

De Nationale Studie
Pedagogische Kwaliteit Kinderopvang
Eindrapport Project 0 en 1

R.E.L. de Kruif, H.J. Vermeer, R.G. Fukkink, J.M.A. Riksen-Walraven, L.W.C.
Tavecchio, M.H. van IJzendoorn, & J. van Zeijl

Subsidiegever:

Ministerie van Sociale Zaken en Werkgelegenheid

Uitvoerder:

Nederlands Consortium Kinderopvang Onderzoek



Januari 2007

INHOUDSOPGAVE

	Voorwoord	1
	Samenvatting	3
I	Inleiding	9
II	Enquête onder direct betrokkenen bij de Nederlandse kinderopvang	11
	1 Inleiding	11
	2 Methode	15
	3 Resultaten	17
	4 Conclusie	22
III	Ontwikkeling en validering van de NCKO-leidsterschalen en NCKO- kindschalen	24
	1 Inleiding	24
	2 Methode	34
	3 Resultaten	45
	4 Discussie en conclusies	57
IV	Slotwoord	64
	Referenties	66

VOORWOORD

Het in dit rapport beschreven onderzoek werd uitgevoerd in verschillende fasen en met verschillende groepen respondenten die wij op deze plaats graag willen bedanken voor hun deelname. In de eerste plaats gaat onze dank uit naar de ouders, leidsters, leidinggevend en 'externe' deskundigen die ons door middel van een enquête lieten weten welke onderdelen van het NCKO-kwaliteitsmodel vanuit hun perspectief belangrijk zijn bij het meten van de pedagogische kwaliteit in de kinderopvang. De validering van onderdelen van het NCKO-instrumentarium zou onmogelijk geweest zijn zonder deelname van de leidinggevend en, leidsters en ouders van de kinderdagverblijven die we bezochten voor trainingsdoeleinden, voor het uitproberen van de procedures in een voorstudie, en voor het verzamelen van de daadwerkelijke data in de hoofdstudie. De leidinggevend en danken we speciaal voor de hulp bij de organisatie rondom de bezoeken van de observatoren. De leidsters zijn we veel dank verschuldigd voor het zeer bereidwillig toelaten van de verschillende observatoren in de groep tijdens twee verschillende bezoeken. Daarnaast maakten zij tijd vrij voor een vraagesprek aan het einde van de observatie en voor het invullen van een uitgebreide vragenlijst. De ouders in deze kinderdagverblijven willen we bedanken voor de toestemming die zij verleenden voor het filmen van hun kind(eren) tijdens ons bezoek.

We danken drs. E.J. Reiling voor het verzorgen van de ITERS-R training en willen daarnaast de volgende medewerkers bedanken voor hun bijdrage aan de organisatie en dataverzameling van dit onderzoek: Elsbeth van Barneveld, Ciska van Buuren, Alev Ceylan, Annemieke Constandse, Marieke Hammerstein, Katrien Helmerhorst, Mathilde van Houdt, Anna Lont, Emma van Manen, Marloes van der Meer, Martine Meesters, Sandra Scholten, Judith Schellekens, Canan Temur, Jourlilete Willems, Tine Willekens en Su'en Kwok. Zij gingen in alle vroegte op pad om te observeren of te filmen in de kinderdagverblijven en/of zaten urenlang achter een tv-scherm om video-opnames van leidsters en kinderen te beoordelen.

Dit project werd uitgevoerd door het NCKO, bestaande uit prof. dr. L.W.C. Tavecchio, prof. J.M.A. Riksen-Walraven, prof. dr. M.H. van IJendoorn, dr. R.E.L. de Kruif, dr. H.J. Vermeer, dr. R.G. Fukkink en dr. J. van Zeijl.

Deze studie is financieel mede mogelijk gemaakt door het Ministerie van Sociale Zaken en Werkgelegenheid.

SAMENVATTING

Project 0

In een eerste studie (Project 0) heeft het NCKO de resultaten van eerder wetenschappelijk onderzoek op het gebied van de kinderopvang in Nederland geïnventariseerd. In deze studie, waarvan de resultaten zijn neergelegd in het boek *De kwaliteit van de Nederlandse kinderopvang* (Van IJzendoorn, Tavecchio, & Riksen-Walraven, 2004), brachten we bestaande kennis over de kwaliteit van kinderopvang in Nederland systematisch in kaart, omschreven we wat we bedoelen met ‘kwaliteit van de kinderopvang’ en evalueerden we diverse instrumenten voor het meten van kwaliteit in de kinderopvang. Uit deze studie bleek dat er maar heel weinig onderzoek is gedaan naar de kwaliteit van de Nederlandse kinderopvang, naar factoren die de kwaliteit kunnen bevorderen en naar de invloed die kinderopvang van hogere of lagere kwaliteit heeft op de ontwikkeling van Nederlandse kinderen. Ook ontbrak een adequaat instrument voor het meten van de kwaliteit van de Nederlandse kinderopvang.

Het ontwikkelen van een goed instrument vereist allereerst duidelijkheid met betrekking tot het te meten *begrip*: wat verstaan we onder ‘pedagogische kwaliteit’? Om deze vragen te beantwoorden heeft het NCKO een *kwaliteitsmodel* voor de kinderopvang ontwikkeld, dat elders uitgebreid is beschreven (Riksen-Walraven, 2004). In dit model worden twee typen kwaliteitsindicatoren onderscheiden. Op de eerste plaats gaat het om indicatoren van de *proceskwaliteit* van de opvang, ofwel de kwaliteit van het feitelijke zorg- en opvoedingsproces en van de directe ervaringen van de kinderen. Daarnaast worden er *structurele* kwaliteitsindicatoren onderscheiden, dat wil zeggen regelbare kenmerken van de kinderopvang, zoals groeps grootte, leidster-kind ratio en opleiding/training van de leidsters, die van invloed zijn op de proceskwaliteit en beschouwd kunnen worden als voorwaarden voor het realiseren van een hoge proceskwaliteit. Aangenomen wordt dat de in het model opgenomen proces- en structurele kwaliteitsindicatoren samen een genuanceerd en compleet beeld geven van de pedagogische kwaliteit van de opvang in een kinderdagverblijf en dat beide groepen kwaliteitsindicatoren daarom zouden moeten worden opgenomen in een instrument om de pedagogische kwaliteit van de opvang te meten.

Een op het NCKO-kwaliteitsmodel gebaseerd meetinstrument voor de pedagogische kwaliteit van de kinderopvang moet zich zowel richten op de proceskwaliteit als op structurele kwaliteitskenmerken. Bij de meting van de proceskwaliteit zou vooral aandacht besteed moeten

worden aan zes interactievaardigheden van leidsters (sensitieve responsiviteit, respect voor de autonomie van de kinderen, structureren en grenzen stellen, praten en uitleggen, ontwikkelingsstimulering en het begeleiden van interacties tussen kinderen) en aan twee kenmerken van de kinderen (welbevinden en betrokkenheid). Als belangrijke structurele kwaliteitskenmerken werden in het NCKO-kwaliteitsmodel tien kenmerken genoemd, waaronder pedagogisch beleid, pedagogische ondersteuning, leidster-kind ratio, groepsgrootte, en leidster- en groepsstabiliteit.

Project 1

Het NCKO heeft zich in de tweede fase van het onderzoek (Project 1) gericht op de ontwikkeling en validering van een standaard-instrument om de pedagogische kwaliteit van kinderopvang op wetenschappelijk verantwoorde wijze te meten.

Enquête

Het NCKO-kwaliteitsmodel reflecteert de visie van wetenschappers op pedagogische kwaliteit en hoe die het beste gemeten kan worden. Een belangrijke vraag is in hoeverre deze visie wordt gedeeld door andere betrokken partijen zoals ouders, leidinggevenden en leidsters in de kinderopvang. Alvorens te starten met de ontwikkeling van een meetinstrument op basis van het model heeft het NCKO daarom eerst een *enquête* gehouden onder verschillende groepen betrokkenen om na te gaan of de in het model opgenomen kwaliteitskenmerken ook in bredere kring als belangrijk worden beschouwd.

De enquête, die vooraf ging aan de constructie van het instrumentarium, liet zien dat de diverse betrokkenen het belang onderschrijven van de verschillende onderdelen die in het model onderscheiden worden. Ouders, leidsters, leidinggevenden en ‘externe’ deskundigen (563 in totaal in de steekproef) hebben opvallend gelijkgestemde oordelen over wat belangrijk is bij het meten van de kwaliteit van de kinderopvang en wat opgenomen dient te worden in een kwaliteitsinstrument.

Ontwikkeling instrument

De volgende fase in het NCKO-onderzoek betrof de ontwikkeling van het instrument voor de bepaling van de pedagogische kwaliteit van kinderopvang. Op basis van het NCKO-kwaliteitsmodel (zie Riksen-Walraven, 2004) is een breed instrumentarium ontwikkeld voor het

meten van de pedagogische kwaliteit van de kinderopvang. In dit instrument is aandacht voor structurele kwaliteitskenmerken, proceskwaliteit en kenmerken van het kind in het kinderdagverblijf. De leidsterschalen meten de sensitieve responsiviteit van de leidster, respect voor autonomie van de kinderen, structureren en grenzen stellen, praten en uitleggen, ontwikkelingsstimulering en de begeleiding van de interacties tussen de kinderen. Met de kindschalen wordt het welbevinden en de betrokkenheid van de kinderen in het kinderdagverblijf bepaald.

Valideringsstudie

De valideringsstudie had tot doel het toetsen van de betrouwbaarheid en validiteit van de NCKO-leidster- en kindschalen. Het ontwikkelde NCKO-instrumentarium is hiertoe beproefd bij 145 leidsters en 623 kinderen uit 72 groepen en 46 kinderdagverblijven. In dit onderzoek hebben we drie veelgebruikte *betrouwbaarheids*indicatoren berekend, namelijk:

- (1) De *inter-observator* betrouwbaarheid van de schalen ofwel de mate waarin verschillende (getrainde) observatoren het eens zijn bij het toekennen van scores op de schalen op basis van dezelfde observatie-episode.
- (2) De *test-hertest* betrouwbaarheid van de schalen, ofwel de mate waarin observatie op verschillende meettijdstippen dezelfde schaalscores oplevert. Hierbij wordt ervan uitgegaan dat het gemeten kenmerk (bijvoorbeeld de sensitieve responsiviteit van de leidster) relatief stabiel is over de tijd, zodat inconsistentie van de observatie-uitkomsten over tijd inderdaad duidt op een gebrek aan precisie van het meetinstrument, en niet op een werkelijke verandering in het geobserveerde fenomeen. In deze valideringsstudie is gekozen voor twee metingen met een tussentijd van ongeveer drie maanden.
- (3) De *interne consistentie* van de schalen, ofwel de mate waarin de schalen onderling samenhangen. Omdat we ervan uitgaan dat de zes leidsterschalen verschillende aspecten van hetzelfde begrip meten, namelijk “interactievaardigheden”, werd verwacht dat de scores van de leidsters op de verschillende schalen onderling significant zouden samenhangen. Omdat het om *verschillende* aspecten van hetzelfde begrip gaat, werd echter geen erg hoge correlatie verwacht. Om dezelfde reden werd een significante doch matige correlatie verwacht tussen de twee kindschalen.

De *validiteit* van een meetinstrument geeft aan of het instrument inderdaad datgene lijkt te meten wat het beoogt te meten. In deze valideringsstudie hebben we ons gericht op drie typen validiteit, namelijk:

-
- (1) De *convergente validiteit* van de schalen, ofwel de mate waarin de scores op de schalen samenhangen (“convergeren”) met scores die verkregen zijn met behulp van andere instrumenten die hetzelfde pretenderen te meten. Om de convergente validiteit van de NCKO-leidsterschalen te bepalen hebben we de scores van de leidsters op deze schalen gerelateerd aan inhoudelijk vergelijkbare scores die op basis van dezelfde observatie-episodes toegekend zijn met behulp van drie andere veelgebruikte instrumenten voor beoordeling van de kwaliteit van leidstergedrag. De convergente validiteit van de NCKO-kindschalen voor welbevinden en betrokkenheid is bepaald door de scores op deze schalen te relateren aan scores die op basis van dezelfde observatie-episodes zijn toegekend met behulp van twee andere inhoudelijk vergelijkbare meetinstrumenten. Zeer hoge correlaties werden niet verwacht, omdat de NCKO-schalen juist ontworpen zijn om net iets andere leidster- en kindkenmerken te meten dan de bestaande schalen.
 - (2) De *discriminante validiteit* van de schalen, ofwel de mate waarin zij zich onderscheiden van meetinstrumenten die *iets anders* pretenderen te meten. Om de discriminante validiteit van de NCKO-leidster- en kindschalen te bepalen, hebben we de schaalscores gerelateerd aan bepaalde temperamentkenmerken van de leidsters en kinderen die theoretisch niet gerelateerd zouden moeten zijn aan datgene wat de NCKO-schalen meten.
 - (3) De *predictieve validiteit* van de schalen, ofwel het vermogen van de schalen om iets te voorspellen wat ze theoretisch zouden moeten voorspellen. In de valideringsstudie zijn verbanden onderzocht om de predictieve validiteit van de NCKO-schalen te toetsen.

De leidsterschalen blijken betrouwbaar (gelet op de inter-observator betrouwbaarheid, test-hertest betrouwbaarheid of stabiliteit, en de onderlinge samenhang) en valide (gelet op de convergente validiteit, discriminante c.q. onderscheidende validiteit, en predictieve validiteit). Eén subschaal, *structureren en grenzen stellen*, scoort onvoldoende op één betrouwbaarheidsaspect, namelijk de inter-observator betrouwbaarheid. De valideringsstudie laat verder zien dat de schaal voor het beoordelen van het welbevinden van de kinderen in een groep betrouwbaar is gelet op de inter-observator betrouwbaarheid en de onderlinge samenhang (namelijk met betrokkenheid), maar laag scoort op stabiliteit. De validiteit van deze schaal is redelijk. De kindschaal voor het meten van de betrokkenheid is betrouwbaar en valide.

De ervaringen en uitkomsten van de valideringsstudie zijn ten slotte aangewend voor de (verdere) aanpassing van de instrumenten. Voor de onderdelen die niet bleken te voldoen aan de gestelde valideringscriteria, is geïnventariseerd welke alternatieven beschikbaar zijn.

Conclusie

De valideringsstudie heeft een betrouwbaar en valide instrument voor toekomstig onderzoek opgeleverd dat beproefd is in de praktijk van de Nederlandse kinderopvang. Dit instrument legt daarmee een fundament voor het meten van de pedagogische kwaliteit van de kinderopvang in Nederland. Het ontwikkelde instrumentarium maakt het mogelijk structurele kwaliteitskenmerken, proceskwaliteit en kindkenmerken te bepalen én deze met elkaar in verband te brengen, recht doend aan het kwaliteitsmodel voor de kinderopvang en de opvattingen die leven in het Nederlandse veld. Ook is het mogelijk nieuwe bevindingen uit toekomstig onderzoek te (blijven) verbinden met uitkomsten uit reeds uitgevoerd peilingsonderzoek in Nederland. De bevindingen uit toekomstige peilingen kunnen dus worden afgezet tegen de vorige peilingen uit 1995, 2001 en de kwaliteitsmeting uit 2005 (zie Vermeer e.a., 2005) om een vergelijking te maken in de tijd. Wel moet worden opgemerkt dat nieuwe metingen met een nieuw en breder instrument een completer – en in die zin: een ander – beeld zullen geven van de pedagogische kwaliteit van kinderdagverblijven. Het blijft tevens mogelijk de kwaliteit van de Nederlandse kinderopvang op kernaspecten te spiegelen aan die van het buitenland, met inachtneming van de unieke kenmerken van de Nederlandse kinderopvangsituatie. Concrete stappen worden op dit moment gezet om een handzaam profiel op te kunnen stellen op basis van het complete instrumentarium. Deze uitsnede uit het omvangrijke instrument moet het mogelijk maken gericht te kunnen werken aan kwaliteitsbepaling en eventuele verbetering van de kernkwaliteit van kinderdagverblijven. In dit profiel wordt aandacht voor structurele kwaliteit gecombineerd met aandacht voor de proceskwaliteit, zoals gemeten met de NCKO-leidster- en kindschalen. Om een breder beeld te verkrijgen van het kinderdagverblijf wordt het profiel verder aangevuld met schalen uit de vertaalde en bewerkte ITERS-R/ECERS-R (zoals Meubilering/inrichting, Activiteiten, Taal, Interacties en Programma).



I. INLEIDING

In het samenwerkingsverband *Nederlands Consortium Kinderopvang Onderzoek* (NCKO) werken pedagogen en ontwikkelingspsychologen van verschillende Nederlandse universiteiten samen aan onderzoek ter bevordering van de pedagogische kwaliteit van de Nederlandse kinderopvang voor 0- tot 4-jarigen. Het onderzoek, dat gesubsidieerd wordt door het Ministerie van Sociale Zaken en Werkgelegenheid (SWZ), is gestart in 2002 en bestaat uit verschillende fasen.

In een eerste studie heeft het NCKO de resultaten van eerder wetenschappelijk onderzoek op het gebied van de kinderopvang in Nederland geïnventariseerd. Uit deze studie, waarvan de resultaten zijn neergelegd in het boek “De kwaliteit van de Nederlandse kinderopvang” (Van IJzendoorn, Tavecchio, & Riksen-Walraven, 2004), bleek dat er maar heel weinig onderzoek is gedaan naar de kwaliteit van de Nederlandse kinderopvang, naar factoren die de kwaliteit kunnen bevorderen en naar de invloed die kinderopvang van hogere of lagere kwaliteit heeft op de ontwikkeling van Nederlandse kinderen. Omdat de Nederlandse kinderopvang in belangrijke opzichten afwijkt van de opvang in landen waar wél wetenschappelijk onderzoek op dit gebied is gedaan, mag niet zonder meer worden aangenomen dat de resultaten van onderzoek in andere landen ook gelden voor de Nederlandse situatie. Daarom is onderzoek in Nederland noodzakelijk.

Om de pedagogische kwaliteit van de kinderopvang te bevorderen en/of op een hoog peil te houden is het allereerst van belang om die kwaliteit te kunnen *meten*. Een goed meetinstrument maakt het mogelijk om de kwaliteit van de opvang te ‘monitoren’ en waar nodig te verbeteren. Daarom heeft het NCKO zich in de tweede fase van het onderzoek gericht op de ontwikkeling en validering van een standaard-instrument om de pedagogische kwaliteit van kinderopvang op wetenschappelijk verantwoorde wijze te meten. In dit rapport wordt daarvan verslag gedaan.

Het ontwikkelen van een goed instrument vereist allereerst duidelijkheid met betrekking tot het te meten *begrip*: wat verstaan we onder ‘pedagogische kwaliteit’? Zijn er meerdere kwaliteitsaspecten te onderscheiden, hoe verhouden die zich tot elkaar en verdienen sommige aspecten bij de meting meer aandacht dan andere? Om deze vragen te beantwoorden heeft het NCKO een *kwaliteitsmodel* voor de kinderopvang ontwikkeld dat uitgebreid is beschreven in Hoofdstuk 5 van het boek *De kwaliteit van de Nederlandse kinderopvang* (Riksen-Walraven, 2004). In dit model worden twee typen kwaliteitsindicatoren onderscheiden. Op de eerste plaats gaat het om indicatoren van de *proceskwaliteit* van de opvang, ofwel de kwaliteit van het feitelijke zorg- en opvoedingsproces en van de directe ervaringen van de kinderen. Daarnaast

worden er *structurele* kwaliteitsindicatoren onderscheiden, d.w.z. reguleerbare kenmerken van de kinderopvang, zoals groeps grootte, leidster-kind ratio en opleiding/training van de leidsters, die van invloed zijn op de proceskwaliteit en beschouwd kunnen worden als voorwaarden voor het realiseren van een hoge proceskwaliteit. Aangenomen wordt dat de in het model opgenomen proces- en structurele kwaliteitsindicatoren tezamen een genuanceerd en compleet beeld geven van de pedagogische kwaliteit van de opvang in een kinderdagverblijf en dat beide groepen kwaliteitsindicatoren daarom zouden moeten worden opgenomen in een instrument om de pedagogische kwaliteit van de opvang te meten.

Het NCKO-kwaliteitsmodel reflecteert de visie van wetenschappers op pedagogische kwaliteit en hoe die het beste gemeten kan worden. Een belangrijke vraag is in hoeverre deze visie wordt gedeeld door andere betrokken partijen zoals ouders, leidinggevend en leidsters in de kinderopvang. Alvorens te starten met de ontwikkeling van een meetinstrument op basis van het model heeft het NCKO daarom eerst een *enquête* gehouden onder verschillende groepen betrokkenen om na te gaan of de in het model opgenomen kwaliteitskenmerken ook in bredere kring als belangrijk worden beschouwd. De opzet en de uitkomsten van deze enquête staan beschreven in Hoofdstuk II van dit rapport.

Een volgende fase in het NCKO-onderzoek betrof de ontwikkeling van het instrument ter bepaling van de pedagogische kwaliteit van kinderopvang en het uitvoeren van een onderzoek om de psychometrische kwaliteiten van dit nieuwe instrument te beproeven. Dit onderzoek, de zogenaamde *valideringsstudie*, is beschreven in Hoofdstuk III van dit rapport. In dit hoofdstuk worden vervolgens ook de implicaties van de uitkomsten van het valideringsonderzoek voor de verdere ontwikkeling van het instrument besproken en wordt een schets gegeven van de definitieve versie van het meetinstrument. Hoofdstuk IV tenslotte vat de voornaamste conclusies uit dit rapport samen en blikkt vooruit op de volgende fase in het NCKO-onderzoek.

II. ENQUÊTE ONDER DIRECT BETROKKENEN BIJ DE NEDERLANDSE KINDEROPVANG

1. INLEIDING

In het kader van de ontwikkeling en validatie van het NCKO-meetinstrument voor het bepalen van de pedagogische kwaliteit van de kinderopvang is een enquête gehouden onder verschillende groepen directe betrokkenen. Leidinggevendenden, leidsters¹, ouders en externe deskundigen konden hun mening geven in een landelijke enquête. De centrale vraag van deze deelstudie was: welke structurele en proceskenmerken worden door alle betrokken partijen belangrijk gevonden als indicatoren van de pedagogische kwaliteit van de kinderopvang (hier: de hele dagopvang voor kinderen van 0 tot 4 jaar). Het antwoord op deze vraag laat zien in hoeverre het NCKO-kwaliteitsmodel aansluit bij opvattingen over kwaliteit die leven bij diverse betrokkenen, zoals het personeel van kinderdagverblijven, ouders die gebruik maken van de kinderopvang en externe deskundigen.

Over deze deelstudie van het NCKO is elders al uitvoerig gerapporteerd. Corresponderend met de groepen die we hebben onderscheiden in het enquête-onderzoek zijn vier artikelen geschreven, namelijk één artikel voor leidinggevendenden in het blad *Management Kinderopvang* (NCKO, 2004a), één voor ouders in het blad van de oudervereniging BOinK (NCKO, 2004b), één voor leidsters in *KIDDO* (NCKO, 2004c) en ten slotte een laatste wetenschappelijke artikel in *Pedagogiek* voor de groep van externe deskundigen (Fukkink, Tavecchio, de Kruif, Vermeer & van Zeijl, 2006). In dit rapport wordt bondig verslag gedaan van deze deelstudie.

Theoretische achtergrond van de enquête: het NCKO-kwaliteitsmodel

Het NCKO-kwaliteitsmodel voor de kinderopvang (voor een uitgebreide beschrijving zie Riksen-Walraven, 2004) is als uitgangspunt gehanteerd bij de constructie van de enquête over pedagogische kwaliteitscriteria. In dit model zijn verschillende factoren onderscheiden die direct of indirect van invloed zijn op de ontwikkeling van het kind in een kinderdagverblijf (zie Figuur 2.1). Twee typen kwaliteitskenmerken zijn onderscheiden: *proceskenmerken*, die het zorg- en opvoedingsproces zelf beschrijven, en *structurele kenmerken*, die betrekking hebben op aspecten

¹ De term ‘leidster’ wordt gebruikt om naar mannelijk en vrouwelijk personeel te verwijzen.

van de opvoedingssetting en waarvan aangenomen kan worden dat zij een goed zorg- en opvoedingsproces bevorderen.

Het perspectief op pedagogische kwaliteit met aandacht voor proceskwaliteit en structurele kwaliteit is door Riksen-Walraven (2004) aangeduid als een *dubbelfocus-benadering*. Zowel de belangrijkste aspecten van de proceskwaliteit als de structurele factoren, die deze proceskwaliteit distaal - 'op afstand' - beïnvloeden, worden immers in kaart gebracht. Samen geven zij een genuanceerd en compleet beeld van de pedagogische kwaliteit van kinderdagverblijven. Een goede opleiding van de leidster, bijvoorbeeld, draagt bij aan een juiste omgang met de kinderen, maar wanneer de leidster op een te grote groep staat is het niet goed mogelijk die juiste omgang te realiseren, hoe zeer bepaalde basisvaardigheden ook worden beheerst. Kortom, aandacht voor de proceskwaliteit op de groep enerzijds en de randvoorwaarden daarvoor anderzijds vullen elkaar aan en dragen elk bij aan een goede analyse van de pedagogische kwaliteit.

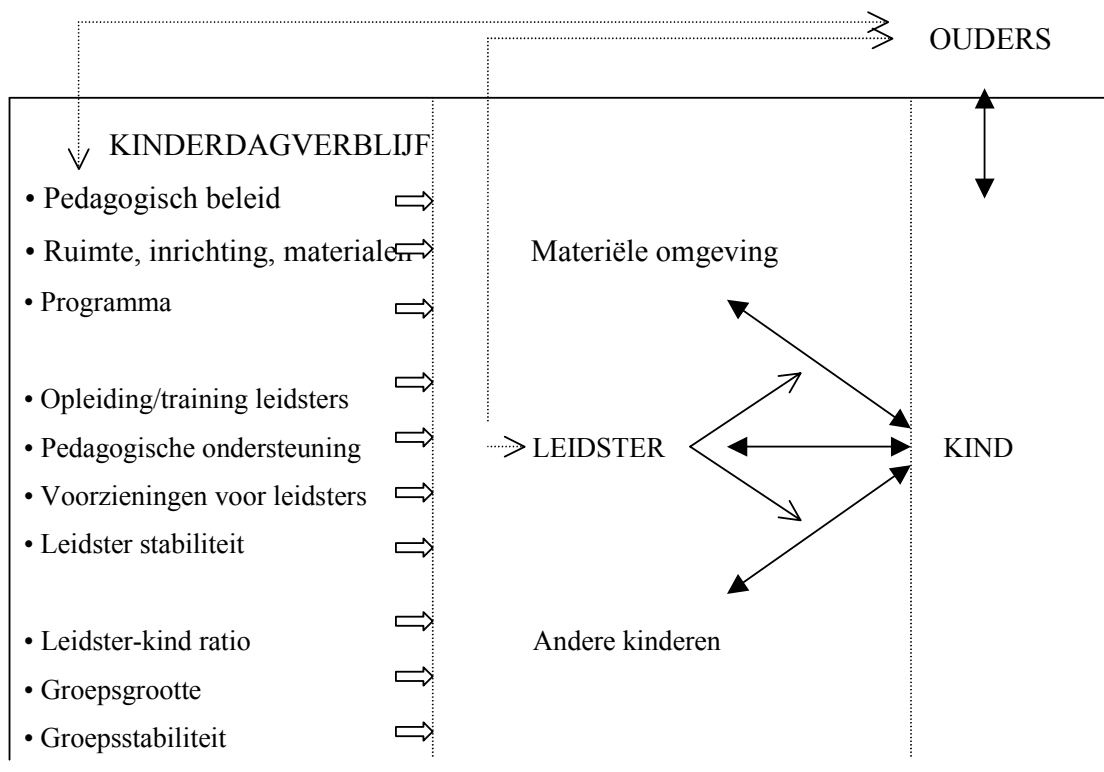
Kenmerken van de proceskwaliteit

Bij de proceskwaliteit staan de directe ervaringen van het kind op het kinderdagverblijf met de leidster en de andere kinderen centraal. Kinderen ontwikkelen zich in interactie met hun omgeving op het kinderdagverblijf, waarbij drie proximale processen onderscheiden kunnen worden. Allereerst is er de interactie van leidster en kind. Daarnaast zijn er interacties tussen het kind en de materiële omgeving en tussen de kinderen onderling. De leidster vormt de spil van de proceskwaliteit, omdat niet alleen haar directe interacties met het kind, maar ook haar invloed op de interacties met de materiële omgeving en van het kind met andere kinderen in hoge mate de ervaringen bepalen die kinderen opdoen op het kinderdagverblijf. Vanwege dit belang worden er in het NCKO-model een zestal interactievaardigheden onderscheiden die naar voren zijn gekomen uit onderzoek naar de invloed van opvoeder-kind interacties op de kinderlijke ontwikkeling. Op de eerste plaats moet een leidster *sensitief* zijn, dat wil zeggen signalen van kinderen opmerken en hier tijdig en adequaat op reageren. Deze sensitieve houding beschermt kinderen tegen stress en bevordert een veilige gehechtheidsrelatie. Daarnaast moet de *autonomie* van een kind gerespecteerd worden met een respectvolle, niet-intrusieve, oftewel niet-'opdringerige', interactiestijl, wat bijdraagt aan het zelfvertrouwen en de eigenwaarde van een kind. Verder is de leidstervaardigheid *structureren en grenzen stellen* van belang, omdat deze zorgt voor een gevoel van veiligheid en competentie bij een kind. *Praten en uitleggen* is ook belangrijk. Een leidster die frequent en op een goede manier verbale interacties heeft met de kinderen, bevordert niet alleen de taalontwikkeling, maar door veel te benoemen en verbaal te

begeleiden wordt ook de cognitieve en sociale ontwikkeling gestimuleerd. *Bevordering van de ontwikkeling* in de vorm van gerichte en passende stimulering bij allerlei dagelijkse activiteiten is een andere belangrijke interactievaardigheid. Ten slotte zou een leidster vaardig moeten zijn in het *begeleiden van de interacties* tussen de kinderen om de unieke sociale omgeving die het kinderdagverblijf biedt ook daadwerkelijk te benutten. Dit betekent onder andere dat een leidster pro-sociaal gedrag aanmoedigt, ingrijpt bij negatieve interacties tussen kinderen, maar ook positief reageert bij positieve interacties.

Naast de zes interactievaardigheden van leidsters zijn in het NCKO-model ook twee kenmerken van de kinderen opgenomen die indicatief kunnen zijn voor de proceskwaliteit van de opvang, namelijk *welbevinden* (zich veilig en ontspannen voelen, genieten) en *betrokkenheid* (gericht, aandachtig en geconcentreerd bezig zijn).

Figuur 2.1
Het NCKO-kwaliteitsmodel



Noot: \longleftrightarrow Proximale processen: ontwikkeling kind in interactie met de directe omgeving
 \Rightarrow Invloed van structurele opvangkenmerken op proximale processen

(Bron: Riksen-Walraven, 2004)

Kenmerken van de structurele kwaliteit

De kenmerken van de structurele kwaliteit (aangegeven in het linkerdeel van Figuur 2.1) betreffen allerlei randvoorwaarden die van directe of indirecte invloed zijn op de dagelijkse verzorging en opvoeding. De groepsgrootte, de leidster-kind ratio en de opleiding van de leidster zijn bekende structurele kenmerken, waarvan het belang voor de pedagogische kwaliteit in diverse empirische studies is aangetoond. Grotere groepen, meer kinderen per leidster en een lagere opleiding hangen samen met een lagere proceskwaliteit. Groepsgrootte blijkt bovendien een negatief effect te hebben dat onafhankelijk is van de leidster-kind ratio. Onder opleiding valt zowel de initiële beroepsopleiding als training daarna. Naast deze drie bekende variabelen, die vanwege hun empirisch gebleken belang in kinderopvangonderzoek ook wel worden aangeduid met de term ‘ijzeren driehoek’ (Tavecchio, 2002), zijn er diverse andere structurele kenmerken van belang voor de pedagogische kwaliteit. Pedagogische ondersteuning van leidsters, bijvoorbeeld gericht op ondersteuning en versterking van leidstervaardigheden en het dagelijkse functioneren op de groep, draagt bij aan het op peil houden en verhogen van de proceskwaliteit. Daarnaast is de stabiliteit van zowel de leidster als van de groep naar voren gekomen als een belangrijke factor. De aanwezigheid van een vertrouwde leidster bevordert bijvoorbeeld de ontwikkeling van een veilige gehechtheidsrelatie tussen leidster en kind en hangt positief samen met het welbevinden van kinderen in een kinderdagverblijf. Het programma, de ruimtelijke inrichting en de materialen zijn met name van belang voor de stimulering op allerlei ontwikkelingsterreinen. Daarnaast is de rol van ouders belangrijk. Ouders hebben allereerst indirect invloed op de opvangkwaliteit door contact met de leidsters bij het brengen en halen van het kind. Daarnaast kunnen zij invloed uitoefenen op de kwaliteit door communicatie met het kinderdagverblijf over het pedagogische beleid. Het kenmerk ‘pedagogisch beleid’, ten slotte, is opgenomen in het model omdat hierin, op basis van een bepaalde pedagogische visie van het kinderdagverblijf, diverse richtlijnen zijn opgenomen die moeten bijdragen aan de kwaliteit. Om recht te doen aan het onderscheid tussen de pedagogische visie en de vertaling daarvan naar het pedagogisch beleid zijn beide aspecten apart bevraagd in de enquête.

2. METHODE

2.1 Steekproef

Er is een landelijke steekproef getrokken, gestratificeerd naar vier doelgroepen die elk een eigen rol in de kinderopvangsector hebben: leidinggevendenden van een kinderdagverblijf, leidsters, ouders en externe deskundigen. De groep leidinggevendenden betreft zowel pedagogen als leiding van kinderdagverblijven, inclusief de zogenaamde ‘competentiemanagers’ die de kinderopvangkoepel Humanitas kent. De leidsters werken op een baby-, peuter- of een verticale groep. Leidinggevendenden, leidsters en ouders zijn in eerste instantie schriftelijk benaderd via kinderdagverblijven.

Voor de enquête is een bestand van in totaal 151 kinderdagverblijven aangelegd door centra te selecteren die vallen onder de drie grote koepels SKON, Catalpa en Humanitas ($n = 84$, 56%) en onder diverse grote kinderopvangaanbieders uit elke provincie van het land ($n = 67$, 44%); in totaal zijn 33 verschillende organisaties betrokken. De 151 geselecteerde kinderdagverblijven ontvingen over de post een pakket met enquêtes voor verschillende groepen: één enquête voor de leidinggevende, drie enquêtes voor de leidster van een baby-, peuter- of verticale groep, en drie enquêtes voor één ouder uit elk van deze groepen. Bij verzending was niet bekend of elk groepstype ook daadwerkelijk voorkwam in het benaderde kinderdagverblijf. Het aantal kinderdagverblijven dat betrokken is in de studie, brengt de leidster- en oudersteekproef in theorie op een maximaal aantal van 453 ($n_{\min} = 151$, $n_{\max} = 453$).

Een elektronische versie van de enquête is verstuurd naar de pedagogen van het Pedagogienplatform, dat 47 leden telt, en naar ouders via de maillijst van Belangenvereniging van Ouders van in de Kinderopvang (BOinK), die zo’n 500 leden telt. Het maximaal aantal bereikte ouders is daarmee – eventuele overlap tussen de groep ouders bij de schriftelijke en elektronische versie buiten beschouwing gelaten – in theorie 953 (453 schriftelijk en 500 via het internet).

Ten slotte is een brede groep van externe deskundigen van diverse instanties benaderd ($n = 141$). Onder de externe deskundigen vallen onder andere medewerkers van expertisecentra (o.a. NIZW, KIND, stichting Pedagogiekontwikkeling voor het Jonge Kind, Keurmerkinstituut HKZ), adviseurs en consultants op het terrein van de kinderopvang (zoals STRIP en Vyvoj), belangenverenigingen (MO-groep en Branchevereniging Ondernemers in de Kinderopvang) en de drie grote landelijke kinderopvangkoepels (SKON, Catalpa en Humanitas). Daarnaast zijn deskundigen benaderd op gemeentelijk (GGD-inspecteurs, Vereniging van Nederlandse

Gemeenten), provinciaal (Provinciale Ondersteuningsorganisaties) en landelijk niveau (Interprovinciaal Platform Kinderopvang).

In totaal zijn in de periode januari-februari 2004 1198 enquêtes per post verstuurd naar 151 leidinggevenden, 453 leidsters, 453 ouders en 141 deskundigen. Via e-mail zijn naar schatting 47 leidinggevenden en 500 ouders geattendeerd op de internetversie van de enquête.

2.2 Constructie van de enquête

Het uitgangspunt voor de voorgelegde kwaliteitsaspecten uit de enquête is, zoals gezegd, het NCKO-kwaliteitsmodel. De vragen uit de enquête hebben betrekking op de hierin onderscheiden structurele kenmerken en proceskenmerken, die ook als duidelijk onderscheiden onderdelen in de enquête zijn opgenomen. Per kenmerk is omwille van de duidelijkheid telkens een korte beschrijving gegeven. Vanuit een exploratief oogmerk is tevens een beperkt aantal overige onderwerpen toegevoegd aan de lijst van aan het kwaliteitsmodel ontleende kwaliteitskenmerken. Deze kenmerken zijn geselecteerd uit vak- en veldpublicaties en uit binnen- en buitenlandse kwaliteitsinstrumenten op het terrein van de kinderopvang (zie De Kruif & Tavecchio, 2004, voor een overzicht).

In de enquête is de respondenten gevraagd een oordeel te geven over in totaal 37 kwaliteitsaspecten: 15 proceskenmerken en 22 structurele kenmerken (zie Tabel 2.1 en 2.3 voor een volledig overzicht). Dit oordeel was tweeledig. Allereerst kon men per aspect aangeven of dit ‘helemaal niet belangrijk (kan zeker gemist worden)’ in een kwaliteitsinstrument voor de kinderopvang, ‘niet zo belangrijk (kan eventueel gemist worden)’, ‘belangrijk (moet toegevoegd worden)’ of ‘heel erg belangrijk (mag zeker niet ontbreken)’. Dit is respectievelijk gescoord als 1, 2, 3 en 4. Hier ging het derhalve telkens om een weging van een kenmerk. Daarnaast werd gevraagd de vijf belangrijkste aspecten te selecteren voor structurele kwaliteit en voor proceskwaliteit (bijvoorbeeld: ‘Kunt u hieronder aangeven welke 5 van de hierboven genoemde structurele kenmerken u het meest belangrijk vindt?’). Hier was dus sprake van een prioritering. De prioritering, die directer inzicht geeft in de accenten die gelegd worden, is toegevoegd om extra reliëf te geven aan de resultaten. Daarnaast kan een vergelijking duidelijk maken in hoeverre de uitkomsten (deels) convergeren bij beide methodes.

2.3 Procedure

De enquête is zowel in een schriftelijke als elektronische versie aan de verschillende doelgroepen aangeboden. De schriftelijke versie is verstuurd per post, inclusief antwoordenvolp. Dezelfde enquête kon worden ingevuld via het internet. Pedagogen die zijn aangesloten bij de maillijst van het Pedagogenplatform en de ouders uit het bestand van BOinK werden via een e-mail met een introductietekst en een link naar een speciale website geattendeerd op de enquête.

2.4 Analyses

De wegingsgegevens uit de enquête zijn geanalyseerd met rechte tellingen voor elke respondentgroep. Voor de prioriteringsgegevens is een top 5 samengesteld. Deze top 5 is gemiddeld over de individuele top 5-gegevens en geeft zo weer welke vijf aspecten, gemiddeld genomen, het vaakst zijn geprioriteerd.

3. RESULTATEN

3.1 Respons

De enquête is ingevuld geretourneerd door 587 respondenten. Doordat 24 respondenten (4%) niet hebben aangegeven tot welke respondentgroep zij behoorden, is de effectieve steekproef echter 563 respondenten (de responspercentages zijn – uitgaand van de maximale steekproefgrootte en derhalve conservatief geschat – voor ouders, leidinggevendenden, leidsters en deskundigen respectievelijk 39, 32, 25 en 62 procent). De respons was relatief laag voor de groep leidinggevendenden en is moeilijk te beoordelen voor de leidster- en oudergroep. Dit laatste is mede het gevolg van de getrapte steekproefprocedure, waarbij meerdere enquêtes voor de leidster- en ouderpopulatie als het ware ‘van bovenaf’ via de leiding de doelgroep moesten bereiken. In geval van een leidinggevende die het bijgevoegde pakket met drie enquêtes voor ouders en leidsters namelijk niet heeft doorgegeven, betekende dit namelijk dat drie leidsters en drie ouders uitvielen. De respons was hoger voor de groep van deskundigen, die direct konden worden benaderd. De non-respons varieert dus per groep en is voor een aantal groepen vrij laag. Van de respondenten heeft 70% de schriftelijke versie ingevuld en 30% de internetversie; de respons voor beide enquête-varianten is daarmee vergelijkbaar.

3.2 Uitkomsten van de enquête

De resultaten staan hieronder samengevat voor proceskwaliteit en structurele kwaliteit. Bij de weging van elk kenmerk op een schaal van 1 = *helemaal niet belangrijk* (kan zeker gemist worden) (score 1), 2 = *niet zo belangrijk* (kan eventueel gemist worden), 3 = *belangrijk* (moet toegevoegd worden) of 4 = *heel erg belangrijk* (mag zeker niet ontbreken) is in de Tabellen 2.1 en 2.3 een onderscheid gemaakt naar gemiddelde scores boven en onder de 3, oftewel naar belangrijk gevonden kenmerken en minder belangrijke kenmerken. Daarnaast is een top 5 van meest genoemde kenmerken samengesteld op basis van de prioritering.

Proceskwaliteit van de kinderopvang

Een opvallende uitkomst is dat vrijwel alle proceskenmerken van pedagogische kwaliteit uit de enquête belangrijk worden gevonden door alle onderscheiden doelgroepen. Binnen de belangrijk gevonden kenmerken wordt met name het belang onderstreept van het welbevinden van het kind, de sensitieve responsiviteit van de leidster, het praten en uitleggen, het tonen van respect voor de autonomie van het kind en de begeleiding van interacties tussen kinderen. Ook positieve interacties tussen de kinderen worden unaniem belangrijk gevonden voor de kinderopvang. Niet aan alles wordt belang gehecht. De uitzondering betreft de stimulering van voorschoolse vaardigheden, die duidelijk lager wordt gewaardeerd, opnieuw door alle groepen (zie Tabel 2.1). Het beeld van overeenstemming tussen leidinggevenden, leidsters, ouders en deskundigen wordt ook zichtbaar bij de prioritering. Een ‘gemiddelde top 5’ voor de vier groepen (zie Tabel 2.2) laat opnieuw duidelijke overeenkomsten zien (tussen haakjes is het percentage respondenten vermeld dat dit aspect heeft geselecteerd). Welbevinden en sensitieve responsiviteit staan onbetwist op de eerste en tweede plaats en worden op enige afstand gevolgd door de overige kenmerken. De resultaten van weging en prioritering uit respectievelijk Tabel 2.1 en 2.2 convergeren. De hoger gewaardeerde kenmerken bij de weging zijn ook de kenmerken die relatief vaak worden geselecteerd bij de prioritering.

Tabel 2.1

Belang proceskenmerken van pedagogische kwaliteit (in alfabetische volgorde) per doelgroep

	Leiding (n = 78)		Leidsters (n = 146)		Ouders (n = 240)		Deskundigen (n = 88)	
	M	SD	M	SD	M	SD	M	SD
Begeleiding interacties tussen kinderen	3.72	0.45	3.70	0.49	3.68	0.50	3.69	0.51
Betrokkenheid	3.18	0.66	3.25	0.56	3.03	0.65	3.23	0.65
Bevorderen zelfstandigheid	3.25	0.52	3.54	0.51	3.22	0.57	3.23	0.54
Individuele verzorging	3.36	0.48	3.61	0.49	3.41	0.63	3.17	0.52
Positieve interacties tussen kinderen	3.64	0.54	3.56	0.57	3.59	0.53	3.56	0.56
Praten en uitleggen	3.75	0.44	3.68	0.48	3.65	0.49	3.60	0.56
Respect voor autonomie	3.66	0.48	3.61	0.49	3.51	0.52	3.66	0.52
Sensitieve responsiviteit	3.83	0.38	3.72	0.45	3.63	0.51	3.74	0.49
Stimulering creativiteit	3.29	0.56	3.38	0.59	3.19	0.60	3.16	0.54
Stimulering motorische ontwikkeling	3.26	0.44	3.52	0.54	3.30	0.62	3.41	0.62
Stimulering spel	3.27	0.57	3.30	0.60	3.18	0.59	3.32	0.67
Stimulering taalontwikkeling	3.32	0.50	3.53	0.53	3.32	0.58	3.41	0.56
Structureren en grenzen stellen	3.24	0.51	3.49	0.55	3.51	0.55	3.43	0.56
Welbevinden	3.92	0.27	3.87	0.34	3.86	0.35	3.86	0.38
Stimulering voorschoolse vaardigheden	2.52	0.60	2.93	0.66	2.84	0.74	2.72	0.70

Structurele kwaliteit

De grote meerderheid van de voorgelegde structurele kenmerken - 18 van de 22 - wordt belangrijk gevonden bij het meten van de pedagogische kwaliteit van de kinderopvang door alle groepen. Ook hier is er sprake van duidelijke overeenstemming tussen alle groepen. Het belang van de pedagogische visie, het contact met de ouders, de inrichting van de ruimte en de kwaliteit van het speelmateriaal worden benadrukt (scores hoger dan 3.5 voor alle groepen), maar diverse andere kenmerken binnen het domein worden eveneens van belang gevonden bij het beoordelen van de pedagogische kwaliteit. Aan vier kenmerken wordt minder belang gehecht en opnieuw – net als bij de proceskenmerken - door alle groepen: culturele en religieuze diversiteit van het

kinderdagverblijf, de inbedding van het centrum in de buurt, flexibiliteit van de opvang en het snoep- en verjaardagsbeleid. In Tabel 2.3 staan de resultaten samengevat.

Tabel 2.2

Top 5 meest belangrijke proceskenmerken (% respondenten dat dit kenmerk selecteerde)

	Leidinggevend (n = 76)	Leidsters (n = 144)	Ouders (n = 240)	Deskundigen (n = 88)
1	Welbevinden Sensitieve responsiviteit (beide 75%)	Welbevinden (75%)	Welbevinden (75%)	Sensitieve responsiviteit (76%)
2	Positieve interacties tussen kinderen (44%)	Sensitieve responsiviteit (59%)	Sensitieve responsiviteit (57%)	Welbevinden (74%)
3	Praten en uitleggen (41%)	Begeleiding interacties tussen kinderen (37%)	Praten en uitleggen (51%)	Respect voor autonomie (46%)
4	Begeleiden van interacties (40%)	Respect voor autonomie (35%)	Structureren en grenzen stellen (49%)	Positieve interacties tussen kinderen (36%)
5	Respect voor autonomie (39%)	Praten en uitleggen (33%)	Respect voor autonomie Begeleiding interacties tussen kinderen (beide 36%)	Praten en uitleggen (33%)

Toelichting: In sommige gevallen delen twee kenmerken een plaats in de top 5.

De vijf meest genoemde structurele kwaliteitsaspecten staan voor elke groep vermeld in Tabel 2.4. Twee kenmerken zijn universeel, namelijk het contact met ouders en de pedagogische visie van het kinderdagverblijf. Leidsterstabiliteit wordt door drie groepen genoemd in de top 5 van belangrijkste structurele kenmerken. De kenmerken waaraan de leidsters prioriteit toekennen, zijn ook de kenmerken die de ouders noemen. Leidinggevend en de deskundigen leggen vanuit hun perspectief iets andere accenten. De leidinggevend benadrukken het belang van teamoverleg en de opleiding en training van leidsters. De deskundigen wijzen iets vaker dan de andere groepen op het belang van pedagogische ondersteuning van de leidsters. De resultaten uit de wegings- en prioriteringsmethode convergeren grotendeels. Structurele kenmerken met de hoogste wegingscores, dicht bij de maximale score van 4, keren terug in de top 5.

Tabel 2.3

Belang structurele kenmerken van pedagogische kwaliteit (in alfabetische volgorde) per groep

	Leiding (n = 77)		Leidsters (n = 145)		Ouders (n = 240)		Deskundigen (n = 87)	
	M	SD	M	SD	M	SD	M	SD
Aandacht voor probleemkinderen	3.18	0.64	3.34	0.55	3.14	0.61	3.28	0.64
Contact met ouders	3.68	0.45	3.77	0.42	3.57	0.50	3.77	0.42
Groepsstabiliteit	3.09	0.59	3.50	0.57	3.29	0.64	3.15	0.62
Groepsgrootte	3.42	0.61	3.80	0.40	3.56	0.57	3.47	0.64
Hygiënebeleid	3.28	0.60	3.69	0.51	3.44	0.62	3.20	0.66
Inrichting binnenruimte	3.63	0.49	3.71	0.47	3.50	0.53	3.56	0.52
Inrichting buitenruimte	3.62	0.49	3.55	0.53	3.42	0.56	3.55	0.55
Kwaliteit speelmateriaal	3.56	0.50	3.77	0.42	3.60	0.52	3.56	0.56
Leidster-kind ratio	3.43	0.59	3.78	0.42	3.60	0.52	3.59	0.60
Leidsterstabiliteit	3.48	0.57	3.74	0.44	3.71	0.49	3.49	0.57
Opleiding/training leidsters	3.52	0.53	3.65	0.52	3.52	0.55	3.44	0.65
Pedagogisch beleid	3.70	0.49	3.71	0.46	3.43	0.58	3.64	0.51
Pedagogische visie	3.84	0.37	3.67	0.47	3.55	0.52	3.68	0.54
Pedagogische ondersteuning leidster	3.20	0.66	3.03	0.64	3.03	0.65	3.38	0.70
Programmastructuur	3.23	0.58	3.46	0.55	3.44	0.55	3.27	0.66
Teamoverleg	3.65	0.55	3.61	0.53	3.29	0.56	3.48	0.59
Veiligheid (fysiek)	3.52	0.62	3.84	0.39	3.74	0.53	3.32	0.67
Voorzieningen leidsters	3.44	0.62	3.64	0.49	3.24	0.54	3.18	0.71
Culturele en religieuze diversiteit	2.67	0.61	2.72	0.68	2.39	0.67	2.86	0.68
Flexibiliteit opvang	2.81	0.72	2.99	0.62	2.94	0.80	2.56	0.70
Inbedding KDV in de buurt	2.76	0.65	2.71	0.71	2.46	0.74	2.81	0.71
Snoep- en verjaardags- beleid	2.61	0.67	2.67	0.64	2.50	0.67	2.43	0.77

Tabel 2.4

Top 5 meest belangrijke structurele kenmerken (% respondenten dat dit kenmerk selecteerde)

	Leidinggeevenden (n = 77)	Leidsters (n = 147)	Ouders (n = 243)	Deskundigen (n = 88)
1	Pedagogische visie (70%)	Leidsterstabiliteit (61%)	Leidsterstabiliteit (62%)	Contact met ouders (53%)
2	Contact met ouders (63%)	Contact met ouders (49%)	Contact met ouders (53%)	Pedagogische visie (52%)
3	Teamoverleg (43%)	Fysieke veiligheid (43%)	Pedagogische visie (49%)	Pedagogische ondersteuning leidsters (44%)
4	Pedagogisch beleid (41%)	Pedagogische visie (39%)	Fysieke veiligheid (44%)	Leidster-kind ratio (40%)
5	Leidsterstabiliteit Opleiding/training leidsters (beide 34%)	Leidster-kind ratio (38%)	Leidster-kind ratio (32%)	Pedagogisch beleid (39%)

Toelichting: In sommige gevallen delen twee kenmerken een plaats in de top 5.

De enquête-resultaten laten overigens ook zien dat het belang van structurele kwaliteit en proceskwaliteit min of meer gelijk gewogen worden door de respondenten. De over alle proceskenmerken gemiddelde score is voor alle groepen weliswaar iets hoger dan voor structurele kenmerken ($t = 7.26, p < .001$), maar het absolute verschil is telkens klein (respectievelijk .08, .04, .10 en .10 voor leidinggeevenden, leidsters, ouders en externe deskundigen) en de scores liggen bovendien op de gehanteerde vierpuntsschaal telkens boven de '3', dat wil zeggen boven de score voor 'belangrijk / moet toegevoegd worden'.

4 CONCLUSIE

De resultaten van dit enquête-onderzoek, met veelal gelijkgestemde oordelen van leidinggeevenden, leidsters, ouders en externe deskundigen, laten zien dat er duidelijk consensus bestaat met betrekking tot het brede begrip 'pedagogische kwaliteit' in de Nederlandse kinderopvang. Het enige verschil dat naar voren kwam bij de wegingsmethode is dat soms één van de groepen – met name de leidsters – een nog groter belang toekende aan sommige kenmerken waarvan het belang door alle groepen werd onderkend. De oordelen liepen pas uiteen wanneer men gedwongen werd prioriteiten aan te geven, maar ook dan waren de verschillen allesbehalve fors; bovendien gaat het hier om verschillende prioriteringen van

kwaliteitskenmerken, die, zoals gezegd, alle unaniem belangrijk werden gevonden. Ouders, leidsters, leidinggevend en deskundigen blijken dus dezelfde zaken belangrijk te vinden als het gaat om de pedagogische kwaliteit van de kinderopvang. Men kijkt als het ware ‘door dezelfde bril’. Bij de dubbelfocus-benadering blijkt men zowel voor ‘dichtbij’ – dat wil zeggen met de blik op het kind en de leidster, oftewel de proceskwaliteit – als voor ‘veraf’ – dat wil zeggen de randvoorwaarden hiervoor, oftewel de structurele kwaliteit – op hetzelfde te letten. De unieke accenten die men soms legt, passen bij de eigen rol in de kinderopvang. De conclusie kan derhalve worden getrokken dat er een groot aantal betrokkenen is, verspreid over de verschillende groepen die kunnen worden onderscheiden, die gezamenlijk een brede opvatting huldigen over wat ‘pedagogische kwaliteit’ is (zie ook Cryer, Tietze & Wessels, 2002; Kärrby & Giota, 1995; Nelson & Garduque, 1991; Sharp & Davis, 1997; Tietze & Viernickel, 2002; zie Feagans & Manlove, 1994 voor een ander resultaat).

Het resultaat van de enquête vormt tevens een ondersteuning van het NCKO-kwaliteitsmodel (Riksen-Walraven, 2004) op het vlak van zowel de proceskwaliteit als de structurele kwaliteit. Voor de proceskwaliteit geldt dat het belang van het welbevinden en de betrokkenheid van kinderen op de crèche worden erkend, evenals alle onderscheiden leidstervaardigheden. De stimulering op diverse ontwikkelingsterreinen wordt eveneens in brede kring gedragen. Ook alle structurele kenmerken uit het model vinden weerklank bij leidinggevend, leidsters, ouders en externe deskundigen. Aan de resultaten van deze enquête kan ten slotte een argument worden ontleend ook aandacht te schenken aan probleemkinderen en het hygiënebeleid, waarvan het belang ook werd onderstreept door alle partijen.

De verschillende betrokken groepen bij de Nederlandse kinderopvang zijn het eens over welke proceskenmerken en structurele kenmerken belangrijk zijn en derhalve een plaats verdienen in een instrument dat deze kwaliteit in kaart moet brengen. De uitkomsten van de enquête betekenen dat een breed instrument moet worden ontwikkeld voor het meten van de kwaliteit van de Nederlandse kinderopvang met expliciete aandacht voor de diverse kwaliteitsaspecten waarvan het belang is onderstreept door de verschillende betrokkenen. De Kruif en Tavecchio (2004) maken duidelijk dat er momenteel geen kwaliteitsinstrument voorhanden is dat dit kwaliteitsspectrum geheel bestrijkt, al zijn er interessante aanzetten hiertoe en is er enige ervaring opgedaan in Nederland met instrumenten voor een aantal aspecten van pedagogische kwaliteit (Fukkink & Van IJzendoorn, 2004a, 2004b).

III. ONTWIKKELING EN VALIDERING VAN DE NCKO-LEIDSTERSCHALEN EN KINDSCHALEN

1. INLEIDING

1.1. Achtergrond

In Hoofdstuk I van dit rapport werd aangegeven dat een op het NCKO-kwaliteitsmodel gebaseerd meetinstrument voor de pedagogische kwaliteit van de kinderopvang zich zowel zou moeten richten op de proceskwaliteit als op structurele kwaliteitskenmerken. Bij meting van de proceskwaliteit zou vooral aandacht besteed moeten worden aan zes interactievaardigheden van leidsters (sensitieve responsiviteit, respect voor de autonomie van de kinderen, structureren en grenzen stellen, praten en uitleggen, ontwikkelingsstimulering en het begeleiden van interacties tussen kinderen) en aan twee kenmerken van de kinderen (welbevinden en betrokkenheid). Als belangrijke structurele kwaliteitskenmerken werden in het NCKO-model tien kenmerken genoemd, waaronder pedagogisch beleid, pedagogische ondersteuning, leidster-kind ratio, groepsgrootte, en leidster- en groepsstabiliteit. Uit de in hoofdstuk II beschreven enquête onder leidinggevendenden, leidsters, ouders, en deskundigen bleek vervolgens dat de genoemde structurele en proceskenmerken ook binnen deze groepen van betrokkenen gezien worden als het meest belangrijk bij het meten van pedagogische kwaliteit en daarom opgenomen dienen te worden in het te ontwikkelen instrument voor het vaststellen van de pedagogische kwaliteit van opvang in kinderverblijven voor 0- tot 4-jarigen.

Voor het meten van de structurele kwaliteitskenmerken behoefden geen nieuwe instrumenten ontwikkeld te worden, omdat betrouwbare en valide meting hiervan relatief eenvoudig is en er voldoende geschikte instrumenten voorhanden zijn. Voor het vaststellen van de leidstervaardigheden en de kindkenmerken die gekozen zijn als indicatoren voor de proceskwaliteit daarentegen bleken geen instrumenten beschikbaar te zijn die zonder meer konden worden toegepast (zie De Kruif & Tavecchio, 2004). Voor het meten van de leidstervaardigheden en van het welbevinden van de kinderen hebben we eerst bestaande instrumenten aangepast en deze nieuwe instrumenten zijn vervolgens in de in dit Hoofdstuk beschreven valideringsstudie op hun betrouwbaarheid en validiteit getoetst. Voor het meten van de betrokkenheid van de kinderen was wel een geschikte schaal beschikbaar, maar ook deze schaal is in de valideringsstudie opgenomen omdat er onvoldoende bekend was over de validiteit en de betrouwbaarheid van de schaal bij toepassing in Nederlandse kinderdagverblijven. In de volgende paragrafen van deze inleiding wordt achtereenvolgens een korte schets gegeven van de

ontwikkeling van de NCKO-leidsterschalen (1.2) en de kindschalen (1.3) en worden de doelstellingen van de valideringsstudie beschreven (1.4).

1.2 De NCKO-leidsterschalen

Voor het vaststellen van de vaardigheden van leidsters in de interactie met kinderen hebben we gekozen voor het werken met *beoordelingsschalen* (Engels: “rating scales”) waarmee op basis van een bepaalde observatie-episode één score (tussen 1 en 7) wordt toegekend die aangeeft in welke mate de leidster gedurende die gehele episode een bepaalde vaardigheid heeft laten zien. Vergeleken met andere gangbare observatiemethoden, zoals het tellen of beoordelen van afzonderlijke gedragingen (bv. hoe vaak gaat de leidster op de juiste manier op een signaal van een kind in, of hoe vaak legt zij iets uit) is deze methode weinig arbeidsintensief, terwijl uit onderzoek toch gebleken is dat zulke beoordelingen heel betrouwbaar kunnen zijn en vaak meer valide zijn dan scores die gebaseerd zijn op meer gedetailleerde gedragsobservaties (zie bv. De Wolff & Van IJzendoorn, 1997).

In eerdere studies naar de leidster-kind interactie in kinderdagverblijven zijn beoordelingsschalen gebruikt die in principe wel geschikt leken om de leidstervaardigheden te beoordelen, zoals de ORCE-schalen die in de Amerikaanse NICHD-studie en in Nederlands onderzoek gebruikt zijn (zie De Kruif & Tavecchio, 2004) en de schalen van Erickson, Sroufe en Egeland (1985) die ook in Nederlandse kinderdagverblijven zijn toegepast (zie Gevers Deynoot-Schaub & Riksen-Walraven, 2006). Een nadeel van deze beoordelingsschalen is echter dat zij ontworpen zijn voor beoordeling van het gedrag van een leidster ten opzichte van één individueel kind, en niet voor haar gedrag ten opzichte van de hele groep kinderen waarmee zij in de betreffende periode te maken heeft gehad. Er zijn wel enkele schalen die speciaal ontworpen zijn om het gedrag van een leidster ten opzichte van meerdere kinderen te beoordelen. Eén daarvan is de veel gebruikte Caregiver Interaction Scale (CIS, Arnett, 1989; zie De Kruif & Tavecchio 2004), met behulp waarvan het gedrag van een leidster op basis van een observatieperiode van enkele uren beoordeeld wordt op 26 verschillende items, zodat een beeld verkregen wordt van het gedrag van de leidster ten opzichte van een groep kinderen. Het nadeel van de CIS is echter dat de schaal niet ontworpen is om de zes interactievaardigheden van leidsters te meten die door ons belangrijk worden geacht voor het welzijn en de ontwikkeling van kinderen.

Een andere set beoordelingsschalen voor het gedrag van leidsters ten opzichte van een *groep* kinderen zijn de schalen die door De Schipper en Riksen-Walraven ontworpen zijn voor een onderzoek naar het effect van de leidster-kind ratio op de kwaliteit van de leidster-kind

interactie (zie De Schipper, Riksen-Walraven, & Geurts, 2006). De schalen zijn gericht op het meten van de leidstervaardigheden die ook in het NCKO-model centraal staan. Deze schalen zijn als voornaamste uitgangspunt gebruikt bij de ontwikkeling van de NCKO-leidsterschalen; daarnaast zijn ook de ORCE en de schalen van Erickson e.a. (1985) gebruikt als inspiratiebron. De zes NCKO-leidsterschalen (alle 7-puntsschalen) meten de volgende vaardigheden:

1. *Sensitieve responsiviteit*

Sensitieve responsiviteit verwijst naar de mate waarin de leidster adequaat ingaat op signalen die aangeven dat het kind zich niet goed voelt of behoefte heeft aan contact of aan een andere vorm van emotionele ondersteuning. Sensitieve responsiviteit vereist dat de leidster oog heeft voor de toestand en de signalen van het kind, dat ze deze goed weet te interpreteren en dat ze er tijdig en adequaat op reageert, zodat het kind zich begrepen, geaccepteerd en veilig voelt.

Een leidster die *hoog* scoort op deze schaal geeft emotionele steun aan alle kinderen die deze steun op een bepaald moment nodig hebben, zowel tijdens stressvolle als niet stressvolle situaties. Zij herkent de signalen van de kinderen en reageert hier over het algemeen tijdig en adequaat op. De leidster toont belangstelling voor de activiteiten van de kinderen en laat zien dat zij zich bewust is van de behoeften, stemmingen en mogelijkheden van de kinderen. Zij vormt een veilige haven voor de kinderen.

Een leidster die *laag* scoort op deze schaal slaagt er niet in emotionele steun aan de kinderen te geven wanneer zij die nodig hebben. Ofwel zij herkent de signalen van de kinderen niet, ofwel zij herkent de signalen, maar reageert niet of heel langzaam of inadequaat. Haar respons op de signalen van de kinderen, als die er al is, past over het algemeen niet bij de situatie, de emoties of het ontwikkelingsniveau van de kinderen. De leidster kan onverschillig of afstandelijk gedrag vertonen.

2. *Respect voor autonomie*

Respect voor de autonomie van een kind verwijst naar de mate waarin een leidster de individualiteit, motieven, en perspectieven van het kind erkent en respecteert.

Een leidster die *hoog* scoort op deze schaal geeft hen zoveel mogelijk de kans om dingen zelf te doen en zelf uit te proberen. Ze stimuleert de kinderen om problemen zelf op te lossen en respecteert ook hun eigen oplossingen en ideeën. Ze commandeert kinderen niet, maar vraagt hen om iets voor haar te doen of om mee te helpen.

Een leidster die *laag* scoort op deze schaal is intrusief, dringt haar eigen ideeën en plannen op en houdt onvoldoende rekening met (of negeert) de intenties of drang naar

zelfstandigheid van de kinderen. Een gebrek aan respect voor de autonomie uit zich in het abrupt ingrijpen in de activiteiten van kinderen, het afpakken van dingen, het zonder aankondiging of toelichting optillen of verplaatsen van kinderen

3. Structureren en grenzen stellen

Structureren en grenzen stellen verwijst naar de vaardigheid van een leidster om aan kinderen duidelijk te maken wat er van hen verwacht wordt en er tevens voor te zorgen dat zij zich daar ook aan houden.

Een leidster die *hoog* scoort op deze schaal maakt de kinderen goed duidelijk wat er van hen verwacht wordt door situaties, activiteiten en taken zodanig te structureren dat zij deze kunnen overzien. Zij zorgt ervoor dat de kinderen zich aan de regels houden door het stellen van voldoende grenzen. Zij doet dat ook tijdig, zodat kinderen niet het gevoel krijgen dat zij voortdurend ‘in overtreding’ zijn; bij het stellen van grenzen is zij bovendien consistent en doortastend. Als kinderen negatief gedrag vertonen, reageert zij adequaat, d.w.z. in overeenstemming met de aard van de overtreding en het ontwikkelingsniveau van de kinderen. Bovendien besteedt ze aandacht aan de consequenties van het negatieve gedrag voor anderen door het geven van uitleg en het aanbieden van alternatieven.

Een leidster die *laag* scoort op deze schaal slaagt er niet in de kinderen duidelijk te maken wat er van hen verwacht wordt. Zij structureert situaties, activiteiten en taken onvoldoende. Het is voor de kinderen niet duidelijk aan welke regels zij zich dienen te houden; het kan zijn dat er teveel of te weinig regels zijn, een inconsistent gebruik van regels, of regels die niet in overeenstemming zijn met het ontwikkelingsniveau van de kinderen. Als kinderen negatief gedrag vertonen, reageert zij inadequaet, d.w.z. dat haar reacties niet in overeenstemming zijn met de aard van de overtreding of het ontwikkelingsniveau van de kinderen. Er kan sprake zijn van zeer negatieve reacties, zoals schreeuwen of fysiek straffen. Ook een leidster die overdreven veel controle op het gedrag van de kinderen probeert uit te oefenen, scoort laag op deze schaal.

4. Praten en uitleggen

Deze vaardigheid betreft de verbale interacties tussen opvoeder en kind, waarbij het niet alleen gaat om de frequentie van die interacties maar ook om de vorm en inhoud ervan.

Een leidster die *hoog* scoort op deze schaal praat vaak met de kinderen, stimuleert de kinderen om hun gedachten en gevoelens te verwoorden en stemt de timing en inhoud van haar verbale interacties af op het begripsniveau van de kinderen en de activiteiten en interesses die de kinderen op dat moment bezig houden.

Een leidster die *laag* scoort op deze schaal praat nauwelijks of niet en reageert nauwelijks of niet op verbale of non-verbale initiatieven van de kinderen. De leidster doet geen moeite om de kinderen te verstaan en stimuleert het taalgebruik van kinderen niet. Ook een leidster die zich voornamelijk negatief uit tegen de kinderen krijgt een lage score op deze schaal.

5. *Ontwikkelingsstimulering*

Ontwikkelingsstimulering heeft vooral betrekking op de stimulering van de persoonlijke competentie van kinderen. Het gaat hier om de “extra” dingen die de leidster doet om motorische, cognitieve en taalontwikkeling en de creativiteit van de kinderen te stimuleren. Zij kan dat bijvoorbeeld doen door de aandacht van kinderen op bepaalde dingen te richten, door nieuwe activiteiten of nieuw spelmateriaal aan te bieden of door op nieuwe mogelijkheden van het spelmateriaal te wijzen. Ze kan taal- of bewegingsspelletjes initiëren of samen met de kinderen muziek maken of naar muziek luisteren.

Een leidster die *hoog* scoort op deze schaal ziet ontwikkelingsstimulering als één van haar doelen in haar interacties met kinderen. Deze leidster grijpt dagelijkse situaties aan om de ontwikkeling te stimuleren, maar biedt ook uitdagende activiteiten en materialen aan die passen bij het ontwikkelingsniveau van de kinderen. Een leidster die goed is in het stimuleren van de ontwikkeling van kinderen, biedt niet alleen veel “extra” stimulering, maar stemt die stimulering ook af op de interesse, het ontwikkelingsniveau en de toestand van de kinderen, waardoor zij overstimulering van de kinderen voorkomt.

Een leidster die *laag* scoort op deze schaal ziet ontwikkelingsstimulering niet als een prioriteit in haar interacties met kinderen. Zij beperkt haar interacties met de kinderen tot verzorgingssituaties en situaties waarin kinderen aangeven hen nodig te hebben, en laat de kinderen vooral vrij spelen. Een leidster die exploratie belemmert, ontmoedigt, beperkt of consequent niet afstemt op de interesse en het ontwikkelingsniveau van de kinderen krijgt ook een lage score op deze schaal.

6. *Begeleiden van interacties*

Begeleiden van interacties verwijst naar de vaardigheid van een leidster om positieve interacties tussen kinderen te begeleiden en bevorderen.

Een leidster die *hoog* scoort op deze schaal gaat op een adequate manier om met positieve interacties van kinderen. Ze merkt spontane positieve interacties op en beoordeelt deze positief. Bovendien probeert ze positieve interacties te stimuleren door veel uitleg te geven, actief

geschikte situaties te creëren om positieve interacties te bevorderen en zelf het goede voorbeeld te geven.

Een leidster die *laag* scoort op deze schaal begeleidt de kinderen niet of nauwelijks in hun interacties met andere kinderen. Ze besteedt nauwelijks of geen aandacht aan positieve interacties en stimuleert de kinderen ook niet in het tonen van positieve interacties. Een leidster die positieve interacties wel opmerkt maar stopt of afremt krijgt ook een lage score op deze schaal.

1.3 De NCKO-kindschalen

Voor het vaststellen van het *welbevinden* en de *betrokkenheid* van de kinderen hebben we, net als bij de leidsterschalen, gekozen voor het werken met beoordelingsschalen waarmee op basis van een bepaalde observatie-episode een score wordt toegekend die de mate van welbevinden of betrokkenheid aangeeft gedurende die gehele episode. Zoals eerder beschreven, heeft deze methode als voordeel dat hij betrouwbaar is en vaak meer valide dan andere observatiemethoden. Daarnaast kozen we voor instrumenten die relatief makkelijk te hanteren zijn (met het oog op de vertaling naar het veld) en zowel op individueel niveau als groepsniveau een betrouwbaar beeld van welbevinden en betrokkenheid kunnen geven.

Terwijl er in Nederland voor het beoordelen van het welbevinden van kinderen in de kinderopvang een aantal schalen in omloop zijn, staat het beoordelen van de betrokkenheid van kinderen in de kinderopvang nog geheel in de kinderschoenen. Met betrekking tot *welbevinden* wordt in de praktijk vooral gebruik gemaakt van de schaal Welbevinden in Situaties (WiS; Balledux, 2001). Voor onderzoeksdoeleinden worden de WelBeVindenLijst (WBVL; Welbevinden Onderzoeksgroep, 1999 in Linting, Groenen, & Tavecchio, 2002), de Welbevinden Observatielijst Leiden (WOL; De Schipper, Van IJzendoorn, & Tavecchio, 2002) en de Welbevinden GroepsSchaal (WGS) van De Schipper & Riksen-Walraven (zie De Schipper e.a., 2006) gebruikt. De WiS, de WBVL en de WGS kunnen op verschillende momenten van de dag (tijdens verschillende situaties) ingevuld worden door een onafhankelijke observator. De algemenere stellingen op de WOL zijn echter bedoeld voor observatoren die het betreffende kind goed kennen en kan daarom alleen door leidsters ingevuld worden. Terwijl de WiS, de WBVL en de WOL gericht zijn op het welbevinden van individuele kinderen, richt de WGS zich op het welbevinden van de gehele groep.

Voor het beoordelen van *betrokkenheid* in de kinderopvang zijn momenteel geen Nederlandse instrumenten beschikbaar. In België wordt de Leuvense Betrokkenheidschaal (Laevers, Aerden, De Bruyckere, Moons, & Silkens, 2003) gebruikt. Deze beoordelingsschaal

wordt al jaren in de praktijk met kleuters toegepast. Sinds kort is er ook een aangepaste versie van deze schaal beschikbaar voor gebruik bij baby's en peuters. Een onafhankelijke observator beoordeelt de betrokkenheid van de individuele kinderen in de groep in meerdere korte episodes over de dag heen en berekent vervolgens een groeps gemiddelde. De beoordelingsschaal kent vijf duidelijk omschreven niveaus voor betrokkenheid. Naast deze betrokkenheidschaal ontwikkelde Laevers en zijn collega's een zusterschaal, de Leuvense Welbevindenschaal (Laevers e.a., 2003), waarbij hetzelfde beoordelingsprincipe wordt gehanteerd als bij de betrokkenheidschaal. Net als de betrokkenheidschaal kent de welbevindenschaal ook vijf duidelijk omschreven niveaus. Resultaten van een betrouwbaarheidsonderzoek gaven aan dat beide schalen op betrouwbare manier het welbevinden en de betrokkenheid van kinderen in de kinderopvang in beeld kunnen brengen (Laevers e.a., 2003). Er zijn echter nog geen gegevens bekend met betrekking tot de validiteit van deze schalen.

Gegeven het gebrek aan Nederlandse betrokkenheidschalen, het veelbelovende rapport met betrekking tot deze Leuvense schalen (Laevers e.a., 2003) en de praktische overweging om schalen te gebruiken die in samenhang met elkaar ontwikkeld zijn, werden zowel de welbevindenschaal als de betrokkenheidschaal, voorafgaande aan het valideringsonderzoek, door het NCKO in een voorstudie gebruikt. De ervaringen uit de voorstudie leidden tot het besluit om de Leuvense Betrokkenheidschaal onveranderd te gebruiken in het NCKO-onderzoek. De 5-puntschaal van de Leuvense Welbevindenschaal werd echter inhoudelijk enigszins aangepast en uitgebreid naar een 7-puntschaal, om nauwkeuriger onderscheid te kunnen maken tussen het welbevinden van kinderen in de vrij brede middencategorie (neutraal welbevinden). Hieronder worden de Leuvense Betrokkenheidschaal en de NCKO-Welbevindenschaal nader omschreven.

NCKO-Welbevindenschaal

Welbevinden kan omschreven worden als "...een algemene positieve toestand waarin een kind zich bevindt. Alledaagse uitdrukkingen zoals 'lekker in je vel zitten', 'het naar je zin hebben' of 'je prettig en op je gemak voelen' benaderen het begrip eigenlijk nog het beste" (Balledux, 2002, p. 23). Gedragingen en uitingen die duiden op welbevinden laten zien dat "...kinderen zich veilig en ontspannen voelen en genieten van de activiteiten waarmee zij bezig zijn" (Riksen-Walraven, 2004, p. 108).

Bij het beoordelen van welbevinden kan men zich laten leiden door een aantal kenmerken van het kindgedrag. Deze kenmerken hoeven niet noodzakelijk steeds allemaal aanwezig te zijn om van welbevinden te kunnen spreken. Een kind dat zich veilig en op zijn gemak voelt, kan de volgende kenmerken vertonen: openheid, flexibiliteit, zelfvertrouwen, weerbaarheid, vitaliteit,

ontspanning, plezier en/of evenwichtigheid. Een kind dat zich niet op het gemak voelt kan kenmerken van onbehagen vertonen, zoals huilen, schreeuwen, (ongerichte) pogingen om aandacht te trekken, onzekerheid en/of gespannenheid.

De NCKO-Welbevindenschaal kent een 7-puntschaal. Een *hoge* score (score 7) geeft aan dat tekenen van welbevinden overduidelijk aanwezig zijn. Onbehagen is in het geheel niet aanwezig. Een *neutrale* score (score 4) geeft aan dat er niet echt sprake is van onbehagen, maar er zijn ook geen echte uitingen van welbevinden. Noch welbevinden, noch onbehagen overheerst. Een *lage* score (score 1) geeft aan dat er duidelijke signalen van onbehagen zijn. Teken van welbevinden komen hier helemaal niet voor.

Leuvense Betrokkenheidschaal

Betrokkenheid kan omschreven worden als het geconcentreerd en aangehouden bezig zijn. Het kind stelt zich open en is gemotiveerd, omdat de activiteit aansluit bij de exploratiedrang en het ontwikkelingsniveau van het kind. Betrokkenheid is te observeren in vele verschillende situaties en is niet leeftijdsgebonden. Al bij baby's kan men uitspraken doen over de mate waarin ze zich op omgevingsprikkels concentreren en zich openstellen voor de werkelijkheid. Czicszenmihayli (1979) spreekt in dit verband van een bijzondere toestand waarin de betrokken persoon zich bevindt, aangeduid met "the state of flow".

Net als bij welbevinden kan men zich bij het beoordelen van betrokkenheid laten leiden door een aantal kenmerken van het kindgedrag. Deze kenmerken hoeven niet noodzakelijk steeds allemaal aanwezig te zijn om van betrokkenheid te kunnen spreken. Een kind dat betrokken is, kan de volgende kenmerken vertonen: concentratie, energie, complexiteit en creativiteit, mimiek en houding, persistentie, nauwkeurigheid, reactietijd, verwoording en voldoening. Een kind dat in de ruimte staart, er afwezig bij zit, lusteloos in een hoekje wat prutst, doelloos rondrentelt en/of met "niets" bezig is, geeft aan niet betrokken te zijn.

De Leuvense Betrokkenheidschaal (Laevers e.a., 2003) kent een 5-puntschaal. Een *hoge* score (score 5) geeft aan dat er sprake is van de grootst mogelijke betrokkenheid. Kenmerken zoals concentratie, persistentie, energie en complexiteit zijn in grote mate aanwezig. Een *neutrale* score (score 3) geeft aan dat het kind gedurende de observatie-episode nagenoeg aangehouden bezig is met een bepaalde activiteit. Echte betrokkenheidssignalen ontbreken echter. Een *lage* score (score 1) geeft aan dat het kind non-actief of met "niets" bezig is. De activiteit lijkt eerder routinematig en wordt uitgevoerd zonder veel inzet van energie.

1.4 Het doel van de valideringsstudie

De in dit hoofdstuk beschreven valideringsstudie had tot doel het toetsen van de twee belangrijkste psychometrische kwaliteitskenmerken van de NCKO-leidster- en kindschalen, namelijk de betrouwbaarheid en de validiteit.

De *betrouwbaarheid* van een meetinstrument zegt iets over de kwaliteit van de meting, met name de mate waarin de meting consistent en herhaalbaar is. Er zijn verschillende manieren om een indruk te krijgen van de betrouwbaarheid van een instrument; in dit onderzoek hebben we drie veelgebruikte betrouwbaarheidsindicatoren berekend, namelijk:

1. De *inter-observer* betrouwbaarheid van de schalen ofwel de mate waarin verschillende (getrainde) observatoren het eens zijn bij het toekennen van scores op de schalen op basis van dezelfde observatie-episode.
2. De *test-hertest* betrouwbaarheid van de schalen ofwel de mate waarin observatie op verschillende meettijdstippen dezelfde schaalscores oplevert. Hierbij wordt ervan uitgegaan dat het gemeten kenmerk (bv. de sensitieve responsiviteit van de leidster) relatief stabiel is over tijd, zodat inconsistentie van de observatie-uitkomsten over tijd inderdaad duidt op een gebrek aan precisie van het meetinstrument, en niet op een werkelijke verandering in het geobserveerde fenomeen. Daarom moeten de twee tijdstippen waarop gemeten wordt om de test-hertest betrouwbaarheid van een meetinstrument te bepalen niet al te ver uiteen liggen in de tijd. In deze valideringsstudie is gekozen voor twee metingen met een tussentijd van drie maanden.
3. De *interne consistentie* van de schalen, ofwel de mate waarin de schalen onderling samenhangen. Het uitgangspunt is hierbij dat schalen die verschillende aspecten van hetzelfde begrip meten (bv. leidstervaardigheden of kindgedrag) onderling significant zullen samenhangen. Omdat het gaat om verschillende aspecten zal de samenhang echter niet erg sterk zijn. Daarnaast werd ook een significante doch matige samenhang verwacht tussen de leidsterschalen enerzijds en de kindschalen anderzijds, omdat beide groepen schalen indicatoren pretenderen te zijn van eenzelfde begrip, namelijk de pedagogische kwaliteit van de kinderopvang.

De *validiteit* van een meetinstrument geeft aan of het instrument inderdaad datgene lijkt te meten wat het bedoelt te meten. Dat kan onderzocht worden door na te gaan of een maat die met het instrument verkregen is (bv. de sensitieve responsiviteit van een leidster of het welbevinden van

een kind) zich in relatie tot andere maten “gedraagt” op een manier die men theoretisch zou verwachten. In deze valideringsstudie hebben we ons gericht op drie typen validiteit, namelijk:

- (4) De *convergente validiteit* van de schalen, ofwel de mate waarin de scores op de schalen samenhangen (“convergeren”) met scores die verkregen zijn met behulp van andere instrumenten die hetzelfde pretenderen te meten. Om de convergente validiteit van de NCKO-leidsterschalen te bepalen hebben we de scores van de leidsters op deze schalen gerelateerd aan inhoudelijk vergelijkbare scores die op basis van dezelfde observatie-episodes toegekend zijn met behulp van drie andere veelgebruikte instrumenten voor beoordeling van de kwaliteit van leidstergedrag. De convergente validiteit van de kindschalen voor welbevinden en betrokkenheid is bepaald door de scores op deze schalen te relateren aan inhoudelijk vergelijkbare scores die op basis van dezelfde observatie-episodes zijn toegekend met behulp van twee andere meetinstrumenten. Om van convergente validiteit te kunnen spreken, moesten de correlaties tussen de NCKO-schalen en de betreffende valideringsinstrumenten tenminste significant zijn. Erg hoge correlaties werden echter niet verwacht, omdat de NCKO-schalen juist ontworpen zijn om net iets andere leidster- en kindkenmerken te meten dan de bestaande schalen; zeer hoge correlaties zouden erop duiden dat de NCKO-schalen hetzelfde meten als de bestaande instrumenten – en dus overbodig zijn. De meetinstrumenten die gebruikt zijn bij de validering van de NCKO-schalen staan beschreven in de methodensectie van dit hoofdstuk (2.3.2) en een gedetailleerd overzicht van de getoetste hypothesen is te vinden in Tabel 3.1 (voor de leidsterschalen) en Tabel 3.2 (voor de kindschalen).
- (5) De *discriminante validiteit* van de schalen, ofwel de mate waarin zij zich onderscheiden van meetinstrumenten die *iets anders* pretenderen te meten. Om de discriminante validiteit van de NCKO-leidster- en kindschalen te bepalen, hebben we de schaalscores gerelateerd aan bepaalde temperamentkenmerken van de leidsters en kinderen die theoretisch niet gerelateerd zouden moeten zijn aan datgene wat de NCKO-schalen meten. Het ontbreken van een significante samenhang tussen de scores op de NCKO-schalen en de temperamentscores zou daarbij duiden op discriminante validiteit.
- (6) De *predictieve validiteit* van de schalen, ofwel het vermogen van de schalen om iets te voorspellen wat ze theoretisch zouden moeten voorspellen. Eerder onderzoek heeft bijvoorbeeld laten zien dat een lage proceskwaliteit van opvang in een kinderdagverblijf samenhangt met het meer vóórkomen van gedragsproblemen bij kinderen die dat kinderdagverblijf bezoeken, en dat een hogere proceskwaliteit samenhangt met een grotere sociale competentie bij de betreffende kinderen. In de valideringsstudie zijn deze

verbanden onderzocht om de predictieve validiteit van de NCKO-schalen te toetsen. Voor een volledig overzicht van de hypothesen zie Tabel 3.1 (voor de NCKO-leidsterschalen) en Tabel 3.2 (voor de NCKO-kindschalen).

2. METHODE

In de valideringsstudie is een onderzoeksopzet gehanteerd waarbij de kinderdagverblijven tweemaal zijn bezocht (test-hertest design). In de eerste ronde is de kwaliteit van de kinderopvang gemeten met het nieuwe NCKO-instrumentarium (de NCKO-leidster- en kindschalen), aangevuld met diverse bestaande instrumenten om ter validering een vergelijking te kunnen maken tussen bestaande instrumenten en het nieuwe instrument. In de tweede ronde is in dezelfde kinderdagverblijven de kwaliteit opnieuw gemeten, nu alleen met de NCKO-leidster- en kindschalen, om de stabiliteit (c.q. test-hertest betrouwbaarheid) van het nieuwe instrument te kunnen bepalen.

2.1 Steekproef

In het najaar van 2004 werden 171 kinderdagverblijven in en rondom Amsterdam, Leiden en Rotterdam per brief benaderd om mee te werken aan het NCKO-onderzoek. Een week na het versturen van de brief werd de leidinggevende van ieder kinderdagverblijf telefonisch benaderd door een medewerker van het NCKO. Deze lichtte het onderzoek nader toe, beantwoordde eventuele vragen van de leidinggevende en vroeg uiteindelijk of de leidinggevende bereid was om mee te werken aan het onderzoek. Leidinggevendenden van in totaal 47 kinderdagverblijven (27% van de benaderde kinderdagverblijven) stemden in met deelname. Redenen om niet mee te doen waren in de meeste gevallen organisatorisch van aard. In een aantal gevallen klopte het op het Internet beschikbare telefoonnummer niet.

De leidinggevendenden selecteerden vervolgens een babygroep, peutergroep en/of verticale groep voor deelname in het onderzoek. In 23 kinderdagverblijven werd in één groep geobserveerd, in 20 kinderdagverblijven werd in twee verschillende typen groepen geobserveerd en in 4 kinderdagverblijven werd in drie verschillende typen groepen geobserveerd. Van de 75 groepen waren er 30 babygroepen, 29 peutergroepen en 16 verticale groepen.

Voor elk kind in de deelnemende groepen werd een toestemmingsbrief voor de ouders naar het kinderdagverblijf gestuurd. In deze brief werden het doel van het onderzoek en de

algemene procedure voor de dataverzameling beschreven. Daarnaast werd toestemming gevraagd aan de ouders om het kind te filmen voor onderzoeksdoeleinden. In het totaal namen 153 leidsters en 698 kinderen uit 75 groepen in 47 kinderdagverblijven deel aan het onderzoek.

2.2 Procedure

De leidsters en kinderen werden ieder in hun eigen groep bezocht door 2 getrainde observatoren van het NCKO. Het bezoek aan de groep duurde van ongeveer 8:00 uur tot na de lunch (tussen 12:00 en 13:00 uur). Observator 1 nam de ITERS-R (in de babygroepen) of de ECERS-R (in de peuter- en verticale groepen) af (zie paragraaf 2.3.2). Hiertoe maakte zij gedurende de ochtend aantekeningen met betrekking tot de dagelijkse ervaringen van de kinderen in de groep (inclusief inrichting van de groep, beschikbare materialen, programma, routines, etc.). Observator 2 filmde de leidsters en de kinderen tijdens de “normale” alledaagse activiteiten. Beide observatoren stelden zich afzijdig op om het groepsgebeuren zo min mogelijk te beïnvloeden. De leidsters werd gevraagd om het reguliere programma te volgen.

Het filmen vond plaats in filmrondes: eerst alle leidsters (leidster 1, leidster 2, etc.), dan alle kinderen (kind 1, kind 2, etc.). De leidsters en kinderen werden steeds gefilmd in een vaste volgorde en in episodes van 8 tot 10 minuten per leidster (afhankelijk van het aantal leidsters) en 2 minuten per kind. De filmer hield een lijst bij met de filmvolgorde en (voor herkenning) een aantal uiterlijke kenmerken van leidsters en kinderen. Tussen iedere filmronde nam de filmer maximaal 10 minuten pauze. Afhankelijk van de groepsgrootte werden er drie tot vier rondes per ochtend gefilmd. Aan het einde van de ochtend vulde Observator 1 een beoordelingsschaal in met betrekking tot de sensitiviteit van iedere leidster in de groep (de *Caregiver Interaction Scale*). Deze observator sloot vervolgens het bezoek af met een vraaggesprek met één van de leidsters voor nadere informatie die niet via directe observatie te verkrijgen was.

Na ongeveer 3 maanden werden alle groepen telefonisch benaderd voor een tweede afspraak. In 6 van de 75 groepen was een herhaling van het bezoek niet mogelijk. Leidsters noemden voornamelijk organisatorische zaken (bijv. te druk, leidsters naar een andere groep) als reden om af te zien van verdere deelname. Uiteindelijk werden 69 groepen een tweede keer bezocht door één observator per groep. Om ervoor te zorgen dat de groepssamenstelling zoveel mogelijk hetzelfde zou zijn, vond het tweede bezoek plaats op dezelfde dag van de week als het eerste bezoek. De observator filmde wederom de leidsters en kinderen in vaste volgorde tijdens alledaagse activiteiten (bv. fruit eten, vrij spel, leidstergerichte activiteit, lunch). Gemiddeld 2 weken na het tweede bezoek ontving iedere leidster een vragenlijst waarin vragen waren

opgenomen met betrekking tot de leidster zelf (achtergrondinformatie en temperament) en vragen met betrekking tot 2 tot 3 kinderen in de groep (achtergrondinformatie, temperament, gedrag en welbevinden in de groep). Het aantal kinderen over wie de leidster informatie gaf was afhankelijk van de groepsgrootte. Het streven was om van de helft van de kinderen in de steekproef extra informatie te verzamelen via de vragenlijst. De kinderen werden per groep aselekt gekozen uit de in de pretest gefilmde kinderen, waarbij rekening werd gehouden met het geslacht van het kind (het streven was een gelijk aantal jongens en meisjes) en de leeftijd (minimale leeftijd van 12 maanden, omdat sommige vragenlijsten alleen geschikt zijn voor deze leeftijdsgroep; zie 2.3.2).

De individuele opnames van de leidsters en kinderen werden op een later tijdstip door getrainde observatoren beoordeeld. De opnames van de leidsters en de kinderen in dezelfde groep werden door onafhankelijke observatoren beoordeeld (dus de codeur van de leidsterschalen was niet gelijk aan de codeur van de kindschalen). Alle leidsteropnames werden beoordeeld op de NCKO-leidsterschalen. Van ongeveer de helft van de leidsters (de eerste leidster die gefilmd werd) werden alle opnames, met het oog op de bepaling van de convergente validiteit, tevens (door een onafhankelijke observator) beoordeeld op de schaal *opdringerigheid van de leidster* van de ORCE (zie 2.3.2). Alle kindopnames werden beoordeeld op de kindschalen *welbevinden* en *betrokkenheid*. In verband met de bepaling van de convergente validiteit werd van ongeveer de helft van de kinderen alle opnames tevens beoordeeld op een tweede maat voor betrokkenheid (*aandacht voor objecten* van de ORCE, zie 2.3.2).

2.3 Instrumenten

2.3.1 Te valideren instrumenten

NCKO-Leidsterschalen

De gefilmde leidster-episodes werden beoordeeld met behulp van de zes NCKO-Leidsterschalen (NCKO, 2004): *sensitieve responsiviteit, respect voor autonomie, structureren en grenzen stellen, praten en uitleggen, ontwikkelingsstimulering en begeleiden van interacties tussen kinderen*.

Deze leidstervaardigheden worden gemeten op een 7-puntschaal, waarbij 7 = *zeer aanwezig* en 1 = *zeer afwezig* (voor een gedetailleerde beschrijving zie paragraaf 1.2)

Zes observatoren werden getraind om de leidsterschalen betrouwbaar te kunnen gebruiken. Observatoren volgden 6 trainingssessies die elk gemiddeld 4 uur duurden (in totaal dus 24 uur). In deze sessies werd aandacht besteed aan het betrouwbaar scoren van de leidster- en kindschalen en aan de procedures rondom het filmen van de leidsters en kinderen. Een

trainingssessie bestond uit het intensief bestuderen van de schaal die op die dag centraal stond en vervolgens het beoordelen en nabespreken van oefenfragmenten. De training werd afgesloten met een toets. De observatoren beoordeelden daartoe 36 fragmenten van 10 minuten (van 9 verschillende leidsters uit baby-, peuter- en verticale groepen), die werden getoetst aan een tevoren bepaalde consensusscore. Deze opnames waren gemaakt in een pilotstudie voorafgaand aan de validatiestudie. De consensusscore voor ieder fragment was bepaald door 3 observatoren (alle medewerkers van het NCKO). Het bepalen van de mate waarin de observatoren het eens waren met de consensusscore (ofwel de inter-observator betrouwbaarheid) behoorde tot één van de doelen van de valideringsstudie. De resultaten van de toets worden daarom beschreven bij de resultaten van de valideringsstudie (paragraaf 3.1.2.) In totaal waren 9 observatoren (6 nieuwe observatoren + 3 al eerder getrainde observatoren) verantwoordelijk voor het beoordelen van de leidstersessies in de test en hertest.

Alleen leidsters en kinderen met minimaal 2 gefilmde episodes werden in de analyses opgenomen. Dit resulteerde in een definitieve steekproef met 145 leidsters en 623 kinderen uit 72 groepen en 46 kinderdagverblijven. Voor elke leidster met twee of meer gefilmde episodes werd een gemiddelde score per leidsterschaal berekend over de gefilmde episodes. Daarnaast werden de scores voor elke schaal gestandaardiseerd en daarna gemiddeld om tot een totaalscore over alle leidsterschalen heen te komen (*NCKO-Totaalscore*).

NCKO-Welbevindenschaal en Leuvense Betrokkenheidschaal

De gefilmde kind-episodes werden beoordeeld op welbevinden en betrokkenheid met behulp van de NCKO-Welbevindenschaal (NCKO, 2004) en de Leuvense Betrokkenheidschaal (Laevers e.a., 2003). De NCKO-Welbevindenschaal kent een 7-puntschaal waarbij 7 = *uitgesproken hoog* en 1 = *uitgesproken laag*. De Leuvense Betrokkenheidschaal (Laevers e.a., 2003) kent een 5-puntschaal waarbij 5 = *volgehouden intense activiteit* en 1 = *geen activiteit*. Zie paragraaf 1.3 voor een gedetailleerde beschrijving.

Tien observatoren werden getraind op het beoordelen van de afzonderlijke kind-episodes. Een trainingssessie bestond uit het intensief bestuderen van de schaal die op die dag centraal stond en vervolgens het beoordelen en nabespreken van oefenfragmenten met betrekking tot deze schaal. De training werd afgesloten met een toets. Observatoren beoordeelden het welbevinden en de betrokkenheid van 60 kinderen uit 7 groepen (3 babygroepen, 3 peutergroepen en 1 verticale groep). De consensusscores werden bepaald door drie observatoren (3 medewerkers van het NCKO). Het bepalen van de mate waarin de observatoren het eens waren met de consensusscore (ofwel de inter-observator betrouwbaarheid) behoorde tot één van de de doelen van de

valideringsstudie. De resultaten van de toets worden daarom beschreven in paragraaf 3.1.2. In het totaal waren 13 observatoren verantwoordelijk voor het beoordelen van de kindepisodes in de test- en hertest. Een belangrijke voorwaarde was dat de codeur die de leidsters of kinderen had beoordeeld in de pretest de beoordeling voor dezelfde groep niet in de hertest zou uitvoeren. Voor elk kind met twee of meer gefilmde episodes werd een gemiddelde score voor welbevinden en een gemiddelde score voor betrokkenheid berekend over alle gefilmde episodes heen. Daarna werden groepsscores voor beide variabelen berekend door de scores van alle kinderen in één groep te middelen.

2.3.2. Instrumenten voor validering

Instrumenten voor convergente validiteit

ITERS-R/ECERS-R

De Infant Toddler Environment Rating Scale – Revised (ITERS-R: Harms, Cryer & Clifford, 2003) en Early Childhood Environment Rating Scale – Revised (ECERS-R: Harms, Clifford & Cryer, 1998) werden (naast andere instrumenten) gebruikt in de toetsing van de convergente validiteit van de leidsterschalen. Deze instrumenten werden vertaald vanuit het Engels en aangepast aan de Nederlandse situatie en zijn ontwikkeld om de proceskwaliteit, oftewel de dagelijkse ervaringen van de kinderen op het kinderdagverblijf, te beoordelen. De originele schalen worden veelvuldig en wereldwijd gebruikt in onderzoeken naar de proceskwaliteit van de kinderopvang.

De ITERS-R is bedoeld voor groepen met kinderen in de leeftijd tot 2,5 jaar (meestal baby- of dreumesgroepen), terwijl de ECERS-R bedoeld is voor groepen met kinderen van 2,5 tot 5 jaar (meestal peutergroepen). In verticale groepen waar de helft of meer van de kinderen 2,5 jaar of ouder is, wordt de ECERS-R gebruikt, in verticale groepen waar minder dan de helft van de kinderen 2,5 jaar oud is, wordt de ITERS-R gebruikt. Elk van deze meetinstrumenten bestaat uit 7 subschalen, te weten: (a) Ruimte en Meubilering, (b) Individuele Zorg, (c) Taal, (d) Activiteiten, (e) Interacties, (f) Programma en (g) Voorzieningen voor Ouders en Staf. De items van beide schalen worden gepresenteerd op een 7-puntsschaal met beschrijvingen voor 1 (inadequaats), 3 (minimaal), 5 (goed) en 7 (uitstekend). *Inadequaats* beschrijft opvang die niet voldoet aan de basale verzorgingsbehoefte, *minimaal* beschrijft opvang die voldoet aan de basale verzorgings- en ontwikkelingsbehoefte, *goed* beschrijft de opvang die verder gaat en gericht is op de ontwikkeling van de kinderen en *uitstekend* beschrijft opvang van hoge kwaliteit met individuele/persoonlijke zorg. De beoordeling wordt gebaseerd op de situatie zoals die

geobserveerd of gerapporteerd wordt gedurende een bezoek van minimaal 3 uur en wordt afgesloten met een vraaggesprek met één van de leidsters van de groep. Voor elk item wordt een score gegeven van 1 tot 7. De totaalscore is de gemiddelde score van alle items.

Alleen de totaalscore en de subschalen Taal, Interacties en Activiteiten werden gebruikt in de validering. De interne consistentie (ofwel de onderlinge samenhang van individuele items binnen iedere schaal) van de totaalscore en subschalen was voldoende tot hoog. Interne consistentie van de ITERS-R schalen is als volgt: Totaal = .80, Taal = .69, Interacties = .71, Activiteiten = .51. De interne consistentie van de ECERS-R schalen is als volgt: Totaal = .84, Taal = .73, Interacties = .76 en Activiteiten = .51.

Vier observatoren werden getraind op de ITERS-R en 4 observatoren werden getraind op de ECERS-R. De training vond plaats vóór de start van de dataverzameling en werd verzorgd door experts op de ITERS-R en ECERS-R. Na een algemene introductie participeerde iedere observator in ten minste 4 veldobservaties onder supervisie van een expert trainer. Iedere observatie werd gevolgd door een nabespreking met de trainer, waarna de inter-observator betrouwbaarheid werd bepaald. Een observator werd beschouwd als betrouwbaar als tijdens ieder van drie opeenvolgende bezoeken aan kinderdagverblijven 80% van haar scores (binnen 1 schaalpunt) overeen kwamen met de score van de trainer. Observatoren bezochten tijdens de training gemiddeld 5 kinderdagverblijven (range 4 – 7). Het gemiddelde betrouwbaarheidspercentage over de drie opeenvolgende bezoeken (>80%) heen was 87% (range 84% tot 89%).

Caregiver Interaction Scale (CIS)

De CIS (Arnett, 1989) werd gebruikt in de toetsing van de convergente validiteit van de leidsterschalen. De CIS meet een breed scala aan leidstergedragingen en wordt om deze reden veelvuldig gebruikt om een beter beeld te krijgen van de kwaliteit van de leidster-kind interacties dan mogelijk is met de ITERS-R en ECERS-R. De CIS bestaat uit 26 items die worden beoordeeld door een getrainde observator op een 4-puntschaal met de volgende ankers: 1 = *helemaal niet van toepassing*, 2 = *enigszins van toepassing*, 3 = *in behoorlijke mate van toepassing*, 4 = *zeer sterk van toepassing*.

Arnett (1989) onderscheidt vier subschalen (positieve interacties, onplezierige interacties, afstandelijkheid en toegeeflijkheid), die in onderzoek echter nog niet vaak gerepliceerd zijn. In een aantal onderzoeken worden drie subschalen gebruikt (zie Whitebook, Howes & Phillips, 1989) of twee (Van IJendoorn e.a., 1996). Vaker nog wordt alleen maar de totaalscore gebruikt

(Burchinal, 2000; Peisner-Feinberg e.a., 1999). In dit onderzoek werd gekozen voor het gebruik van de totaalscore (gemiddelde van alle itemscores).

CIS-observaties werden uitgevoerd door de ITERS-R en ECERS-R observatoren. Training voor de CIS vond plaats tijdens de ITERS-R en ECERS-R trainingen. De observatoren vulden de CIS in direct na het scoren van de ITERS-R of ECERS-R. Een observator werd beschouwd als betrouwbaar als 80% van de scores (binnen 1 schaalpunt) voor drie opeenvolgend beoordeelde leidsters overeenkwamen met de score van de trainer. Een schaalpunt verschil was echter alleen acceptabel tussen de scores 1 en 2 (beschrijving van negatievere gedragingen) en tussen de score 3 en 4 (beschrijving van positievere gedragingen). Het gemiddelde betrouwbaarheidspercentage over de drie opeenvolgende beoordelingen (>80%) heen was 90% (range 85% tot 98%).

Observational Record of the Caregiving Environment (ORCE)

De ORCE (NICHD Early Child Care Research Network, 1996) werd gebruikt in de toetsing van de convergente (eigenlijk: congruente) validiteit van de leidsterschaal *respect voor autonomie* en de kindschaal *betrokkenheid*. Twee ORCE variabelen werden gemeten: (1) *opdringerigheid van de leidster* (ofwel het gebrek aan respect voor autonomie) en (2) *aandacht van kinderen voor objecten* (ofwel betrokkenheid). De ORCE werd speciaal ontwikkeld voor een grootschalig onderzoek naar ervaringen van kinderen in de Amerikaanse kinderopvang. *Opdringerigheid van de leidster* wordt gemeten op een 4-puntschaal met de volgende ankers: 1 = *helemaal niet kenmerkend voor deze leidster*; 2 = *enigszins kenmerkend*; 3 = *tamelijk kenmerkend* en 4 = *zeer kenmerkend voor deze leidster*. *Aandacht voor objecten* wordt gemeten op een 4-puntschaal met de volgende ankers: 1 = *helemaal niet aanwezig*, 2 = *een beetje*, 3 = *tamelijk aanwezig* en 4 = *sterk aanwezig*.

Eén observator was verantwoordelijk voor het beoordelen van de leidstervariabele *opdringerigheid van de leidster*. Een tweede observator was verantwoordelijk voor het beoordelen van de kindvariabele *aandacht voor objecten*. Een observator werd beschouwd als betrouwbaar als ten minste 80% van de scores (binnen 1 schaalpunt) overeenkwam met de score van de trainer. Een schaalpunt verschil was echter alleen acceptabel tussen de scores 1 en 2 (de ‘negatieve’ kant van de schaal) en tussen de score 3 en 4 (de ‘positieve’ kant van de schaal).

Beide observatoren waren onafhankelijk, d.w.z. dat zij verder bij geen enkele NCKO-training of NCKO-observaties betrokken waren en dus blind waren m.b.t. de al gescoorde data en de betekenis hiervan.

Welbevinden Observatie Lijst (WOL)

De WOL (de Schipper, van IJzendoorn, & Tavecchio, 2002) werd gebruikt in het toetsen van de convergente validiteit van de --*Welbevindenschaal* en het toetsen van de predictieve validiteit van de NCKO-leidsterschalen. De WOL werd ontwikkeld om, met behulp van een beoordelingsschaal, het welbevinden van een kind in de groep te beoordelen. De WOL bestaat uit 28 items die door de leidster beoordeeld worden op een 6-puntschaal, variërend van 1 = *situatie komt nooit voor* tot 6 = *situatie komt altijd voor*. In navolging van De Schipper e.a. (2002) werden alleen de 12 kernitems gebruikt om tot een gemiddelde score voor welbevinden te komen (bv: “Dit kind vindt het leuk om naar het kindercentrum te gaan”). De interne consistentie van deze items was .76 (vergelijkbaar met De Schipper e.a.; Cronbach alpha = .81). De WOL maakte onderdeel uit van de vragenlijst voor de leidsters. Iedere leidster beoordeelde het welbevinden van 2 tot 3 kinderen uit de groep.

Instrumenten voor discriminante validiteit

Emotionality, Activity & Sociability Survey (EAS)

De Nederlandse vertaling van de EAS (Buss & Plomin, 1984) werd gebruikt om de discriminante validiteit van de leidsterschalen te bepalen. De EAS is ontwikkeld om het temperament van de leidsters te meten en bestaat uit 20 items die door de leidster zelf worden beoordeeld op een 5-puntschaal, lopend van 1 = *niet typisch voor mij* tot 5 = *heel typisch voor mij*. Buss en Plomin onderscheiden 5 subschalen met ieder 4 items: *Activity, Fearfulness, Distress, Anger* en *Sociability*. De interne consistentie van deze subschalen was in ons onderzoek respectievelijk .49, .31, .47, -.02 en .66. Alleen de subschaal *Sociability* werd daarom gebruikt in de analyses. Deze schaal beschrijft de voorkeur om met anderen te zijn of werken in plaats van alleen (bv.: “Ik werk liever met andere mensen dan in mijn eentje”). De EAS maakte onderdeel uit van de vragenlijst voor de leidsters.

Infant Characteristics Questionnaire (ICQ)

De ICQ (Bates, Freeland & Lounsbury, 1979) werd gebruikt in de toetsing van de discriminante validiteit van de kindschalen. De ICQ werd ontwikkeld om het temperament van jonge kinderen te beoordelen en bestaat uit 33 items vragen over diverse gedragskenmerken van kinderen en één algemeen item met de vraag naar de mate van moeilijkheid van het temperament van het kind. De items worden beoordeeld op een 7-puntschaal. In navolging van De Schipper e.a. (2002) werden de scores op 17 items gemiddeld om de subschaal *Moeilijk Temperament* te berekenen. Hoge

scores op deze items (bv: “Hoe vaak huilt dit kind in het algemeen of windt zich op?”) duiden op een moeilijker temperament. De interne consistentie van de items was .83 (vergelijkbaar met de Schipper e.a.; Cronbach alpha = .87). De ICQ maakte onderdeel uit van de vragenlijst voor de leidsters. Iedere leidster vulde de ICQ in voor 2 tot 3 kinderen uit de groep.

Instrumenten voor predictieve validiteit

Brief Infant Toddler Social Emotional Assessment (BITSEA)

De BITSEA (Briggs-Gowan, Carter, Irwin, Wachtel, & Cicchetti, 2004) werd gebruikt in de toetsing van de predictieve validiteit van zowel de leidster- als de kindschalen. De BITSEA werd ontwikkeld om een indicatie te krijgen van mogelijke sociaal-emotionele problemen en van de competentie van kinderen en bestaat uit 42 items die worden beoordeeld op een 3 puntschaal (0 = *niet waar/zelden*, 1 = *een beetje waar/soms*, 2 = *waar/vaak*). Er zijn twee subschalen te onderscheiden: *Probleemgedrag* (gemiddelde van 31 items, bv: “Doet zich zo vaak pijn dat je hem/haar voortdurend in de gaten moet houden”) en *Competentie* (gemiddelde van 11 items, bv: “Zoekt u op (of andere leidster) als hij/zij overstuur is”). Op basis van normscores voor leeftijd en sekse kan een klinische indicatie verkregen worden van probleemgedrag of competentie (zie Briggs-Gowan & Carter, 2001). In deze valideringsstudie werden de subschalen als continue variabelen gebruikt. De BITSEA is geschikt voor gebruik bij kinderen vanaf 12 maanden. Alleen de scores van kinderen van 12 maanden en ouder zijn meegenomen in de analyses. De BITSEA maakte onderdeel uit van de vragenlijst voor de leidsters. Iedere leidster beoordeelde het gedrag van 2 of 3 aselect gekozen kinderen uit de groep.

Overige instrumenten

Vragenlijst leidsters

Alle leidsters ontvingen een vragenlijst die uit twee delen bestond. Deel 1 bestond uit (a) vragen over de achtergrond van de leidster (leeftijd, sekse, geboorteland, beroepsgerichte opleiding, hoogst genoten opleiding, aantal uren werkzaam per week, aantal jaren ervaring in de kinderopvang, aantal jaren ervaring binnen de geobserveerde groep) en (b) vragen met betrekking tot het temperament van de leidster (EAS). Deel 2 bestond uit 2 of 3 vragenlijsten die de leidster invulde voor 2 of 3 (door de onderzoekers aselect gekozen) kinderen in de groep. Voor ieder van deze kinderen beantwoordde de leidster vragen met betrekking tot de achtergrond van het kind (geboortedatum, sekse, aantal dagen in de groep, hoe lang al in deze groep, bekendheid leidster

met het kind, bijzonderheden in de laatste maand), het temperament van het kind (ICQ), het gedrag van het kind (BITSEA) en het welbevinden van het kind in de groep (WOL).

2.4 Concrete verwachtingen

Voor het bepalen van de *inter-observer betrouwbaarheid* van de leidster- en kindschalen werd de strengste maat gehanteerd - de Intraclass Correlation Coefficient (ICC; two-way mixed, absolute agreement) - met $\geq .70$ als kritische ondergrens. Met betrekking tot de *test-hertest betrouwbaarheid* (stabiliteit) van de leidsterschalen werd verwacht dat de stabiliteit voor de gestandaardiseerde totaalscore hoger zou zijn dan de stabiliteit voor de afzonderlijke schalen. Voor de kindschalen werd verwacht dat de stabiliteit op groepsniveau hoger zou zijn dan de stabiliteit op individueel kindniveau.

Omdat we ervan uitgaan dat de zes leidsterschalen verschillende aspecten van hetzelfde begrip meten, namelijk “interactievaardigheden”, werd verwacht dat de scores van de leidsters op de verschillende schalen onderling significant zouden samenhangen; omdat het om *verschillende* aspecten van hetzelfde begrip gaat, werd echter geen erg hoge correlatie verwacht. Om dezelfde reden werd een significante doch matige correlatie verwacht tussen de twee kindschalen. Tenslotte werd ook een significante doch matige samenhang verwacht tussen de leidsterschalen enerzijds en de kindschalen anderzijds, omdat beide groepen schalen indicatoren pretenderen te zijn van eenzelfde begrip, namelijk de pedagogische kwaliteit van de kinderopvang.

De tabellen 3.1 en 3.2 geven een overzicht van de verwachtingen met betrekking tot de validiteit voor de leidsterschalen en de kindschalen, inclusief een aanduiding van de schalen (scores) die gebruikt werden in deze valideringsstudie.

Tabel 3.1
Verwachtingen validiteit NCKO-Leidsterschalen

Validatie-instrument	Verwachtingen
<i>Convergente validiteit</i>	
▪ CIS-Totaal score	1. Positieve correlatie tussen de NCKO <i>Totaalscore (leidsters)</i> en alle afzonderlijke NCKO-leidsterschalen enerzijds en de <i>CIS-Totaal score</i> anderzijds. De hoogste correlatie wordt verwacht tussen NCKO Sensitieve Responsiviteit en de CIS-Totaal score.
▪ ORCE-opdringerigheid	2. Significante negatieve correlatie tussen NCKO <i>Respect voor Autonomie</i> en <i>ORCE-opdringerigheid</i>
▪ ITERS-R/ECERS-R	3. Significante positieve correlatie tussen de NCKO <i>Totaalscore (leidsters)</i> en alle afzonderlijke NCKO-leidsterschalen enerzijds en de ITERS-R/ECERS-R <i>Totaalscore</i> anderzijds
	4. Significante positieve correlatie tussen NCKO <i>Praten en uitleggen</i> enerzijds en ITERS-R/ECERS-R subschalen <i>Taal</i> en <i>Interacties</i> anderzijds
	5. Significante positieve correlatie tussen NCKO <i>Ontwikkelingsstimulering</i> en ITERS-R/ECERS-R subschalen <i>Taal</i> en <i>Activiteiten</i>
	6. Significante positieve correlatie tussen NCKO <i>Begeleiden van interacties</i> enerzijds en ITERS-R/ECERS-R subschalen <i>Interacties</i> en <i>Taal</i> anderzijds
<i>Discriminante validiteit</i>	
▪ EAS-Sociability	7. Geen significante correlatie tussen de NCKO <i>Totaalscore (leidsters)</i> en de afzonderlijke NCKO-leidsterschalen enerzijds en de EAS-Sociability score (temperament van de leidster) anderzijds
<i>Predictieve validiteit</i>	
▪ WOL	8. Significante positieve correlatie tussen de NCKO <i>Totaalscore (leidsters)</i> en de afzonderlijke NCKO-leidsterschalen enerzijds en de WOL (welbevinden) anderzijds
▪ BITSEA-probleemgedrag & BITSEA-competentie	9. Significante negatieve correlatie tussen de NCKO <i>Totaalscore (leidsters)</i> en <i>BITSEA-probleemgedrag</i>
	10. Significante positieve correlatie tussen de NCKO <i>Totaalscore (leidsters)</i> en <i>BITSEA-competentie</i>

Tabel 3.2
Verwachtingen validiteit NCKO-kindschalen

Validatie-instrument	Verwachtingen
<i>Convergente validiteit</i>	
▪ WOL	11. Significante positieve correlatie tussen de <i>NCKO-Welbevindenschaal</i> en de <i>WOL</i> (welbevinden)
▪ ORCE-aandacht voor objecten	12. Significante positieve correlatie tussen de <i>Leuvense Betrokkenheidschaal</i> en de <i>ORCE-aandacht voor objecten</i> (betrokkenheid)
<i>Discriminante validiteit</i>	
▪ ICQ	13. Geen significante correlatie tussen de <i>NCKO-Welbevindenschaal</i> en de <i>Leuvense Betrokkenheidschaal</i> enerzijds en de ICQ (temperament van de kinderen) anderzijds
<i>Predictieve validiteit</i>	
▪ BITSEA-probleemgedrag & BITSEA-competentie	14. Significante negatieve correlatie tussen de <i>NCKO-Welbevindenschaal</i> en de <i>Leuvense Betrokkenheidschaal</i> enerzijds en <i>BITSEA-probleemgedrag</i> anderzijds
	15. Significante positieve correlatie tussen de <i>NCKO-Welbevindenschaal</i> en de <i>Leuvense Betrokkenheidschaal</i> enerzijds en <i>BITSEA-competentie</i> anderzijds

3. RESULTATEN

3.1. NCKO-Leidsterschalen

3.1.1. Beschrijvende data

In Tabel 3.3 wordt een overzicht gegeven van de gemiddelde scores voor de leidsterschalen op leidsterniveau (over alle gefilmde episodes heen). De waarden voor skewness en kurtosis geven aan dat de scores voor drie van de schalen niet normaal verdeeld zijn; de scores voor *sensitieve responsiviteit* liggen relatief vaak boven het gemiddelde en de scores voor *ontwikkelingsstimulering* en *begeleiden van interacties* liggen relatief vaak onder het gemiddelde. Omdat sommige analyses met betrekking tot de validiteit van de leidsterschalen op groepsniveau zijn uitgevoerd, geeft Tabel 3.3 ook een overzicht van de gemiddelde scores voor leidstervaardigheden op groepsniveau (over alle leidsters in een groep heen). De verdeling van de scores benadert voor alle leidsterschalen een normale verdeling, behalve voor *begeleiden van*

interacties. Voor deze laatste schaal liggen de scores ook op groepsniveau relatief vaak onder het gemiddelde.

Tabel 3.3
Beschrijvende data NCKO-leidsterschalen (7-puntsschalen)

	<i>M</i>	<i>SD</i>	Range	Skewness*	Kurtosis*
Leidsterniveau (<i>N</i> = 145)					
Sensitieve responsiviteit	4.44	.90	1.67 – 6.50	-2.50	.49
Respect voor autonomie	4.23	.83	2.00 – 6.25	-.81	.91
Praten en uitleggen	3.38	.92	1.25 – 6.00	.55	-.18
Ontwikkelingsstimulering	2.20	.87	1.00 – 4.75	2.93	-.56
Begeleiden van interacties	1.60	.75	1.00 – 4.75	9.01	10.01
Groepsniveau (<i>N</i> = 72)					
Sensitieve responsiviteit	4.41	.74	2.50 – 6.13	-1.81	.23
Respect voor autonomie	4.21	.68	2.00 – 5.67	-1.47	1.80
Praten en uitleggen	3.37	.78	1.50 – 5.25	.07	.04
Ontwikkelingsstimulering	2.19	.66	1.00 – 3.67	1.01	-1.50
Begeleiden van interacties	1.60	.64	1.00 – 3.17	3.56	1.29

*Skewness = skewness/standard error; kurtosis = kurtosis/standard error. Bij een normale verdeling vallen deze waarden tussen -2.00 en 2.00.

3.1.2. Betrouwbaarheid

Inter-observator betrouwbaarheid

De observatoren voldeden aan de gestelde eis ($ICC \geq .70$ op iedere schaal) met betrekking tot de inter-observator betrouwbaarheid voor de leidsterschalen. Dit geeft aan dat de schaalwaarden zó duidelijk geformuleerd zijn, dat getrainde observatoren vrijwel dezelfde beoordeling geven als zij de leidster in actie zien. De gemiddelde ICC-waarden over alle observatoren heen was als volgt: *sensitieve responsiviteit* = .85, *respect voor autonomie* = .81, *praten en uitleggen* = .76, *ontwikkelingsstimulering* = .80, *begeleiden van interacties* = .83. De ICC-waarden voor de schaal *structureren en grenzen stellen* kwamen niet boven de kritische grens. Deze schaal kon niet voldoende variatie tussen leidsters aangeven en wordt momenteel herzien. De schaal is verder niet meegenomen in de gerapporteerde analyses.

Test-hertest betrouwbaarheid

Voor het bepalen van de test-hertest betrouwbaarheid (of stabiliteit) van de leidsterschalen waren de scores beschikbaar van 108 leidsters, die op twee verschillende dagen (met gemiddeld 3 maanden tussen de twee observatiemomenten) in hun eigen groep gefilmd werden. Tabel 3.4 laat zowel de beschrijvende data van de test en de hertest zien als de test-hertest correlatie per leidsterschaal. Er waren geen statistisch significante verschillen tussen de gemiddelde test en hertest scores.

Tabel 3.4

Gemiddelde scores en standaarddeviaties voor de NCKO-leidsterschalen op de twee meetmomenten en test-hertest betrouwbaarheid (N = 108)

	Test	Hertest	Test-hertest correlatie
Leidstervaardigheid	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>R</i>
Sensitieve responsiviteit	4.48 (.90)	4.30 (.99)	.42**
Respect voor autonomie	4.22 (.86)	4.33 (.82)	.48**
Praten en uitleggen	3.40 (.93)	3.27 (.96)	.48**
Ontwikkelingsstimulering	2.18 (.90)	2.06 (.81)	.44**
Begeleiden van interacties	1.63 (.79)	1.58 (.63)	.36**
NCKO-Leidstervaardigheden (Totaal) ^a	0 (.84)	0 (.82)	.60**

^a Gestandaardiseerd gemiddelde van alle leidstervaardigheden

** $p < .01$

De test-hertest betrouwbaarheid is, zoals verwacht, het hoogst voor de gestandaardiseerde totaalscore van de leidstervaardigheden. De betrouwbaarheid is het laagst voor *begeleiden van interacties*.

Interne consistentie.

Tabel 3.5 laat de onderlinge samenhang zien tussen de leidsterschalen. De resultaten bevestigen de eerder gestelde verwachting: de NCKO-leidsterschalen hangen significant en positief, maar

niet te sterk met elkaar samen. Elke leidsterschaal hangt enigszins samen met de andere leidsterschalen (met gedeelde varianties die variëren van 19 tot 56%), maar meet ook unieke variantie. Cronbach's alpha (een andere maat om interne consistentie aan te geven) voor de vijf eerder genoemde leidsterschalen was .88. Deze waarde geeft aan dat de leidsterschalen gezamenlijk een betrouwbare indicatie geven van het begrip "leidstervaardigheden".

Tabel 3.5
Correlaties tussen de NCKO- leidsterschalen (N = 145)

	1	2	3	4
1. Sensitieve responsiviteit	-			
2. Respect voor autonomie	.69**	-		
3. Praten en uitleggen	.75**	.66**	-	
4. Ontwikkelingsstimulering	.57**	.56**	.65**	-
5. Begeleiden van interacties	.47**	.44**	.59**	.53**

** $p < .01$

3.1.3. Validiteit

Zoals aangegeven in Hoofdstuk I, zijn de NCKO-leidsterschalen getoetst op convergente validiteit, discriminante validiteit en predictieve validiteit. De verwachtingen met betrekking tot deze drie valideringsmethoden werden in Tabel 3.1 gepresenteerd. Tabel 3.6 laat de beschrijvende data zien van alle variabelen die (naast de leidsterschalen zelf) gebruikt zijn tijdens de validering. Omdat de ITERS-R en de ECERS-R op groepsniveau worden gemeten en de scores op de WOL en de BITSEA ook op groepsniveau zijn geaggregeerd, zijn de leidsterschalen in de analyses met deze variabelen, eveneens samengevoegd tot een groepsscore (het gemiddelde over alle leidsters in de groep); zie Tabel 3.3 voor beschrijvende data van de leidstervaardigheden op groepsniveau (de waarde bij 'n' heeft hier dus betrekking op het aantal groepen). De verwachtingen met betrekking tot de NCKO-leidsterschalen zijn getoetst met behulp van correlaties.

Convergente validiteit

De convergente validiteit van de NCKO-leidsterschalen werd getoetst aan de hand van correlaties van de NCKO-leidsterschalen met de CIS, een variabele van de ORCE en een aantal subschalen van de ITERS-R/ECERS-R. De gegevens in Tabel 3.7 laten zien dat de verwachting met betrekking tot de CIS is uitgekomen (zie Tabel 3.1, verwachting 1). Er bleek een statistisch significant positief verband te bestaan tussen alle NCKO-leidsterscores enerzijds en de CIS-

Totaalscore anderzijds, waarbij voor wat betreft de afzonderlijke NCKO-leidsterschalen het verband met *sensitieve responsiviteit* en de CIS-Totaalscore het sterkst was.

Tabel 3.6
Beschrijvende data voor de instrumenten gebruikt ter validering van de NCKO-leidsterschalen

Validatie-instrumenten	<i>N</i>	<i>M</i>	<i>SD</i>	Range	Mogelijke Schaalwaarden
<i>Convergente validiteit</i>					
▪ CIS-Totaal	130	3.12	.36	1.88 – 3.73	1 – 4
▪ ORCE-opdringerigheid	65	1.35	.61	1.00 – 3.00	1 – 4
▪ ITERS-R/ECERS-R Totaal*	72	3.21	.56	1.66 – 4.34	1 – 7
▪ ITERS-R/ECERS-R Taal*	72	3.93	1.21	1.00 – 6.25	1 – 7
▪ ITERS-R/ECERS-R Interacties*	72	4.45	1.40	1.00 – 7.00	1 – 7
▪ ITERS-R/ECERS-R Activiteiten*	72	2.46	.53	1.33 – 3.89	1 – 7
<i>Discriminante validiteit</i>					
▪ EAS-Sociability	110	3.57	.61	1.00 – 5.00	1 – 5
<i>Predictieve validiteit</i>					
▪ WOL*	60	4.84	.32	4.08 – 5.86	1 – 6
▪ BITSEA-probleemgedrag*	60	.25	.08	.11 - .49	0 – 2
▪ BITSEA-competentie*	60	1.40	.24	.73 – 1.76	0 – 2

*Scores op groepsniveau. Groepsscores voor leidstervaardigheden gebruikt in analyses met deze variabelen.

De relatie tussen de leidstervaardigheid *respect voor autonomie* en de met de ORCE gemeten *opdringerigheid van de leidster* was, in overeenstemming met de verwachting, negatief (zie Tabel 3.1, verwachting 2). Verwachtingen 3, 4, 5 en 6 beschrijven de relaties tussen NCKO-leidsterschalen en de totaalscore en een aantal subschalen van de ITERS-R/ECERS-R. Verwachting 3 (er bestaat een positieve relatie tussen alle NCKO-leidsterschalen enerzijds en de ITERS-R/ECERS-R totaalscore anderzijds) werd bevestigd.

De leidsterscores voor *praten en uitleggen* en *begeleiden van interacties* hingen, zoals verwacht, positief samen met de ITERS-R/ECERS-R subschalen Taal en Interacties (verwachtingen 4 en 6). Terwijl de subschaal Taal vooral items bevat die te maken hebben met

taalstimulering (helpen taal te begrijpen en taal te gebruiken, aanwezigheid en gebruik van boeken, informeel taalgebruik, gebruik van taal om logisch denken te stimuleren), ligt in de subschaal Interacties het accent op supervisie, discipline, de sensitiviteit van leidster-kind interacties en het stimuleren van interacties tussen kinderen onderling. Deze aspecten tonen enige overlap met de eerder genoemde leidstervaardigheden.

Tabel 3.7
Resultaten Validatie NCKO-Leidsterschalen

Validatie-instrumenten	LV-Totaal	SR	RA	PU	OS	BI
<i>Convergente validiteit</i>						
▪ CIS-Totaal	.50**	.48**	.34**	.46**	.44**	.32**
▪ ORCE-opdringerigheid	-.45**	-.49**	-.44**	-.30**	-.21**	-.45**
▪ ITERS-R/ECERS-R Totaal	.49**	.47**	.25*	.44**	.43**	.43**
▪ ITERS-R/ECERS-R Taal	.48**	.42**	.34*	.43**	.44**	.33**
▪ ITERS-R/ECERS-R Interacties	.45**	.43**	.22	.36**	.44**	.39**
▪ ITERS-R/ECERS-R Activiteiten	.25*	.24*	.11	.27*	.20	.22
<i>Discriminante validiteit</i>						
▪ EAS-Sociability	.03	.04	-.02	-.01	.02	.08
<i>Predictieve validiteit</i>						
▪ WOL	.05	-.10	.06	.03	.02	.21
▪ BITSEA-probleemgedrag	-.05	.01	-.07	-.16	.09	-.10
▪ BITSEA-competentie	.36**	.25	.34**	.36**	.30*	.26*

Noot: Grijs vlak = verwachting geformuleerd, vet gedrukt = verwachting uitgekomen.

* p < .05, ** p < .01

De laatste verwachting met betrekking tot de convergente validiteit van de NCKO-leidsterschalen betreft *ontwikkelingsstimulering* (verwachting 5). *Ontwikkelingsstimulering* hing, zoals verwacht, positief samen met de ITERS-R/ECERS-R subschaal Taal, maar hing, tegen de verwachting in, niet statistisch significant samen met de subschaal ITERS-R/ECERS-R Activiteiten.

Discriminante validiteit

De discriminante validiteit van de NCKO-leidsterschalen werd getoetst met behulp van correlaties met de subschaal Sociability van de EAS. De verwachting (zie Tabel 3.1, verwachting 7) dat er *geen* verband zou bestaan tussen de NCKO-leidsterschalen en het temperament van de leidster werd bevestigd. De resultaten geven aan dat leidsters die de voorkeur geven aan het samen zijn en werken met anderen niet per definitie ook kwalitatief goede interacties met anderen (in dit geval kinderen) hebben.

Predictieve validiteit

De predictieve validiteit van de NCKO-leidsterschalen werd getoetst op groepsniveau. De algemene verwachting was dat het gemiddelde leidstergedrag (NCKO *Totaalscore*) zou samenhangen met het kindgedrag (welbevinden, probleemgedrag, competentie). Alleen de verwachting met betrekking tot de competentie van kinderen (zie Tabel 3.1, verwachting 10) werd deels bevestigd; in groepen met gemiddeld een hogere totaalscore op het leidstergedrag was de gemiddelde groepsscore voor competentie hoger dan in groepen met een lagere totaalscore op het leidstergedrag. Het gemiddelde leidstergedrag hing niet samen met het welbevinden en probleemgedrag van de kinderen.

3.2. NCKO-Kindschalen

3.2.1. Beschrijvende data

In Tabel 3.8 wordt een overzicht gegeven van de gemiddelde scores voor het welbevinden en de betrokkenheid op kindniveau (over alle sessies heen, $N = 623$ kinderen) en op groepsniveau (over alle kinderen in één groep heen, $N = 72$ groepen). De waarden voor skewness en kurtosis geven aan dat de scores (zowel op kind- als op groepsniveau) voor welbevinden boven het schaalgemiddelde liggen en dat de scores voor betrokkenheid onder het schaalgemiddelde van de totale groep liggen. Opvallend is de beperkte spreiding van scores.

Tabel 3.8
Beschrijvende data NCKO-kindschalen

Variabele	<i>M</i>	<i>SD</i>	Range	Schaalwaarden	Skewness*	Kurtosis*
Kindniveau (<i>N</i> = 623)						
Welbevinden	4.27	.50	1.5 – 6.00	1 – 7	-4.55	12.05
Betrokkenheid	2.38	.60	1.0 – 4.33	1 – 5	2.36	.36
Groepsniveau (<i>N</i> = 72)						
Welbevinden	4.24	.26	3.38 - 5.05	1 – 7	-1.33	4.33
Betrokkenheid	2.36	.33	1.61 – 3.14	1 – 5	-.48	.23

* Skewness = skewness/standard error; kurtosis = kurtosis/standard error. Bij een normale verdeling vallen deze waarden tussen -2.00 en 2.00.

3.2.2 Betrouwbaarheid

Inter-observator betrouwbaarheid

De observatoren voldeden aan de voorspelde verwachting ($ICC \geq .70$ op iedere schaal) met betrekking tot de inter-observator betrouwbaarheid voor de kindschalen. De gemiddelde ICC-waarden over alle observatoren heen waren als volgt: *welbevinden* = .84, *betrokkenheid* = .79, wat aangeeft dat er op beide schalen een hoge overeenstemming was tussen de observatoren en de consensusscore.

Test-hertest betrouwbaarheid

Voor het bepalen van de test-hertest betrouwbaarheid van de kindschalen waren de scores van 319 kinderen beschikbaar uit 58 groepen. De kinderen werden gefilmd op twee verschillende dagen (gemiddeld 3 maanden uit elkaar) in hun eigen groep. Omdat de kindscores samengevoegd worden om tot een indicatie van het welbevinden en de betrokkenheid op groepsniveau te komen, werd de test-hertest betrouwbaarheid ook op groepsniveau bepaald. Tabel 3.9 laat de beschrijvende data van de test en van de hertest zien en tevens de test-hertest correlatie.

De gegevens in Tabel 3.9 laten zien dat de scores voor welbevinden en betrokkenheid tijdens de hertest hoger waren dan tijdens de pretest. Voor welbevinden zijn deze verschillen statistisch significant (kindniveau: $t(318) = -5.24, p < .001$; groepsniveau: $t(57) = -4.56, p < .001$), maar voor betrokkenheid niet. Dit uit zich ook in de test-hertest betrouwbaarheid. De test-hertest betrouwbaarheid voor welbevinden is, op zowel kind- als groepsniveau, onvoldoende. De test-hertest betrouwbaarheid voor betrokkenheid is laag op kindniveau en, in overeenstemming met de verwachting, voldoende op groepsniveau. Deze resultaten geven aan dat welbevinden, in dit

onderzoek, geen stabiel kenmerk is. Betrokkenheid (mits gemeten op groepsniveau) daarentegen wel.

Tabel 3.9

Gemiddelde scores en standaarddeviaties en test-hertest correlaties voor de NCKO-kindschalen

Kindschaal	Test	Hertest	Test-hertest correlatie
	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>r</i>
Kindniveau (<i>N</i> = 319)			
Welbevinden	4.26 (.51)	4.44 (.46)	.17*
Betrokkenheid	2.38 (.59)	2.43 (.52)	.31**
Groepsniveau (<i>N</i> = 58)			
Welbevinden	4.21 (.33)	4.45 (.30)	.14
Betrokkenheid	2.33 (.38)	2.37 (.36)	.48**

* $p < .05$ ** $p < .001$

Interne consistentie

Net als bij de NCKO-leidsterschalen verwachtten we ook bij de kindschalen een positieve, maar matige, *onderlinge samenhang*. Welbevinden en betrokkenheid worden in het NCKO-kwaliteitsmodel immers gezien als verschillende aspecten van kindgedrag. De samenhang tussen welbevinden en betrokkenheid (gemeten met de NCKO-Kindschalen) was $r = .43$ ($p < .05$) op kindniveau en $r = .59$ ($p < .001$) op groepsniveau, wat in overeenstemming is met de verwachting.

Tabel 3.10 laat de samenhang zien tussen de NCKO-leidster- en kindschalen. De verwachting dat de kindschalen positief doch matig samenhangen met de leidsterschalen werd gedeeltelijk bevestigd. Het welbevinden van de kinderen hangt significant samen met hogere leidsterscores voor *sensitieve responsiviteit*, *respect voor autonomie* van de kinderen en *praten en uitleggen* maar niet met *ontwikkelingsstimulering* of het *begeleiden van interacties* tussen kinderen. De betrokkenheid van de kinderen hangt significant samen met hogere leidsterscores voor *respect voor autonomie* van de kinderen en *praten en uitleggen*, maar niet met de overige leidsterschalen. Het gemiddelde leidstergedrag (NCKO *Totaalscore*) hangt wel significant samen met welbevinden, maar niet met betrokkenheid.

Tabel 3.10
Correlaties NCKO-leidsterschalen en NCKO-kindschalen

NCKO-Leidsterschalen	NCKO-kindschalen	
	Welbevinden	Betrokkenheid
Sensitieve responsiviteit	.24**	.14
Respect voor autonomie	.35**	.23*
Praten en uitleggen	.36**	.29*
Ontwikkelingsstimulering	.15	.04
Begeleiden van interacties	.17	.12
NCKO-Totaalscore (leidsters)	.31**	.20

* $p \leq .05$, ** $p < .01$

3.2.3. Validiteit

Net als de leidsterschalen werden ook de kindschalen getoetst op convergente validiteit, discriminante validiteit en predictieve validiteit. De verwachtingen met betrekking tot deze drie valideringsmethoden werden al eerder in Tabel 3.2 gepresenteerd. Tabel 3.11 laat de beschrijvende data zien van alle variabelen die (naast de kindschalen zelf) gebruikt zijn bij de validering. Tabel 3.12 geeft de valideringsresultaten weer.

Convergente validiteit

De convergente validiteit van de NCKO-kindschalen (NCKO-Welbevindenschaal en de Leuvense Betrokkenheidschaal) werd getoetst aan de hand van correlaties. De verwachtingen 11 (Er bestaat een positief verband tussen de NCKO-Welbevindenschaal en de WOL) en 12 (Er bestaat een positief verband tussen de Leuvense Betrokkenheidsschaal en de ORCE variabele *aandacht voor objecten*) werden bevestigd (zie Tabel 3.12). Het verband voor de welbevindenschaal is echter zwak.

Tabel 3.11

Beschrijvende data voor de instrumenten gebruikt ter validering van de NCKO-kindschalen

	N	M	SD	Range	Mogelijke Schaalwaarden
<i>Convergente validiteit</i>					
▪ WOL	246	4.87	.49	3.42 – 5.92	1 – 6
▪ ORCE-aandacht voor objecten	328	2.00	.55	1.00 – 3.50	1 – 4
<i>Discriminante validiteit</i>					
▪ ICQ	246	3.31	.69	1.47 – 5.47	1 – 7
<i>Predictieve validiteit</i>					
▪ BITSEA-probleemgedrag	248	.25	.15	0 - .93	0 – 2
▪ BITSEA-competentie	248	1.42	.33	.45 – 2.00	0 – 2

Tabel 3.12

Resultaten Validatie NCKO-kindschalen

Validatie-instrumenten	Welbevinden	Betrokkenheid
<i>Convergente validiteit</i>		
▪ WOL	.19**	.10
▪ ORCE-aandacht voor objecten	.18**	.42**
<i>Discriminante validiteit</i>		
▪ ICQ	-.05	-.08
<i>Predictieve validiteit</i>		
▪ BITSEA-probleemgedrag	-.04	-.16**
▪ BITSEA-competentie	.25**	.31**

Noot: Grijs vlak = verwachting geformuleerd, vet gedrukt = verwachting uitgekomen.

** p = .01

Discriminante validiteit

De discriminante validiteit van de NCKO-kindschalen werd getoetst met behulp van correlaties met de ICQ (een temperamentsmaat). De niet-significante, lage correlaties ondersteunen de verwachting (zie Tabel 3.2, verwachting 13) dat er geen verband zou bestaan tussen de NCKO-kindschalen en het temperament van de kinderen.

Predictieve validiteit

De predictieve validiteit van de kindschalen werd getoetst aan de relatie met de variabelen probleemgedrag en competentie van de BITSEA. De resultaten in Tabel 3.12 ondersteunen wel de predictieve validiteit van de betrokkenheidschaal, maar niet (geheel) die van de NCKO-welbevindenschaal. De Leuvense Betrokkenheidschaal bleek, in overeenstemming met verwachtingen 14 en 15, significant gerelateerd aan de BITSEA-scores voor probleemgedrag en competentie: hogere scores op de Leuvense Betrokkenheidschaal hingen inderdaad samen met hogere scores voor competentie en lagere scores voor probleemgedrag zoals beoordeeld met de BITSEA. Voor de NCKO-Welbevindenschaal werd verwachting 15 wel bevestigd, maar hypothese 16 niet: hogere scores op de NCKO-welbevindenschaal hingen wel samen met hogere scores voor competentie, maar niet met lagere scores voor probleemgedrag op de BITSEA.

3.3 Overzicht van de resultaten

In Tabel 3.13 zijn de hierboven beschreven resultaten van de valideringsstudie nog eens overzichtelijk bij elkaar geplaatst. De resultaten voor de verschillende NCKO-*Leidsterschalen* (m.u.v. de schaal *structureren en grenzen stellen*) zijn samengevat in één kolom, omdat het patroon van uitkomsten voor al deze leidsterschalen hetzelfde was. Omdat voor de schaal *structureren en grenzen stellen* de inter-observator betrouwbaarheid onvoldoende was, is deze schaal verder niet in het onderzoek betrokken. Voor de twee NCKO-*Kindschalen* waren de resultaten niet gelijk; daarom zijn deze in aparte kolommen opgenomen. In de tabel is met plussen en minnen aangegeven of de resultaten van het onderzoek positieve dan wel negatieve evidentie vormen voor de verschillende aspecten van betrouwbaarheid en validiteit die we in deze studie onder de loep namen (voor omschrijvingen zie par. 1.4 van dit hoofdstuk).

Tabel 3.13

Overzicht resultaten valideringsonderzoek NCKO-Leidster- en -Kindschalen

	NCKO- Leidsterschalen *	NCKO- Kindschalen	
		Welbevinden	Betrokkenheid
<i>Betrouwbaarheid</i>			
Inter-observator betrouwbaarheid	+	+	+
Test-hertest betrouwbaarheid	+	-	+
Onderlinge samenhang	+		+
<i>Validiteit</i>			
Convergente validiteit	+	+	+
Discriminante validiteit	+	+	+
Predictieve validiteit	+/-	+/-	+

* Met uitzondering van de schaal structureren en grenzen stellen

+ Resultaten ondersteunen betrouwbaarheid/validiteit van het instrument

- Resultaten ondersteunen betrouwbaarheid/validiteit van het instrument niet

4. DISCUSSIE EN CONCLUSIES

De resultaten van de valideringsstudie laten zien dat de NCKO-leidsterschalen, met uitzondering van één schaal, betrouwbaar en valide zijn en daarmee op een wetenschappelijk verantwoorde manier de kwaliteit van leidster-kind interacties in kinderdagverblijven in beeld kunnen brengen. Met betrekking tot de NCKO-kindschalen zijn de resultaten minder eenduidig. Terwijl de betrokkenheidschaal zowel betrouwbaar als valide blijkt te zijn, voldoet de welbevindenschaal niet helemaal aan de verwachtingen wat betreft de getoetste psychometrische kwaliteiten. De resultaten worden in de volgende paragrafen nog eens uiteengezet en besproken.

Betrouwbaarheid

De resultaten met betrekking tot de *inter-observator betrouwbaarheid* zagen er voor vijf van de zes leidsterschalen en voor beide kindschalen goed uit; de observatoren behaalden de vooraf vastgestelde kritische grens ruimschoots. Dit geeft aan dat de schaalwaarden voor deze instrumenten zó duidelijk geformuleerd zijn dat getrainde observatoren vrijwel dezelfde beoordeling geven als zij het gedrag van de leidsters of kinderen in de groep observeren. Voor de

leidsterschaal *structureren en grenzen stellen* werd de kritische grens echter niet bereikt. Het voornaamste probleem was dat de schaal niet voldoende verschillen tussen leidsters te zien gaf. Het is mogelijk dat dit meer te maken heeft met de situatie waarin we de leidsters filmden dan met de psychometrische kwaliteiten van de schaal. Structuren en grenzen stellen is een vaardigheid die bij uitstek zichtbaar wordt als leidsters iets van kinderen verwachten, zoals in een instructiesituatie. De schalen voor *structureren en grenzen stellen* van Erickson e.a. (1985) en van De Schipper en Riksen-Walraven (zie De Schipper e.a., 2006), die het voornaamste uitgangspunt vormden voor de NCKO-leidsterschalen, zijn beide met succes gebruikt in met name dit soort instructiesituaties. In het huidige onderzoek wilden we echter een beeld krijgen van de *alledaagse* leidster-kind interacties in allerlei situaties die zich door de dag heen voordoen. We hebben er daarom voor gekozen om de leidsters op bepaalde tijden te filmen, ongeacht de activiteit die zij op dat moment uitvoerden. De situatie die zich daarbij verreweg het meeste bleek voor te doen was vrij spel; een situatie waarbij de kinderen zelf de richting in het spel bepalen en leidsters zich veelal afzijdig houden wanneer de kinderen niet speciaal een beroep op hen doen. Dit had tot gevolg dat op veel van de gefilmde episodes nauwelijks sprake was van structureren en grenzen stellen door de leidsters, simpelweg omdat dit op dat moment niet nodig was. Omdat *structureren en grenzen stellen* in het NCKO-model als een belangrijke leidstervaardigheid wordt beschouwd en omdat deze vaardigheid ook door verschillende groepen betrokkenen in het veld (met name ouders en leidsters) als een belangrijk kenmerk van goede kinderopvang wordt gezien (zie Hoofdstuk 2, Tabel 2.1), is besloten om de schaal voorlopig in het NCKO-instrument te behouden en om de filmprocedure en zo nodig de schaal zelf aan te passen om de betrouwbaarheid van de observaties te verhogen. In het vervolg zal meer gericht worden gefilmd in situaties die een beroep doen op de vaardigheid van leidsters tot structureren en grenzen stellen, zoals *transities* tussen alledaagse activiteiten, bijvoorbeeld de transitie van vrij spel naar lunch, waarbij onder andere opgeruimd moet worden en de kinderen hun handen moeten wassen.

De *test-hertest resultaten* laten zien dat gebruik van de NCKO-leidsterschalen en de Leuvense betrokkenheidschaal op twee verschillende tijdstippen (met een tussentijd van 3 maanden) overeenkomstige resultaten opleveren. Voor de welbevindenschaal bleek dit echter niet het geval. De zwakke samenhang tussen de twee achtereenvolgende metingen van het welbevinden van de kinderen kan globaal twee dingen betekenen. Op de eerste plaats is het mogelijk dat de observaties op zichzelf wel een goed beeld geven van het welbevinden van de kinderen op de betreffende dag, maar dat dit welbevinden voor een groot deel mede bepaald wordt door andere factoren dan de kwaliteit van de opvang, zoals de gezondheidstoestand van het kind of de situatie thuis. Verandering van het welbevinden betekent dan niet per definitie dat de

kwaliteit van de opvang veranderd is, maar kan ook duiden op verandering in andere factoren die het welbevinden van het kind beïnvloeden. Een tweede mogelijkheid is dat het welbevinden van kinderen zo instabiel is over de dag heen, dat de relatief korte observatietijd (er is 3 à 4 maal 2 minuten per kind gefilmd) geen representatief beeld oplevert van het welbevinden van het kind op die dag. De goede test-hertest betrouwbaarheid voor de betrokkenheidsschaal en de betere resultaten voor deze schaal met betrekking tot de predictieve validiteit suggereren dat betrokkenheid een stabiel kindkenmerk is, dat in dezelfde beperkte observatietijd wèl betrouwbaar kan worden gemeten. Aangezien langer filmen van de kinderen in het kader van een kwaliteitsmeting op een kinderdagverblijf praktisch gezien niet haalbaar is, gaan we onderzoeken of het welbevinden van de kinderen op een meer betrouwbare manier kan worden vastgesteld met behulp van een andere schaal, namelijk de welbevindenschaal die in het onderzoek van De Schipper et al. (2006) met succes is toegepast om het welbevinden van een *groep* kinderen over een periode van 5-15 minuten te beoordelen; het gebruik van deze schaal kan in principe goed worden ingepast in de NCKO-kwaliteitsmeting.

De *onderlinge samenhang* tussen de vijf afzonderlijke NCKO-leidsterschalen en tussen de twee afzonderlijke NCKO-kindschalen ondersteunde de betrouwbaarheid van deze schalen. De samenhang tussen de NCKO-leidster- en -kindschalen ten slotte was niet geheel volgens verwachting. De gestandaardiseerde totaalscore voor leidstervaardigheden hing wel samen met welbevinden, maar niet met betrokkenheid. Daarnaast bleken de leidsterschalen *ontwikkelingsstimulering* en *begeleiden van interacties* niet samen te hangen met de kindschalen. Het is mogelijk dat de spreiding op beide schalen te beperkt is om effecten te laten zien op het kindgedrag. De samenhang tussen welbevinden en betrokkenheid enerzijds en verschillende aspecten van het leidstergedrag anderzijds biedt toch enige ondersteuning voor de validiteit van het NCKO-meetinstrument als geheel - verschillende schalen meten verschillende aspecten van eenzelfde construct (de proceskwaliteit van kinderopvang).

Validiteit

De verwachtingen met betrekking tot de *convergente validiteit* van de leidster- en kindschalen werden grotendeels bevestigd. Ook hier werden geen zeer hoge correlaties verwacht gegeven het feit dat de NCKO-schalen ontworpen waren om andere leidster- en kindkenmerken te meten dan de bestaande schalen. De correlaties tussen de NCKO-schalen en de valideringsinstrumenten bleken vrijwel allemaal significant en matig sterk te zijn, hetgeen de convergente validiteit van de schalen bevestigt. Alleen de correlatie tussen de NCKO-schaal *ontwikkelingsstimulering* en de ITERS-R/ECERS-R subschaal Activiteiten was, in tegenstelling tot de verwachting (verwachting

5) niet significant. Nadere inspectie van de items binnen de subschalen van de ITERS-R en ECERS-R geeft aan dat deze resultaten mogelijk te maken hebben met de opbouw van deze subschalen. Terwijl de subschaal Taal (die wel een verband liet zien met de NCKO-schaal *ontwikkelingsstimulering*) al op minimaal niveau (score 3) indicatoren voor taalstimulering bevat, zijn indicatoren voor de subschaal Activiteiten op dit niveau nog gericht op de aanwezigheid en toegankelijkheid van materialen. Indicatoren voor ontwikkelingsstimulering komen pas op een hoger niveau (score 5) aan de orde. De gemiddelde taalscore lag *boven* het minimale niveau ($M = 3.93$) wat aangeeft dat leidsters gebruik maakten van materialen en activiteiten om taal en logisch redeneren te stimuleren. De gemiddelde score voor Activiteiten lag echter *onder* het minimale niveau ($M = 2.46$), wat aangeeft dat er over het algemeen niet voldoende geschikte materialen beschikbaar waren. Omdat niet voldaan werd aan het minimale niveau, werd de mate van ontwikkelingsstimulering van de leidster tijdens activiteiten en met materialen niet beoordeeld. Dit bevestigt nog eens dat de NCKO-leidsterschalen, waarmee de interactievaardigheden van de leidster onafhankelijk van andere factoren beoordeeld kunnen worden, inderdaad in een lacune voorzien.

Wat betreft de *convergente validiteit* van de NCKO-kindschalen bleek de correlatie tussen de NCKO-welbevindenschaal en het valideringsinstrument de *WOL* weliswaar significant, maar toch veel lager dan de correlatie tussen de Leuvense Betrokkenheidschaal en het valideringsinstrument de *ORCE*-schaal *aandacht voor objecten*. Dit verschil is waarschijnlijk toe te schrijven aan het feit dat de beoordeling van het welbevinden van de kinderen met de NCKO-schalen en met de *WOL* gebaseerd is op twee heel verschillende observatieperioden (6 à 8 minuten per kind bij de NCKO-schaal en 3 weken voor de *WOL*), terwijl de beoordeling van de betrokkenheid van het kind m.b.v. de NCKO-schaal en de *ORCE*-schaal geschiedde op basis van *dezelfde* videofragmenten. De lage correlatie tussen de NCKO-welbevindenschaal en de *WOL* bevestigt het vermoeden (eerder geuit bij de bespreking van de lage test-hertest betrouwbaarheid van de NCKO-welbevindenschaal) dat het welbevinden van jonge kinderen geen erg stabiel kenmerk is en dat er waarschijnlijk veel meer observatietijd nodig is om met behulp van de NCKO-welbevindenschaal een goed beeld van dit kindkenmerk te krijgen.

De resultaten met betrekking tot de *discriminante validiteit* van de leidster- en kindschalen waren eenduidig. Het ontbreken van een significante samenhang tussen de NCKO-schalen en temperamentskenmerken van de leidsters (Tabel 3.1, verwachting 7) laat zien dat de NCKO-leidsterschalen een ander construct meten dan temperament. Specifieker, leidsters die de voorkeur geven aan het samenzijn of werken met anderen in plaats van alleen te zijn of werken

(die dus hoog scoorden op de temperament-subschaal *sociability*) blijken daarmee niet automatisch de vaardigheden te bezitten die een goede leidster-kind interactie typeren. De onderzoeksresultaten bevestigden ook de discriminante validiteit van de NCKO-kindschalen: er werd geen significante samenhang gevonden tussen het welbevinden en de betrokkenheid van kinderen enerzijds en het temperament van de kinderen anderzijds (Tabel 3.2, verwachting 13). De NCKO-welbevindenschaal en de Leuvense betrokkenheidsschaal meten dus niet of de kinderen een gemakkelijk of moeilijk temperament hebben.

De *predictieve validiteit* van de NCKO-schalen werd ondersteund door hun significante positieve correlatie met de sociale competentie van de kinderen zoals door de leidsters beoordeeld met behulp van de BITSEA: kinderen wier leidsters als meer vaardig werden beoordeeld op de NCKO-leidsterschalen en kinderen die tijdens de observaties meer welbevinden en betrokkenheid lieten zien, gaven blijk van hogere niveaus van sociale competentie. Het feit dat een hogere proceskwaliteit zoals gemeten met de NCKO-leidsterschalen, *niet* het welbevinden van de kinderen zoals gemeten met de WOL en *niet* het probleemgedrag van de kinderen zoals gemeten met de BITSEA voorspelde, was in tegenstelling tot de verwachting. Kinderen die een minder vaardige leidster hadden vertoonden dus *niet* een lager niveau van welbevinden over de afgelopen drie weken en zij vertoonden *geen* hogere niveaus van probleemgedrag. Deze uitkomst kan beschouwd worden als contra-evidentie voor de validiteit van de NCKO-leidsterschalen, maar zij zou ook kunnen betekenen dat de leidsterschalen wél een goede indicatie vormen van de proceskwaliteit van de opvang (hetgeen gesuggereerd wordt door de andere uitkomsten van de valideringsstudie), maar dat in Nederland – in tegenstelling tot in de Verenigde Staten – de proceskwaliteit van de kinderopvang op dit moment *werkelijk* niet gerelateerd is aan het welbevinden van kinderen en aan het voorkomen van gedragsproblemen. Een verklaring daarvoor zou kunnen liggen in verschillen tussen beide landen op het gebied van de opvang in kindercentra. De gegevens van de valideringsstudie zijn niet toereikend om hierover uitsluitel te geven.

Conclusie en opzet van het complete NCKO-instrument

De resultaten van de valideringsstudie bevestigen in het algemeen de betrouwbaarheid en validiteit van de NCKO-leidster- en kindschalen. Voor twee van de schalen (*structureren en grenzen stellen* en *welbevinden*) waren de uitkomsten van dien aard dat aanpassing van de gebruikte observatieprocedure of van de schaal geïndiceerd is. Momenteel wordt aan deze aanpassing gewerkt.

De NCKO-leidster- en kindschalen zullen – als directe indicatoren van het pedagogische proces – de kern vormen van het NCKO-instrument voor de meting van de pedagogische kwaliteit van de kinderopvang. De scores op deze schalen weerspiegelen het meest direct de kwaliteit van de dagelijkse ervaringen van de kinderen.

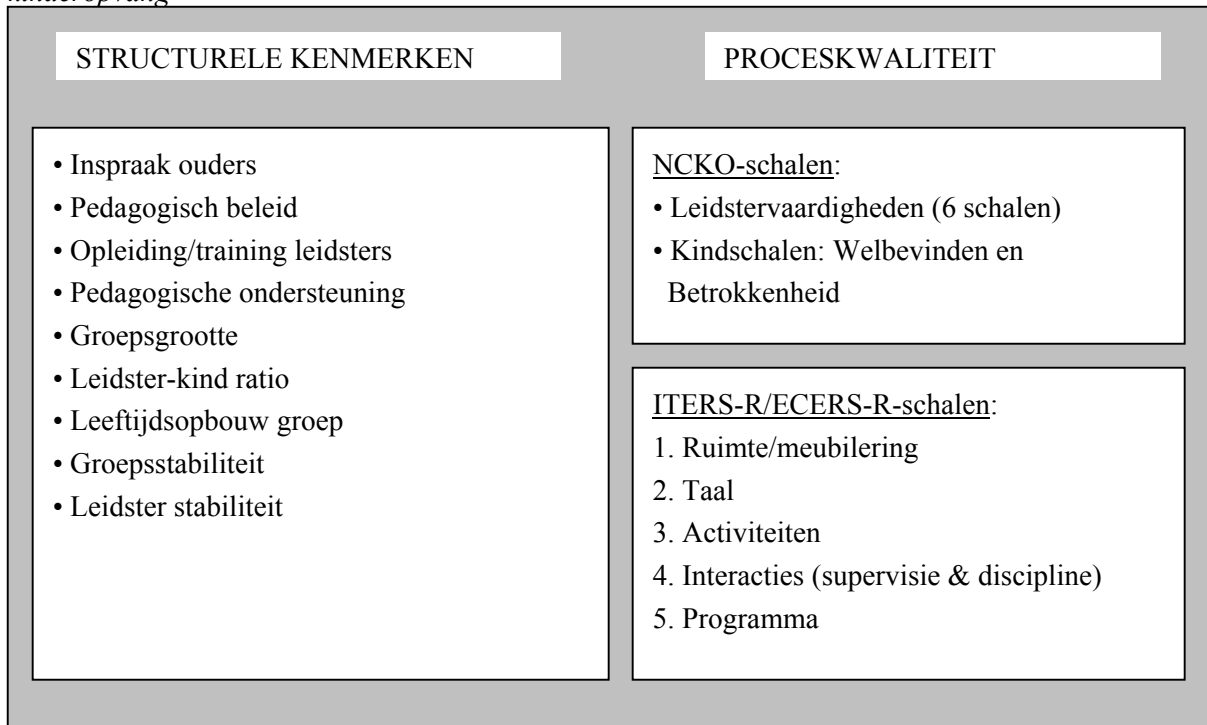
Naast de leidster- en kindschalen zullen in het NCKO-instrument ook enkele schalen uit de ITERS-R/ECERS-R worden opgenomen die informatie geven over aspecten van de proceskwaliteit die met de leidster- en kindschalen minder goed observeerbaar zijn. Het gaat daarbij om een vijftal schalen die betrekking hebben op de inrichting (*Meubilering en inrichting*), op de beschikbaarheid en het gebruik van educatieve materialen (*Activiteiten en Taal*), op aspecten van het leidstergedrag die beter in vivo dan op video-opnamen zijn te beoordelen (*Interacties*, d.w.z. supervisie en disciplineren) en de structuur en inhoud van het dagprogramma (*Programma*). Zoals aangegeven in Figuur 3.1 vormen deze vijf schalen, naast de NCKO-leidster- en kindschalen, in het NCKO-instrument een tweede domein van de proceskwaliteit. De ITERS-R/ECERS-R schaal *Individuele zorg* wordt niet in het NCKO-instrument opgenomen, omdat de betreffende interactiekwaliteiten van de leidsters tijdens de individuele zorg al vertegenwoordigd zijn in de NCKO-leidsterschalen.

Zoals te zien in Figuur 3.1 heeft het derde domein van het NCKO-instrument betrekking op *structurele* kwaliteitskenmerken. Zoals aangegeven in het NCKO-kwaliteitsmodel (zie Hoofdstuk I en II van dit rapport) gaat het daarbij om reguleerbare kenmerken van de opvang die van invloed zijn op de proceskwaliteit en die beschouwd kunnen worden als voorwaarden voor het realiseren van een hoge proceskwaliteit.

Het uiteindelijke NCKO-instrument voor de meting van de pedagogische kwaliteit van de kinderopvang zal dus een kwaliteitsprofiel opleveren met indicatoren in drie domeinen. Dit kwaliteitsprofiel is bedoeld om een genuanceerd en compleet beeld te geven van de pedagogische kwaliteit en om als uitgangspunt te dienen voor de verbetering en bewaking van de pedagogische kwaliteit van de opvang in Nederlandse kinderdagverblijven.

Figuur 3.1.

Overzicht van het complete NCKO-instrument voor de meting van de pedagogische kwaliteit van kinderopvang



IV. SLOTWOORD

De valideringsstudie heeft een betrouwbaar en valide instrument voor toekomstig peilingsonderzoek opgeleverd dat beproefd is in de praktijk van de Nederlandse kinderopvang. Dit instrument legt daarmee een fundament voor het meten van de kwaliteit van de kinderopvang in Nederland. Het ontwikkelde instrumentarium maakt het mogelijk structurele kwaliteitskenmerken, proceskwaliteit en kindkenmerken te bepalen én deze met elkaar in verband te brengen, recht doend aan het kwaliteitsmodel voor de kinderopvang en de opvattingen die leven in het Nederlandse veld. Ook is het mogelijk nieuwe bevindingen uit toekomstig onderzoek te (blijven) verbinden met uitkomsten uit reeds uitgevoerd peilingsonderzoek in Nederland. De bevindingen uit toekomstige peilingen kunnen dus worden afgezet tegen de vorige peilingen uit 1995, 2001 en de kwaliteitsmeting uit 2005 (zie Vermeer e.a., 2005) om een vergelijking te maken in de tijd. Wel moet worden opgemerkt dat toekomstige metingen met een nieuw en breder instrument een completer – en in die zin: een ander – beeld zullen geven van de pedagogische kwaliteit van kinderdagverblijven. Het blijft tevens mogelijk de kwaliteit van de Nederlandse kinderopvang op kernaspecten te spiegelen aan die van het buitenland, met inachtneming van de unieke kenmerken van de Nederlandse kinderopvangsituatie. Concrete stappen worden op dit moment gezet om een handzaam profiel op te kunnen stellen op basis van het complete instrumentarium. Deze uitsnede uit het omvangrijke instrument moet het mogelijk maken gericht te kunnen werken aan kwaliteitsbepaling en eventuele verbetering van de kernkwaliteit van kinderdagverblijven. In dit profiel wordt aandacht voor structurele kwaliteit gecombineerd met aandacht voor de proceskwaliteit, zoals gemeten met de NCKO-leidster- en kindschalen. Om een breder beeld te verkrijgen van de proceskwaliteit in een kinderdagverblijf wordt het profiel verder aangevuld met schalen uit de vertaalde en bewerkte ITERS-R/ECERS-R (zoals Meubilering/inrichting, Activiteiten, Taal, Interacties en Programma).

Met gebruikmaking van het ontwikkelde instrument kan nu ook een valide antwoord worden gegeven op de wetenschappelijke vraag die aan de wieg heeft gestaan van het longitudinale NCKO-project en die – tot op de dag van vandaag – alle partijen bezig houdt: wat is de pedagogische kwaliteit van de Nederlandse kinderopvang? Wat zijn de sterkere kanten en wat zijn de zwakkere kanten? En wat zijn de effecten van de Nederlandse kinderopvang op de ontwikkeling van kinderen, lettend op de pedagogische kwaliteit van de individuele opvangcentra? Deze vragen zijn niet beantwoord in dit onderzoek, waar immers de validering van een nieuw instrument centraal stond. Anders gezegd, ‘de toets is getoetst’ (en dus niet de

Nederlandse kinderopvang). De descriptieve toetsgegevens uit dit rapport lenen zich dan ook nadrukkelijk alleen voor een interpretatie vanuit valideringsperspectief. Bezien vanuit dit perspectief is deze studie niets meer, maar ook niets minder ('goed gereedschap is het halve werk') dan een noodzakelijke 'tussenstop'. De beantwoording van de bovengenoemde vragen blijft daarmee wel actueel, ook omdat de Nederlandse kinderopvangsector aan verandering onderhevig is (of beter gezegd: blijft).

Dit rapport is het samenvattende verslag van de valideringsstudie. Een deel van de resultaten uit dit onderzoek is echter al eerder gepubliceerd. Afgestemd op de onderscheiden doelgroepen, zijn de resultaten uit de enquête al in verschillende tijdschriften gepubliceerd (zie NCKO, 2004a, b, c; Fukkink e.a., 2005). Ook in de nabije toekomst zullen wij, waar mogelijk, op vergelijkbare wijze rapporteren over de opzet en pedagogische achtergrond van het nieuwe instrument. Gezien de aandacht voor leidstervaardigheden vormen met name de leidsters en de opleiders van leidsters hier een voor de hand liggende doelgroep.

Niet alleen de onderzoeksbevindingen, maar ook het instrument zelf zal worden 'vertaald' naar de doelgroep met de geplande ontwikkeling van een veldinstrument, dat afgeleid wordt van het wetenschappelijke instrument. Het laatste instrument stond centraal in deze studie en in dit verslag. De inhoud van dit rapport (maar ook de precieze vormgeving van de instrumenten) heeft dus nadrukkelijk alleen betrekking op de wetenschappelijke versie van het NCKO-instrument en niet op deze toekomstige versie, die bovendien in een heel andere context zal worden gebruikt.

REFERENTIES

- Arnett, J. (1989). Caregivers in day-care centers: Does training matter? *Journal of Applied Developmental Psychology, 10*, 541-552.
- Ballex, M. (2005). *Werken aan welbevinden. Evalueren door observeren*. 3e druk. Amsterdam: SWP.
- Bates, J.E., Freeland, C.A.B., & Lounsbury, M.L. (1979). Measurement of infant difficultness. *Child Development, 50*, 794-803.
- Briggs-Gowan, M. J. & Carter, A. S. (2001). *Brief-Infant Toddler Social and Emotional Assessment (BITSEA) Manual. Version 1.0*. University of Massachusetts, Boston.
- Briggs-Gowan, M. J., Carter, A. S., Irwin, J. R., Wachtel, K., & Cicchetti, D. V. (2004). The Brief-Toddler Social and Emotional Assessment: Screening for social-emotional problems and delays in competence. *Journal of Pediatric Psychology, 29*, 143-155.
- Buss, A & Plomin, R (1984). *Temperament: Early developing personality traits*. Hillsdale, NJ: Lawrence Erlbaum.
- Burchinal, M.R., Roberts, J., Riggins Jr., R., Zeisel, S.A., Neebe, E., & Bryant, D. (2000). Relating quality of center-based child care to early cognition and language development longitudinally. *Child Development, 71*, 339-357.
- Czicszenmihayli, M. (1979). The concept of flow. In B. Sutton-Smith (red.), *Play and learning*. New York: Gardner Press.
- Cryer, D., Tietze, W., & Wessels, H. (2002). Parents' perceptions of their children's child care: a cross-national comparison. *Early Childhood Research Quarterly, 17*, 259-277.
- De Kruif, R.E.L., Reiling, E.J. & Tavecchio, L.W.C. (2006). *Infant/Toddler Environment Rating Scale – NL*. Nederlandse vertaling van Harms, T., Cryer, D. & Clifford, R. M. (2003) *Infant/Toddler Environment Rating Scale – Revised*. Universiteit van Amsterdam.
- De Kruif, R.E.L., Reiling, E.J. & Tavecchio, L.W.C. (2006). *Early Childhood Environment Rating Scale – NL*. Nederlandse vertaling van Harms, T., Clifford, R. M. & Cryer, D. (1998) *Early Childhood Environment Rating Scale – Revised*. Universiteit van Amsterdam.
- De Kruif, R.E.L., & Tavecchio, L.W.C. (2004). Instrumenten voor meting van kwaliteit van kinderopvang: beschrijving en evaluatie. In M.H. van IJzendoorn, L.W.C. Tavecchio, & J.M.A. Riksen-Walraven (Red.), *De kwaliteit van de Nederlandse kinderopvang* (pp. 124-153). Amsterdam: Boom.
- De Schipper, J.C., Van IJzendoorn, M.H., & Tavecchio, L.W.C. (2002). Infant's well-being and problem behavior in center day care: Relations with stability of care, quality of care and family management stress. Manuscript submitted for publication. Leiden University.

-
- De Schipper, E.J., Riksen-Walraven, J.M.A., & Geurts, S.A.E. (2006). Effects of child-caregiver ratio on caregiver-child interaction in child-care centers: an experimental study. *Child Development, 77*(4), 861-874.
- De Wolff, M. & Van IJzendoorn, M.H. (1997). Sensitivity and attachment: A meta-analysis on parental antecedents of infant attachment. *Child Development, 68*, 571-591.
- Erickson, M.F., Sroufe, L.A., & Egeland, B. (1985). The relationship between quality of attachment and behavior problems
- Feagans, L.V., & Manlove, E.E. (1994). Parents, infants, and day-care teachers: Interrelations and implications for better child care. *Journal of Applied Developmental Psychology, 15*, 585-602.
- Fukkink, R.G, Tavecchio, L.W.C, de Kruif, R.E.L., Vermeer, H.J., & van Zeijl, J. (2006). Criteria voor kwaliteit van kinderopvang: Visies van sleutelfiguren. *Pedagogiek, 25*(4), 243-261.
- Fukkink, R.G. & Van IJzendoorn, M.H (2004a). Dertig jaar onderzoek naar de Nederlandse kinderopvang: algemene trend. In R. van IJzendoorn, L. Tavecchio, & M. Riksen-Walraven (Red.), *De kwaliteit van de Nederlandse kinderopvang* (pp. 31-48). Amsterdam: Boom.
- Fukkink, R.G. & Van IJzendoorn, M.H (2004b). Dertig jaar onderzoek naar de Nederlandse kinderopvang: centrale thema's. In M.H. van IJzendoorn, L.W.C. Tavecchio, & J.M.A. Riksen-Walraven (Red.). *De kwaliteit van de Nederlandse kinderopvang* (pp. 49-76). Amsterdam: Boom.
- Gevers Deynoot-Schaub & Riksen-Walraven (2006). *Young Children's Behavior and Experiences in Child Care Centers: A Longitudinal Study*. Amsterdam: SCO-Kohnstamm Instituut van de Faculteit der Maatschappij en Gedragwetenschappen, Universiteit van Amsterdam.
- Harms, T., Clifford, R.M., & Cryer, D. (1998). *Early Childhood Environment Rating Scale-Revised*. New York: Teachers College Press.
- Harms, T., Cryer, D., & Clifford, R.M. (2003). *Infant/Toddler Rating Scale-Revised*. New York: Teachers College Press.
- Kärrby, G., & Giota, J. (1995). Parental conceptions of quality in daycare centers in relation to quality measured by the ECERS. *Early Child Development and Care, 110*, 1-18.
- Laevers, F., Aerden, I., De Bruyckere, G., Moons, J. & Silkens, K. (2003). *Belevingsonderzoek bij het jonge kind in opvangsituaties buiten huis. Eindrapport*. Leuven: Expertisecentrum Ervaringsgericht Onderzoek.

-
- Linting, M., Groenen, P. J. F., & Tavecchio, L. W. C. (2002). *Maar hoe voelen de kinderen zich? Een onderzoek naar meetinstrumenten voor welbevinden van twee- tot vierjarigen in de kinderopvang*. Universiteit Leiden: Intern rapport.
- NCKO (2004a). Vraag Nederlands Consortium Kinderopvang Onderzoek: Wat is pedagogische kwaliteit? *Management Kinderopvang*, 10(8/9), 34-37.
- NCKO (2004b). Wat vinden ouders bepalend bij 'pedagogische kwaliteit?'. *BOinK*, 9(4), 8-9.
- NCKO (2004c). Wat vinden leidsters belangrijk bij pedagogische kwaliteit? *KIDDO*, 8, 24-27.
- NCKO (2004d). *NCKO-Leidsterschalen*. Amsterdam: NCKO
- NCKO (2004e). *NCKO-Welbevindenschaal*. Amsterdam: NCKO
- Nelson, F., & Garduque, L. (1991). The experience and perception of continuity between home and day care from the perspective of child, mother, and caregiver. *Early Child Development and Care*, 68, 99-111.
- NICHD Early Child Care Research Network (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly*, 11, 269-306.
- Peisner-Feinberg, E.S., Burchinal, M.R., Clifford, R.M., Yazejijian, N., Culkin, M.L., Zelazo, J., Howes, C., Byler, P., Kagan, S.L., Rustici, J. (1999). *The children of the Cost, Quality and Outcome Study go to school: Technical Report*. Chapel Hill, NC: University of North Carolina.
- Riksen-Walraven, J.M.A. (2004). Pedagogische kwaliteit in de kinderopvang: doelstellingen en kwaliteitscriteria. In M.H. van IJzendoorn, L.W.C. Tavecchio, & J.M.A. Riksen-Walraven (Red.), *De kwaliteit van de Nederlandse kinderopvang* (pp. 100-123). Amsterdam: Boom.
- Sharp, C., & Davis, C. (1997). *Parents' view of pre-school quality matters*. Slough (England): National Foundation for Educational Research.
- Tavecchio, L.W.C. (2002). *Van Opvang naar Opvoeding. De emancipatie van een unieke opvoedingsmilieu*. Oratie, Universiteit van Amsterdam. Amsterdam: Vossiuspers UvA.
- Tietze, W., & Viernickel, S. (2002). *Pädagogische Qualität in Tageseinrichtungen für Kinder; Ein nationaler Kriterienkatalog*. Weinheim: Beltz.
- Van IJzendoorn, M.H., Tavecchio, L.W.C., Riksen-Walraven, J.M.A. (2004). *De kwaliteit van de Nederlandse kinderopvang*. Amsterdam: Boom.
- Van IJzendoorn, M.H., Tavecchio, L. W. C., Verhoeven, M. J. E., Reiling, E. J., & Stams, G. J. J. (1996). De kwaliteit van de Nederlandse kinderopvang. *Nederlands Tijdschrift voor Opvoeding, Vorming en Onderwijs*, 12, 286-313.

-
- Vermeer, H.J., Van IJzendoorn, M.H., De Kruif, R.E.L., Fukkink, R.G., Tavecchio, L.W.C., Riksen-Walraven, J.M.A., & Van Zeijl, J. (2005). *Kwaliteit van Nederlandse kinderdagverblijven: Trends in kwaliteit in de jaren 1995-2005*. Amsterdam: NCKO.
- Whitebook, M., Howes, C., & Phillips D. (1989). *Who cares? Child care teachers and the quality of child care in America: Final report. National Child Care Staffing Study*. Oakland, CA: Child Care Employee Project.