de diverse niveaus van Kirkpatrick.

Daarnaast levert dit rapport informatie over (het type van) constructen die gemeten kunnen worden en informatie over kwaliteitscriteria van meetinstrumenten (zoals betrouwbaarheid en bruikbaarheid). Het raamwerk geeft daarmee veel informatie die toegepast kan worden door scholen, opleidingsinstanties, militaire eenheden en/of researchinstellingen voor het bepalen van een onderzoeksdesign en voor het kiezen een goede (combinatie van) maten en meetinstrumenten bij evaluaties van leerinterventies en oefeningen. Afhankelijk van onderzoeksdoel, het domein en contextuele randvoorwaarden dient een keuze te worden gemaakt voor de indicatoren en meetinstrumenten die moeten worden ingezet. Een hulpmiddel dat helpt bij het maken van goede keuzes (richtlijnen, beslisriteria, stappenplan) zou in de toekomst kunnen worden ontwikkeld.

In dit project is het evaluatieraamwerk als geheel nog niet toegepast op een leerinterventie voor teams. Alleen de teameffectiviteit zelf is in een pilot gemeten. In de toekomst zou toepassing op een leerinterventie kunnen plaatsvinden op basis van een teambuilding of teamtraining. Dit biedt inzicht in de toepasbaarheid van dit evaluatieraamwerk op teams.

Het prototype *TeamScan* als peilstok voor het snel en herhaald meten van team-effectiviteit, is specifiek van belang voor het leiding geven aan dynamische teams, juist vanwege het dynamische karakter van deze teams (tijdelijk, variabel, flexibel, wisselend van samenstelling etc.). De teamleider heeft altijd de taak om in te grijpen als hij constateert dat de teameffectiviteit wordt bedreigd. *TeamScan* kan de teamleider daarbij helpen zodat tijdig wordt geacteerd, bijvoorbeeld bij verstoringen

op de coördinatie mechanismen en kerncomponenten. Het team kan zich dan tijdig door middel van kleine interventies herstellen van verstoringen. Door op gezette tijden de vragenlijst in te (laten) vullen (wat heel weinig tijd en inspanning kost) kan de teamleider voortdurend de kwaliteit van het teamfunctioneren volgen en krijgt hij tevens inzicht welke aspecten daarvan niet optimaal verlopen. Ook wordt zo inzicht verkregen in het verloop van het teamfunctioneren door de tijd heen, mogelijk als gevolg van veranderingen binnen het team, in de teamomgeving of door een leerinterventie. Op basis daarvan kan beter feedback worden gegeven aan de teamleden en kunnen eventueel vervolg-interventies worden geïntroduceerd om hierop in te spelen (adaptief, op maat trainen). Bij CZSK bestaat belangstelling om het instrument in 2014 te gaan gebruiken bij het Operational Headquarters van de EU missie ATALANTA als hulpmiddel voor *Training and Evaluation*.

Wat hierbij nu nog ontbreekt zijn duidelijke handvatten voor leiders om op basis van de uitkomsten in te grijpen. Wat moet hij doen als het onderlinge vertrouwen wegvalt, of wanneer mentale modellen te veel dreigen uiteen te lopen? In vervolgonderzoek zouden deze handvatten kunnen worden ontwikkeld. Daarnaast moet verder worden nagedacht over de manier waarop de output over de tijd wordt gevisualiseerd. Om adequaat te kunnen ingrijpen moet de teamleider een goed overzicht hebben hoe het teamproces verloopt, op welke punten het goed gaat en waar er verbeterpunten zijn. Het gaat hierbij met name om het weergeven van de dynamiek over de tijd op zowel individueel teamlid niveau als het niveau van het gehele team.

Kortom, de kennis van dit rapport en het bijbehorende evaluatieraamwerk en meetinstrument kunnen worden gebruikt voor het doen van een breed scala aan (militaire) effectstudies. Dat kan betrekking hebben op effecten van leerinterventies, maar ook (snel) onderzoek naar het verloop van teamfunctioneren door de tijd of effecten van allerlei andere factoren die van invloed kunnen zijn op teameffectiviteit. Hierbij kan worden gedacht aan effecten van verandering van taken, team-samenstelling, nieuwe (geavanceerde) opleidingsleermiddelen, omgevingsfactoren, structuur van de trainingsorganisatie, nieuwe inhoudelijke inrichting van trainings-trajecten en combinaties hiervan. Voor dit type onderzoek kan *TeamScan* door research-instellingen worden gebruikt voor het longitudinaal meten en analyseren van effecten hiervan op teamfunctioneren en teameffectiviteit.

7 Referenties

- Aiken, L.R. (1996). *Rating scales and checklists. Evaluating behavior, personality and attitudes*. New York: Freeman and company.
- Akl, E.A., Pretorius, R.W., Sackett, K., Scott Erdley, W. Bhoopath, P., Alfarah, Z., & Nemann, H.J. (2010). The effect of educational games on medical students' learning outcomes: a systematic review. *BEME guide no 14*.
- Alliger, G. M., & Janak, E. A. (1989). Kirkpatrick's levels of training criteria: thirty years later. *Personnel Psychology, 42*, 331–342.
- Alliger, G. M., Tannenbaum, S. I., Bennett, W., Traver, H., & Shotland, A. (1997). A meta-analysis of the relations among training criteria. *Personnel Psychology, 50*, 341–358.
- Alvarez, K., Salas, E., & Garofano, C.M. (2004). An integrated model of training evaluation and effectiveness. *Human resource development review 3(4)*, 385-416.
- Appleton, J. , Christenson, S. L., Kim, D., & Reschly, A.L. (2006). Measuring cognitive and psychological engagement: validation of the student engagement instrument. *Journal of School Psychology, 44*, 427-445.
- Baldwin, T. T., & Ford, J. K. (1988). Transfer of training: a review and directions for future research. *Personnel Psychology, 41*, 63-105.
- Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York: Freeman.
- Barnard, L., Lan, W.Y., Yen, M.T., Osland Paton, V., & Lai, S. (2009). Measuring self-regulation in online and blended learning environments. *Internet and higher education, 12*, 1-6.
- Barnard, Y.F., Veldhuis, G.J., & Rooij, C.G.M. van (2001). Evaluation in practice: identifying factors for improving transfer of training in technical domains. *Studies in Educational Evaluation, 27*, 269-290.
- Bates, R. (2004). A critical analysis of evaluation practice: the Kirkpatrick model and the principle of beneficence. *Evaluation and Program Planning 27*, 341–347.
- Bonnstetter B, Suiter J, Widrick, R. (2001) The universal Language DISC. A reference manual. Scottsdale, AZ: Target Training International, 1993.
- Brouwer, A.M., Korteling, J.E. (2012). *Neuroscience for selection and training of military personel*. Report TNO-DV 2012 A020. Soesterberg: TNO Defense Security and Safety.
- Campbell, D.T. & Stanley J.C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago, Illinois: Rand McNally & Company.
- Cannon-Bowers, J. A., Salas, E., Tannenbaum, S. I., & Mathieu, J. E. (1995). Toward theoretically based principles of training effectiveness: a model and initial empirical investigation. *Military Psychology, 7*, 141–164.
- Cannon-Bowers, J. A., Tannenbaum, S. I., Salas, E., & Volpe, C. E. (1995). Defining team competencies and establishing team training requirements. In R. Guzzo & E. Salas (Eds.), *Teams: Their training and performance* (pp. 101-124). Norwood, NJ: Ablex.
- Caro, P.W. (1977). *Some factors influencing air force simulator training effectiveness* HUMRRO Technical Report tr-77-2. Alexandria, Virginia: Human Resources Research Organisation.

- Cohn, J., Kay, S., Milham, L., Bell Carroll, M., Jones, D. Sullivan, J., & Darken, R. (2009). Training effectiveness evaluation: from theory to practice. In D. Schmorow, J. Cohn, & D. Nicholson (Eds). *The PSI Handbook of Virtual Environments for Training and Education*. Pp 157-172.
- Connolly T, Stansfield M., & Hainey, T. (2009). Towards the development of a game-based learning evaluation framework. In: *Games-based learning advancements for multisensory human computer interfaces: Techniques and effective practices*.
- Digman, J. M. (1990). Personality structure: Emergence of the five factor model. *Annual Review Psychology* , 41, 417-440.
- Driskell, J. E., & Salas, E. (1992). Collective behaviour and team performance. *Human Factors*, 34, p. 277-288.
- Essens, P., Vogelaar, A., Mylle, J., Blendell, C., Paris, C., Halpin, S., et al., (2005). *Military command team effectiveness: Model and instrument for assessment and improvement*. Soesterberg: Nato HF.
- Fisher, M.J., & King, J. (2009). The self-directed learning readiness scale for nursing education revisited: a confirmatory factor analysis. *Nurse Education Today*.
- Fredericks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74, 59– 109.
- Freeman, J., Stacy, W., Olivares, O. (2009) Measurement and assessment for training in virtual environments. In: D. Schmorow, J. Cohn & D. Nicholson (Eds), *The PRI Handbook of Virtual Environments for Training and Education Vol 1 (Learning requirements and metrics)*. Westport, Connecticut – London: Praeger Security International. pp 236-250.
- Gipps, C.V. (1994). *Beyond testing: Towards a theory of educational assessment*. London-Washington DC: The Falmer Press.
- Guion, R.M. (1998). *Assessment, measurement and prediction for personnel decisions*. Mahwah NJ: Lawrence Erlbaum Associates.
- Harteveld, C. (2011). *Triadic game design: Balancing reality, meaning and play*. London, UK: Springer.
- Hays, R. T. (2005). *The effectiveness of instructional games: a literature review and discussion*. Technical Report 2005-004. Naval Air Warfare Training Systems Division. Orlando, U.S.A.
- Hockey, G.R. (1997). Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. *Biological psychology*, 45, 73-93.
- Holton, E.F. (1996). The flawed four-level evaluation model. *Human resource development quarterly* 7, 5-21.
- Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From input-process-output models to IMOI models. *Annual Review Psychology*. 56, 517–543.
- Kirkpatrick, D. L. (1976). Evaluation of training. In R. L. Craig (Ed.), *Training and development handbook: A guide to human resource development*. New York: McGraw Hill.
- Kirkpatrick, D.I. (2006). *Evaluating Training Programs: The Four Levels* 3rd edition San Francisco: Berrett-Koehler.

- Korteling, J.E., & Oprins, E. (2012). True evidence for transfer of training of serious gaming. *GATE Magazine*, 2012.
- Korteling, J.E., Emmerik, M.L. van (2013). *Optimale inzet van simulatiemiddelen voor O&T bij Defensie*. Rapport TNO 2013 R10590. Soesterberg: Behavioural & Societal Sciences.
- Korteling, J.E., Helsdingen, A.S., & Theunissen, N.C.M. (2013). Serious Games @ Work: Learning job-related competences using serious gaming. In D. Derks & A. Bakker & (Eds) *The Psychology of Digital Media at Work*. Psychology Press LTD / Taylor & Francis Group. Pp 123-144.
- Korteling, J.E., Oprins, E.A.P.B., & Kallen, V.L. (2013). Measurement of Training Effectiveness for training simulations. In Z. Whang (Ed). RTO-MP-SAS-095 *Cost-Benefit Analysis of Military Training* Paper presented at the SAS Workshop held in Amsterdam, the Netherlands, June 5-6, 2012 [http://www.cso.nato.int/abstracts.aspx].
- Korteling, J.E., Jetten, A.M. & Pieters, S. (2014). *Measurement of military mental competences*. Report TNO-in press. Soesterberg: Behavioral & Societal Sciences.
- Kraiger, K. (2002). Decision-based evaluation. In: K. Kraiger (Ed), *Creating, implementing and managing effective training and development* (pp. 331-375). San Francisco CA: Jossey-Bass.
- Kraiger, K., Ford, J.K., Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. *Journal of applied Psychology* 78 (2), 311-328.
- Krokos, K. J., Baker, D. P., Alonso, A., & Day, R. (2009). Assessing Team Processes in Complex Environments: Challenges in Transitioning Research to Practice. In E. Salas, G. F. Goodwin, & C. S. Burke, *Team effectiveness in complex organizations* (pp. 383-408). New York: Routledge.
- Lisa M. Schilling, MD, John F. Steiner, MD, MPH, Kristy Lundahl, MBA, and Robert J. Anderson, MD. *Academic Medicine*, Vol. 80, No. 1 / January 2005, p. 51-56.
- Marks, M. A., Sabella, M. J., Burke, C. S., & Zaccaro, S. J. (2002). The impact of cross-training on team effectiveness. *Journal of Applied Psychology*, 87(1), 3-13.
- Mayer, I. (2012). Towards a comprehensive methodology for the research and evaluation of serious games. *Procedia Computer Science* 15, p. 233 – 247.
- Merrienboer, J.J.G. van, & Kirschner, P. (2007). *Ten steps to complex learning: a systematic approach to four-component instructional design*. Mahwah NJ: Lawrence Erlbaum Associates.
- Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
- Mills, M.C., Pascual, R.G., Blendell, C., Verral, M.G. (1999). *Supporting distributed and ad-hoc team interaction: annex B leading distributed and ad hoc team (the team work guide)*, London: DERA.
- O'Neil, H. F., Wainess, R., & Baker, E. L. (2005). Classification of learning outcomes: evidence from the computer games literature. *The Curriculum Journal*, 16(4), 455–474. doi: 10.1080/09585170500384529.
- Oprins, E. (2008). *Design of a competence-based assessment system for air traffic control training*. Doctoral dissertation, Maastricht University.
- Oprins, E.A.P.B. & Korteling, J.E. (2014). Transfer of training of an instructional serious game: The effectiveness of the Cashier Trainer, In: Y. Cai and S. L. Goei

- (eds.), *Simulations, Serious Games and Their Applications, Gaming Media and Social Effects*. Singapore: Springer Science+Business Media DOI: 10.1007/978-981-4560-32-0_15.
- Oprins, E., Bakhuis Roozeboom, M., Visschedijk, G., Kistemaker, L. (2013). *Effectiviteit van serious gaming in het onderwijs*. TNO rapport (R10415). TNO: Soesterberg.
- Pervin, L. A. (2003). *The science of personality* (2 ed.). New York: Oxford university press.
- Pintrich, P.R., & Groot, W.V. de (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of educational psychology, 82*(1), 33-40.
- Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion, 28*(2), 147–169.
- Reschly, A., & Christenson, S. L. (2006). Prediction of dropout among students with mild disabilities: a case for the inclusion of student engagement variables. *Remedial and Special Education, 27*(5), pp. 276 - 292.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological Correlates of University Students' Academic Performance: A Systematic Review and Meta-Analysis. *Psychological Bulletin 138* (2), 353–387.
- Roe, R.A. (2002). Competenties: een sleutel tot integratie in theorie en praktijk van de A&O psychologie [Competences – A key towards integration in theory and practice in W&O Psychology]. *Gedrag en Organisatie, 15*(4), 203 – 224.
- Roe, R.A. (2005). The design of selection systems: contexts, principles, issues. In A. Evers, O. Smit, & N. Anderson (Eds.), *Handbook of personnel selection* (pp. 73-97). Oxford: Blackwell.
- Russell, V. J., Ainley, M., & Frydenberg, E. (2005). *Schooling issues digest: Student motivation and engagement*.
[Http://www.dest.gov.au/sectors/school_education/publications_resources/schooling_issues_digest/schooling_issues_digest_motivation_engagement.htm](http://www.dest.gov.au/sectors/school_education/publications_resources/schooling_issues_digest/schooling_issues_digest_motivation_engagement.htm).
- Salas, E., Milham, L.M., & Bowers, C.A. (2003). Training evaluation in the military: misconceptions, opportunities, and challenges. *Military Psychology, 15*, 3-16.
- Salas, E., Sims, D. E., & Burke, C.S. (2005). Is there a "big five" in teamwork? *Small Group Research, 36*(5), p. 555-599.
- Salas, E., Rosen, M. A., Burke, C. S., & Goodwin, G. F. (2009). The Wisdom of Collectives in Organizations: An Update of the Teamwork Competencies. In E. Salas, G. F. Goodwin, & C. S. Burke, *Team effectiveness in complex organizations* (p. 39-79). New York: Routledge.
- Seng, K.H., & Fraser, B.J. (2008). Using classroom psychosocial environment in the evaluation of adult computer courses in Singapore. *Technology, pedagogy and education, 17*(1), 67-81.
- Sitzmann, T. (2011). A meta-analytic examination of the instructional effectiveness of computer-based simulation games. *Personnel Psychology, 64*, 489-528.
- Stewart, G. L., & Manz, C. C. (1995). Leadership for self-managingwork teams: a typology and integrative model. *Human Relations, 48*, 747-770.
- Stubbé, H.M., Theunissen, N.C.M. (2008). Self-directed learning in a ubiquitous learning environment: a meta-review. *Proceedings of Special Track on Technology Support for Self-Organised Learners 2008*, pp.5-28.

- Tams, S. (1008). Self-directed social learning: the role of individual differences. *Journal of management development*, 27(2), 196-213.
- Tannenbaum, S., Cannon-Bowers, J., Salas, E., & Mathieu, J. (1993). *Factors that influence training effectiveness: A conceptual model and longitudinal analysis* (Tech. Rep. No. 93-011). Orlando, FL: Naval Training Systems Center, Human Systems Integration Division.
- Venrooij, W. (2009). *Development of the Critical Teamwork Competence Questionnaire*. Master Thesis TU/e, Eindhoven.
- Zaccaro, S. J., Marks, M. A., O'Connor-Boes, J., & Costanza, D. (1995). *The nature and assessment of leader mental models*. 95-3 Bethesda, Management Research Institute.
- Zijlstra, F.R.H. (1993). *Efficiency in work behaviour: a design approach for modern tools*. Doctoral dissertation, University of Delft.

REPORT DOCUMENTATION PAGE

(MOD-NL)

1. DEFENCE REPORT NO (MOD-NL)	2. RECIPIENT'S ACCESSION NO	3. PERFORMING ORGANIZATION REPORT NO
-		TNO 2014 R10243
4. PROJECT/TASK/WORK UNIT NO	5. CONTRACT NO	6. REPORT DATE
053.01961	-	February 2014
7. NUMBER OF PAGES	8. NUMBER OF REFERENCES	9. TYPE OF REPORT AND DATES COVERED
56 (incl 1 appendices, excl RDP & distribution list)	72	Final
10. TITLE AND SUBTITLE		
Evaluatie van leerinterventies en teamfunctioneren in dynamische teams		
11. AUTHOR(S)		
Dr. J.E. Korteling, Dr. E.A.P.B. Oprins, Ir. W. Venrooij.		
12. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)		
TNO, P.O. Box 23, 3769 ZG Soesterberg, The Netherlands Kampweg 5, Soesterberg, The Netherlands		
13. SPONSORING AGENCY NAME(S) AND ADDRESS(ES)		
Ministry of Defence, The Hague, The Netherlands		
14. SUPPLEMENTARY NOTES:		
The classification designation Ongerubriceerd is equivalent to Unclassified, Stg. Confidentieel is equivalent to Confidential and Stg. Geheim is equivalent to Secret.		
15. ABSTRACT (MAXIMUM 200 WORDS (1044 BYTE))		
<p>Dit rapport betreft het verder ontwikkelen van methoden voor het evalueren van samengestelde dynamische teams op efficiëntie en effectiviteit en voor het bepalen van de leeropbrengst van leerinterventies. In dit verband levert het rapport op basis van de literatuur een generiek evaluatieraamwerk voor het meten van de effectiviteit van leerinterventies en de theoretische achtergrond hiervan. Deze kennis kan worden gebruikt voor evalueren van leerinterventies en het meten van transfer of training en voor het doen van evaluatie- en effectonderzoek. Het raamwerk is niet alleen bedoeld voor het meten van interventie-effecten, maar vooral ook op het krijgen van inzicht in de onderliggende processen van die effecten. Daarnaast levert het informatie over constructen die gemeten moeten worden, issues t.a.v. verschillende typen meetmethoden en kwaliteitscriteria, zoals de betrouwbaarheid en validiteit van proces- en prestatie-maten. Vervolgens wordt dieper ingegaan op methode-ontwikkeling voor het daadwerkelijk evalueren van dynamische teams. Op basis van literatuuronderzoek, o.a. naar modellen voor team-functioneren en team adaptiviteit, wordt een meetinstrument, getiteld <i>TeamScan</i> hiervoor ontwikkeld en beproefd. Dit meetinstrument is te gebruiken voor zowel het evalueren van het resultaat van team-trainingsinterventies als van het evalueren van teamfunctioneren on-the job. Door de relatief eenvoudige wijze waarop longitudinale, herhaalde metingen kunnen uitgevoerd, kunnen juist de dynamische aspecten van teams door de tijd heen goed in kaart worden gebracht. Tot slot werd <i>TeamScan</i> als prototype d.m.v. pilots i.s.m. andere projecten binnen het programma beproefd.</p>		
16. DESCRIPTORS	IDENTIFIERS	
Training, measurement, transfer of training, teams, training interventions, dynamic teams, military training	Evaluation framework, Teamscan, dynamic teams	
17a. SECURITY CLASSIFICATION (OF REPORT)	17b. SECURITY CLASSIFICATION (OF PAGE)	17c. SECURITY CLASSIFICATION (OF ABSTRACT)
Ongerubriceerd	Ongerubriceerd	Ongerubriceerd
18. DISTRIBUTION AVAILABILITY STATEMENT	17d. SECURITY CLASSIFICATION (OF TITLES)	
Subject to approval MOD-NL	Ongerubriceerd	

Distributielijst

Onderstaande instanties/personen ontvangen een volledig exemplaar van het rapport.

- 1 DMO/SC-DR&D
standaard inclusief digitale versie bijgeleverd op cd-rom
- 2/3 DMO/DR&D/Kennistransfer
- 4/5 Programmabegeleider Defensie,
Afdeling Operationele Gereedstelling / Team Opleiden
Directie Aansturen Operationele Gereedstelling
Majoor J.F.M. Bosman
LKol R. Pleijsant.
- 6 Projectbegeleider Defensie,
J.F.J. Bor, Kromhoutkazerne, Utrecht
- 7 Bibliotheek KMA
- 8 Programmaleider TNO, vestiging Soesterberg
Dr N.C.M. Theunissen
- 9 TNO, vestiging Soesterberg,
(Archief)
- 10 TNO Business Line Manager Human Effectiveness, vestiging Soesterberg
Prof dr A. Bronkhorst
- 11/14 TNO, vestiging Soesterberg,
Dr J.E. Korteling
Dr. E.A.P.B. Oprins
Drs. W. Venrooij
Drs. W.S.M. Piek