



Moreel schoolklimaat en morele competentie bij scholieren van het voortgezet onderwijs: convergerende en discriminerende validiteit

Projectnummer 94742

Stichting voor onderzoek van het Onderwijs (SVO)
Adviescommissie Niet-Gebonden Onderzoek (ANGO)

D. Brugman
L.W.C. Tavecchio
B.J. Van Os
K. Høst
T. Meijer
M. Van Roosmalen

September, 1995
Vakgroep Algemene Pedagogiek
FSW, RUL
Wassenaarseweg 52, 2300 RB Leiden

Vakgroep Ontwikkelingspsychologie
FSW, UU
Heidelberglaan 2, 3584 CS Utrecht

VOORWOORD

"At this point, the performance variables that determine fluctuation of stage use have only begun to be delineated, and this represents a particularly important direction for further research." (Colby & Kohlberg, 1987, p. 8).

De kwaliteit van het moreel klimaat van een school kan invloed hebben op de ontwikkeling van de leerlingen in het algemeen en hun sociomorele ontwikkeling in het bijzonder. Tot dit laatste rekenen we ook de wijze waarop leerlingen oordelen en zich gedragen in praktische morele situaties, niet alleen in de school, maar ook daarbuiten. Het moreel klimaat van een school (klas, groep) wordt gezien als een intermediaire variabele in de relatie tussen het individueel moreel besef (competentie) en het gedrag van scholieren in situaties in en rond de school, het praktisch moreel denken en handelen (performance). De vraag naar de kwaliteit van het moreel schoolklimaat is niet alleen maatschappelijk actueel, maar is als bepalende factor van morele performance ook theoretisch van belang, getuige het bovenstaande citaat.

Het onderzoeksproject "Moreel schoolklimaat en de morele competentie van scholieren" (Brugman, Høst, Van Roosmalen & Tavecchio, 1994) heeft een schriftelijk afneembaar instrument ontwikkeld voor het vaststellen van de wijze waarop leerlingen van het voortgezet onderwijs het moreel klimaat op hun school waarnemen, zowel bij zichzelf als bij andere leerlingen (Moreel Schoolklimaat Instrument, MSI). De vragenlijst is bedoeld om interventies te evalueren die beogen dit klimaat te verbeteren en aanwijzingen te geven om zulke interventies te implementeren of te optimaliseren. Het instrument is gebaseerd op het conceptuele raamwerk van Kohlberg, Power en Higgins. Tevens wordt door dit project een schriftelijk af te nemen instrument gebruikt voor het vaststellen van het individueel moreel ontwikkelingsniveau van de leerlingen: Sociomoral Reflection Objective Measure-Short Form. In het project werden de beide instrumenten afgenomen op een groot aantal scholen van voortgezet onderwijs. De bedoeling van het project is o.a. om op korte termijn een uitspraak te kunnen doen over de vraag of het moreel klimaat van scholen van voortgezet onderwijs zoals de leerlingen dit waarnemen ook achterblijft bij het individueel moreel oordeelsvermogen van de leerlingen. Het onderzoek wordt gesubsidieerd door de Stichting GedragsWetenschappen (PEDON-NWO).

Het doel van het hier gerapporteerde onderzoek is een uitspraak te kunnen doen over de convergerende en discriminerende validiteit van de instrumenten die beogen moreel klimaat en morele competentie te meten. Er bestaat zowel een schriftelijke als een mondelinge versie van de instrumenten voor de waarneming van het moreel klimaat en

het moreel oordeelsvermogen van de leerlingen. Van de Adviescommissie voor Niet Gebonden Onderzoek (ANGO) van de Stichting voor onderzoek van het Onderwijs werd subsidie verkregen om deze studie uit te voeren.

In dit rapport wordt ook ingegaan op de theoretische veronderstellingen en vragen die bij de empirische vraagstelling in het geding zijn. Enkele van deze veronderstellingen zullen in het kader van het NWO-project nog worden getoetst. De gegevensverzameling vond plaats in het najaar 1994.

Daan Brugman en Louis Tavecchio zijn verantwoordelijk voor de opzet van het onderzoek als respectievelijk projectleider en supervisor. Eerstgenoemde schreef in hoofdzaak hoofdstuk 1 en gedeelten van hoofdstuk 2 en 5. Karin Høst organiseerde de afname van de interviews op de scholen in aansluiting op de schriftelijke instrumenten en de invoer van de gegevens. De afname van deze schriftelijke instrumenten vond plaats onder leiding van stagiaires werkzaam bij het NWO project. Tineke Meijer en Marinka Fintelman namen de interviews af, scoorden de uitspraken op de SRM - SF en voerden de gegevens ervan in. Mechteld van Roosmalen scoorde de uitspraken op de MSI-i. Hun aandeel in de verslaggeving is in hoofdzaak terug te vinden in de beschrijving van de instrumenten in hoofdstuk 2 en hoofdstuk 1. Bartjan van Os nam de gegevensverwerking voor zijn rekening. De hoofdstukken 3, 4 en 5 zijn van zijn hand. Over alle hoofdstukken heeft intensief overleg plaats gevonden.

We bedanken Marga Brugman van het LISWO voor de wijze waarop ze zorg droeg voor de complexe personele ondersteuning die dit onderzoek vereiste. Dankzij Til Ruigrok die de meeste bandopnamen van de interviews uittipte kon de codering van de interviews vlot verlopen. Tenslotte bedanken we de scholen, docenten en leerlingen voor hun bereidwillige deelname aan het onderzoek.

Augustus 1995

INHOUD

1. THEORETISCH KADER EN PROBLEEMSTELLING	6
1.1. Morele competentie, morele performance en moreel klimaat	6
1.1.1. Competence: algemeen of specifiek?.....	8
1.1.2. Onderscheid tussen competence en performance.....	12
1.1.3. Competence - performance en typen dilemma's.....	14
1.1.4. Ontwikkeling van moreel klimaat als interventiestrategie	15
1.2. De stadiumgewijze ontwikkeling van morele competentie, performance en klimaat	17
1.2.1. De ontwikkeling van de morele competentie.....	17
1.2.2. Een Piagetiaanse herziening van Kohlbergs stadium model door John Gibbs.....	21
1.2.3. De ontwikkeling van het moreel klimaat.....	23
1.2.4. De ontwikkeling van de morele performance	29
1.3. Constructen en meetmethoden	30
1.4. Probleemstelling.....	34
2. OPERATIONALISERING EN INSTRUMENTALISERING VAN KERNBEGRIPPEN	36
2.1. Morele competentie: de Sociomorele Reflectie	36
2.1.1. De Sociomoral Reflection Measure - Short Form (SRM-SF).....	37
2.1.2. Sociomoral Reflection Objective Measure - Short Form (SROM-SF)	39
2.2. Moreel klimaat: Zelf en Ander	40
2.2.1. Moreel schoolklimaat instrument-interview-productie (MSI-i).....	41
2.2.2. Moreel schoolklimaat instrument -herkenning (MSI-s).....	43
2.3. Concurrerende constructen	46
2.3.1. Groninger Intelligentie Test	46
2.3.2. Sociale wenselijkheidsschaal	47
2.4. Omschrijving van de onderzoeksgroep en van de selectieprocedure	48
2.5. Afname	49
2.5.1. Afname procedure schriftelijke instrumenten	49
2.5.2. Afnameprocedure mondelinge instrumenten.....	50
3. BESCHRIJVING RESULTATEN PER INSTRUMENT	51
3.1. Inleiding.....	51

3.2. Resultaten Moreel Schoolklimaat Instrument - Interview (MSI-i).....	52
3.2.1. Uitval	52
3.2.2. Descriptie op item niveau.....	55
3.2.3. Schaalgegevens	57
3.3. Moreel Schoolklimaat Instrument - Schriftelijk (MSI-s).....	59
3.3.1. Uitval	59
3.3.2. Descriptie op item niveau.....	60
3.3.3. Schaalgegevens	64
3.4. Sociomoral Reflection Measure - Short Form (SRM-SF)	66
3.4.1. Uitval	66
3.4.2. Descriptie op item-niveau.....	67
3.4.3. Schaalgegevens	68
3.5. Sociomoral Reflection Objective Measure - Short Form (SROM-SF).....	69
3.5.1. Uitval	69
3.5.2. Descriptie op item-niveau.....	70
3.5.3. Schaalgegevens	71
3.6. Groninger Intelligentie Test (GIT).....	72
3.6.1. Uitval	72
3.6.2. Descriptie en schaalgegevens.....	72
3.7. Sociale wenselijkheidsschaal	74
3.7.1. Uitval	74
3.7.2. Descriptie op item-niveau.....	74
3.7.3. Schaalgegevens	75
4. VALIDITEIT VAN DE CONSTRUCTEN.....	77
4.1. Inleiding	77
4.2. Het onderscheid tussen zelfperceptie en perceptie van medeleerlingen.	79
4.3. Kwalitatieve analyse van de convergente en discriminante validiteit	86
4.3.1. Convergente validiteit	86
4.3.2. Discriminante validiteit.....	87
4.4. Kwantitatieve analyse van convergente en discriminante validiteit.....	89
4.5. De relatie tussen Morele competentie, Inhoud en Stadium van de norm.....	93
5. SAMENVATTING EN DISCUSSIE	96

REFERENTIES.....	i
BIJLAGE A. Overzicht van de scoringsregels voor het stadium van de collectieve norm en voorbeeld vragen MSI-i	vi
BIJLAGE B. Voorbeeld van het helpen dilemma uit de MSI-s	ix
BIJLAGE C. Items van het Stadium van gemeenschapsgevoel	xi
BIJLAGE D. Voorbeeld-vragen SRM-sf.	xiii
BIJLAGE E. Voorbeeld-vragen SROM-sf.....	xiv

1. THEORETISCH KADER EN PROBLEEMSTELLING

1.1. Morele competentie, morele performance en moreel klimaat

In alledaagse morele situaties laten mensen zich in hun gedrag ten opzichte van anderen dikwijls sterker leiden door de morele oordelen die gelden in de groep, gemeenschap, institutie of sociale context waarvan ze in deze situatie deel uitmaken dan door oordelen ingegeven door hun eigen moreel besef. In de Amerikaanse vakliteratuur is als dramatisch voorbeeld van dit verschijnsel het My Lai bloedbad beschreven gedurende de oorlog in Vietnam, waarbij soldaten de inwoners van een heel dorp doodden (Kelman & Hamilton, 1989). Het gedrag van deze soldaten kan volgens Higgins, Power & Kohlberg (1984) niet worden begrepen of verklaard uit een onvolgroeid individueel moreel besef, wèl uit de participatie in een groep, de binding met de groep en de normen van deze groep.

Het deel uitmaken van een groep kan -afhankelijk van de normen van die groep en de onderlinge verbondenheid in de groep- in een praktische situatie zowel ondersteuning betekenen om zich conform het eigen oordeel te gedragen, als ondermijning. Deelname in een groep impliceert mede verantwoordelijkheid. Van groepsleden wordt verwacht dat ze de normen van de groep handhaven en verdedigen. Als personen falen zich conform de verwachtingen te gedragen ontvangen ze sancties wat gepaard kan gaan met gevoelens van schaamte of schuld. Bij succes wordt solidariteit ervaren. De groep werkt als een extern drukmiddel.

De binding met de groep is niet het enige kenmerkende van deelname in een groep dat van belang is voor het gedrag. De waarden en normen van de groep zijn in dat opzicht ten minste van even groot belang: "Norms that are widely shared and strongly enforced may be unfair or woefully lacking in their moral significance." (Power, Higgins & Kohlberg, 1989, p. 142). Beide facetten, binding met de groep en de normen en waarden van de groep, zijn opgenomen in ons Moreel Schoolklimaat Instrument (MSI, zie hoofdstuk 2).

De werking van de groep als extern drukmiddel verdient afzonderlijke studie in morele situaties omdat daarbij een individueel moreel besef in het geding is met een karakteristiek ontwikkelingsverloop. Dit besef kan tegenwicht bieden om deel te nemen aan onrechtvaardig handelen in een groepscontext. Morele situaties worden hier opgevat als sociale situaties waarin belangen en perspectieven van verschillende personen of partijen met elkaar in conflict zijn.

Volgens Colby en Kohlberg (1987, p. 8) worden morele oordelen van lagere niveaus dan waarover het individu de beschikking heeft uitsluitend gebruikt in situaties met een beduidende neerwaartse druk. Hiervan is bijvoorbeeld sprake in openbare scholen voor voortgezet onderwijs in de Verenigde Staten: "Ons onderzoek wijst uit dat de culturen van

gewone, grote, openbare scholen voor voortgezet onderwijs een (...) negatief effect hebben op de morele oordelen van leerlingen over problemen die zich in en rond school afspelen." (Power, 1988, p. 202, zie ook Higgins, 1991). Het is voorstelbaar dat ook op veel Nederlandse scholen voor voortgezet onderwijs een neerwaartse druk optreedt.

In vergelijking met leerlingen van de basisschool gaat bij leerlingen van het voortgezet onderwijs de moraal van de groep een belangrijke rol spelen voor hun gedrag. Een voorbeeld is een scholier die iets steelt om zich te conformeren aan de normen van zijn of haar groep. Het moreel klimaat van een school (klas, groep) wordt gezien als een van de intermedierende variabelen in de relatie tussen het individueel morele besef (competentie) en het gedrag van scholieren in situaties in en rond de school, het praktisch moreel denken en handelen (performance).

De begrippen 'competence' en 'performance' verwijzen voor ons naar het onderscheid tussen bekwaamheid en het gebruik ervan (Levine, 1979). Het onderscheid sluit aan bij de algemene overtuiging dat het concrete gedrag van zelf en anderen uitwendige verschijnselen zijn van andere, meer persistente en dieperliggende aspecten van de mens en daar op enigerlei wijze het gevolg van zijn. Het begrip 'competentie' wordt o.a. gebruikt op het gebied van de intelligentie, motivatie, identiteit en gedragsaanpassing. In de psychologie staat op dit moment vooral de beleving van de eigen competentie in de belangstelling als factor die de performance en de mentale gezondheid beïnvloedt (Sternberg & Kolligian, 1990).

We zullen eerst deze begrippen competence en performance toelichten vanuit de optiek van onderzoekers in de Kohlbergiaanse traditie. De belangstelling voor het competence-performance onderscheid komt voort uit de vraag naar de relatie tussen moreel besef en moreel gedrag. In het alledaags taalgebruik wordt deze relatie gewoonlijk opgevat als zouden mensen niet handelen op een wijze die overeenkomt met hun moreel oordeel, mensen doen niet of zelden wat ze vinden dat ze zouden moeten doen. Deze kloof tussen oordeel en gedrag wordt binnen het Kohlbergiaanse raamwerk opnieuw gedefinieerd als het verschil in (a) stadium van moreel oordelen op de klassieke, hypothetische vraagstukken en het stadium van de praktische, context-gebonden morele oordelen, (b) stadium van moreel oordelen op vragen naar wat men zou moeten of behoren te doen en het stadium van uitspraken op vragen naar wat men zou doen, (c) stadium van de uitspraken waarin wordt gezegd wat men zelf vindt te moeten doen of zou doen en het stadium van uitspraken waarin wordt gezegd wat anderen menen te moeten doen of zouden doen. De betekenis van de uitspraak van Kohlberg c.s. over een (negatieve) invloed van het moreel klimaat op het gedrag en de empirische onderbouwing voor die uitspraak zal eerst worden onderzocht.

Niet alleen voor de school, maar ook voor andere instituties (leger, tehuis, gevangenis, bedrijf en kantoor) zijn aanwijzingen gevonden dat oordelen over morele vraagstukken die zich afspelen binnen het instituut worden gekenmerkt door een lager stadium dan oordelen over hypothetische morele vraagstukken, die 'context-vrij' zijn (Kohlberg, Scharf & Hickey, 1974, Rest & Narvèez, 1994, Sprinthall, Stewart, Siemienska & Jasinska-Kania, 1993, Walker, De Vries & Trevethan, 1987, Weber, 1990; vgl. ook Evans, 1986, p. 474, Rest, 1986a, Siegal, 1980). Verder zijn de oordelen die individuen zelf geven in contextuele morele vraagstukken van een hoger stadium dan de oordelen die ze bij andere deelnemers waarnemen (Kohlberg et al., 1974, Higgins et al., 1984). Tenslotte hangen deze individuele contextuele oordelen sterker samen met de oordelen die ze bij anderen waarnemen (het moreel klimaat van dat instituut) dan met de oordelen kenmerkend voor het niveau van het individueel moreel besef. De modus van het gedrag en oordeel in instituties kan werken als selectiesocialisatie: zowel leden met een laag niveau van moreel besef als die met een hoog niveau worden verwijderd (Ponemon, 1992).

1.1.1. Competence: algemeen of specifiek?

In de literatuur worden twee benaderingen onderscheiden in de beschrijving en bepaling van competentie, één welke het als een molair of unitair begrip opvat en één welke meer specifieke kenmerken benadrukt. Deze opvattingen hangen samen met verschillende opvoedingsmodellen: de eerste benadrukt de spontane ontwikkeling van het kind, de tweede de overdracht van kennis en vaardigheden door de volwassene aan het kind.

De morele competentie in de zin van Kohlberg heeft o.i. betrekking op de bekwaamheid om over morele vraagstukken te reflecteren (Pritchard, 1991, p. 44). Als integratief begrip omschrijven Waters en Sroufe (1983, p. 80) competentie als "de bekwaamheid flexibele, aangepaste reacties te geven en te coördineren en het scheppen en benutten van mogelijkheden in de omgeving". Ontwikkeling van competentie betekent dus een toename van flexibiliteit, enz.

Volgens Waters en Sroufe impliceert hun definitie dat het individu zelf kan bijdragen aan de situatie of de mogelijkheid heeft om te reageren, de mogelijkheid of noodzaak om te reageren herkent, beschikt over alternatieven, uit alternatieven kan kiezen, gemotiveerd is om te reageren, in zijn of haar reactie volhardt of deze verandert als de omstandigheden dit vereisen en met zijn of haar reactie kan inspelen op de situatie. Een bezwaar van zulke molaire definities is volgens Waters en Sroufe dat ze geen mogelijkheid bieden om tot meting over te gaan.

Anderzijds bestaan er definities van competenties als specifieke vaardigheden, zoals in de informatieprocesbenadering. Zulke specifieke vaardigheden zijn situatie- en leeftijdspecifiek en daardoor waarschijnlijk van minder belang om de aanpassing van het zich ontwikkelende individu begrijpelijk te maken. Een voorbeeld van deze opvatting is te vinden in de kritiek van Yeates en Selman (1989) op de trainingsprogramma's in de traditie van Spivack en Shure die kinderen vaardigheden leren voor het oplossen van sociale problemen zonder de (inter) persoonlijke aanpassing op de langere termijn in ogenschouw te nemen. Het model dat Yeates en Selman zelf presenteren is een voorbeeld om de tegenstelling tussen de twee benaderingen tot sociale competentie (de eigen oorspronkelijk molaire, structuralistische benadering van Selman en de specifieke, functionalistische benadering van Spivack en Shure) te overbruggen.

Piaget was de unitaire opvatting toegedaan op het intellectuele domein. Deze opvatting lijkt op het intellectuele domein steeds minder aanhang te vinden, althans voor (zeer) jonge kinderen: "Children's competence is not a general structure underlying diverse intellectual activities, as Piaget claimed. Neither does it contain general principles that the child possesses at an early age and that direct the development of competence (...). Rather, it must be understood as a variety fragments of knowledge, connected to specific contexts. Only gradually, as the child is initiated in increasingly more encompassing perspectives on reality, does competence acquire a more general character." (Elbers, 1991a, p. 92, zie ook Byrnes, 1995, p. 4).

Op het domein van sociale ontwikkeling en in het bijzonder van moraliteit heeft Kohlberg (1969, p.349) de molaire opvatting van competentie verdedigd: "Er is een fundamentele eenheid van persoonsorganisatie en ontwikkeling, het ego of het zelf. Hoewel er verschillende gebieden zijn van de persoonsontwikkeling (..) zijn deze gebieden verenigd door hun gemeenschappelijke verwijzing naar *een enkel begrip van zelf in een enkele sociale wereld*. Sociale ontwikkeling is in essentie de herstructurering van het (1) begrip van zelf, (2) in relatie tot begrippen van andere mensen, (3) opgevat als deel uitmakend van een gemeenschappelijke sociale wereld met sociale standaards. Naast de eenheid van het niveau van sociale ontwikkeling dankzij de algemene cognitieve ontwikkeling (de *g* factor in intelligentietests), is er verder eenheid dankzij de gemeenschappelijke factor van ego-ontwikkeling". Dit standpunt verwijst naar het stadium criterium van structureel geheel en wordt in later werk herhaald: "There is a general factor of maturity of moral judgment much like the general factor of intelligence in cognitive tasks." (Kohlberg, 1981, p. 40), "...there is a unitary developmental process involved in the development of both [moral] judgment and action" (Kohlberg & Candee, 1984, p. 506).

Voor het bestaan van die algemene factor wijst Kohlberg o.a. op resultaten van factoranalyse (zie ook Oppenheimer, 1978). Byrnes (1995, p. 5) beschouwt dit als zeer discutabel. Geloofwaardige redenen voor het bestaan van "g" zijn volgens hem: (a) consistent voorspellende waarden van het gedrag van baby's voor de latere ontwikkeling, (b) positieve samenhangen tussen (cognitieve) toetsen die weinig gemeenschappelijk hebben voor wat betreft inhoud of strategieën, (c) hoge correlaties tussen reactietijden op ongelijke toetsen, (d) tweelingstudies die laten zien dat ongeveer 50% van de variantie in IQ scores kan worden toegeschreven aan erfelijkheidskenmerken.

Voor de ontwikkeling van moreel besef zoals hier opgevat (zie par. 1.2) is punt (b) van belang voorzover het om gevarieerde inhoud gaat bij hypothetische, contextvrije morele dilemma's. Bovendien vonden Walker et al. (1987) een grote mate van consistentie in het stadium van oordelen over hypothetische en zelf ervaren dilemma's (Pearson $r = .83$). Echter, Krebs, Vermeulen, Carpendale en Denton (1991, zie ook Krebs & Van Hesteren, 1994) laten daarentegen zien dat situatiespecifieke kenmerken in de dilemma's zijn aan te wijzen en in de condities waaronder de dilemma's worden afgenomen, die variatie in het stadium van oordelen bewerkstelligen. Zowel door Walker et al. als door Krebs et al. werd het Moral Judgment Interview (MJI) gebruikt, dat uit open vragen bestaat. De redenen die respondenten geven op deze vragen worden gecodeerd volgens een uitgebreid handboek (Colby et al., 1987). Als met de schriftelijk af te nemen Defining Issues Test wordt gewerkt, die meerkeuze vragen bevat met naar de stadia gesleutelde antwoorden, is de samenhang met het stadium van oordelen in contextgebonden dilemma's meestal lager; Brugman (1994) vond een hoogste correlatie van .62.

De stadia van moreel oordelen van Kohlberg zijn dieptestructuren die een georganiseerd geheel vormen. Elke structuur biedt een denkkader voor moraliteit. Kohlberg onderscheidt in navolging van Aristoteles weliswaar verdelende rechtvaardigheid (eerlijk delen), procedurele rechtvaardigheid (rechten en plichten) en vergeldende rechtvaardigheid (straffen en herverdelen), maar formuleert geen specifieke stadia voor die gebieden.

Dit in tegenstelling tot onderzoekers als Damon en Turiel die voor een domeinspecifieke benadering hebben gekozen. Deze benadering gaat uit van de veronderstelling dat sociale begrippen gestructureerd zijn binnen verschillende deelgebieden van sociale kennis die kwalitatief verschillende aspecten van de sociale interacties van kinderen kunnen verklaren. Damon heeft stadia beschreven in de ontwikkeling van het denken over verdelende rechtvaardigheid (zie het overzicht van Ter Laak & Aleva, 1994) en Turiel in de ontwikkeling van het denken over sociale conventies (Turiel, 1983, Nucci & Weber, 1991). Niet alleen is er sprake van een opsplitsing in

meerdere inhouden, ook wordt de ontwikkeling niet langer beschreven via 'volle' stadia, maar via substadia (Berkowitz & Keller, 1994), een uitwerking van de gedachte dat een vol stadium verschil in niveau van oordelen minder stimulerend is voor de ontwikkeling dan een kleiner verschil (Berkowitz, 1985). Als men bedenkt dat de ontwikkeling van een vol stadium bij moreel oordelen zo'n 8 jaar in beslag neemt, lijkt zo'n benadering zinvol.

In het overzicht dat Elbers (1991a) biedt van het competentiebegrrip op het cognitieve domein valt het belang van de methode op en de interpretatie van de onderzoekers van de gehanteerde methode. De volgende observatie van Chapman en Chandler (1992, p. 262) is wat dat betreft zeer to the point: "The problem as we see it is that investigators choosing to use different response criteria in assessing a given competence often commit themselves unaware to different conceptual criteria at the same time. Instead of measuring the same competence with different assessment procedures, they often end up measuring competencies which are conceptual distinct."

Op het domein van moraliteit zijn de methoden bijna steevast verbaal van aard. De respondent krijgt dilemma's voorgelegd en vervolgens aan de hand daarvan een serie vragen. Elbers (1991) ziet aan dergelijke verbale methoden belangrijke nadelen verbonden voor het vaststellen van de competentie. Zijn suggestie is methoden aan te wenden waar de produktie van oordelen zich voordoet in spontane interacties.

Kohlberg onderbouwt zijn molaire stellingname met gegevens verzameld met het Moral Judgment Interview. Dit instrument wordt individueel afgenomen bij jeugdige adolescenten en vooral volwassenen (Colby & Kohlberg, 1987, p. 115 e.v.). Rest verdedigt een positie waarin plaats is voor multiple competenties en onderbouwt deze stellingname met gegevens verkregen met behulp van een schriftelijk instrument, dat groepsgewijs afgenomen wordt bij personen in dezelfde leeftijdsbereik als de MJJ. Een schriftelijk instrument dat groepsgewijs afgenomen wordt en speciaal bedoeld is voor jongere personen (vanaf 13-14 jaar), de Sociomoral Reflection Objective Measure, komt nog uitgebreid ter sprake. Damon en Turiel hebben hun onderzoek verricht bij kinderen (4-10 jarigen). Damon (Damon & Killen, 1982) volgt de door Elbers voorgestane werkwijze: kinderen interacteren spontaan over het verdelen van een beloning na het uitvoeren van een taak. Ook bij Power et al. (1989) vinden we spontane interacties, nu echter over morele vraagstukken die zich afspelen op de 'just community' scholen. In dat geval worden de uitspraken van leerlingen als voorbeelden opgevat van de morele schoolcultuur, eventueel van de individuele morele performance, maar niet van de morele competentie van de leerlingen.

Om een moreel oordeel te geven moet een persoon eerst onderkennen dat er sprake is van een morele situatie. Rest (1986) en Kohlberg (Kohlberg & Candee, 1984) hebben een

discussie gevoerd over de relatie tussen morele perceptie en morele competentie. Rest (1983) beschouwt de morele perceptie als het eerste proces van zijn 'vier-componenten model'. De componenten zijn: (a) Morele sensibiliteit, bestaande uit morele perceptie en morele interpretatie van de situatie: wat kan men doen en wat zijn de gevolgen van de verschillende handelingen voor derden? (b) Morele competentie: wat is juist om te doen in dit geval? - het toepassen van morele idealen in de situatie om de handeling te bepalen. (c) Morele motivatie of waardenprioritering: worden morele waarden belangrijker gevonden dan andere waarden, bijvoorbeeld economische? (d) Morele actie: beschikt de persoon over het vermogen en de vaardigheden de intentie uit te voeren? Deze vier processen berusten op een logisch onderscheid; ze vormen niet de vier opeenvolgende fasen in een beslissingsproces: daarvoor worden ook allerlei interacties tussen de componenten gepostuleerd, zoals tussen morele competentie en de interpretatie van de situatie. Echter, wel wordt op basis van dit model gesteld dat om tot een moreel oordeel te komen men eerst moet inzien dat er sprake is van een morele situatie. Proces (a) zou dus logischer wijze vooraf gaan aan proces (b). Kohlberg en Candee (1984) presenteren een model dat compatibel is met het model van Rest, maar zijn de opvatting toegedaan dat de morele competentie de perceptie leidt.

De discussie tussen Rest en Kohlberg lijkt ons in het licht van de hierboven genoemde facetten van competentie tenminste ten dele van conceptuele aard. Vanuit het perspectief van Waters en Sroufe (1983) is er sprake van de uitwerking van enkele facetten van het competentiebegrip. Het model van Rest heeft onderzoek gegenereerd naar hoe zich deze facetten tot elkaar verhouden. Een van de vragen die rijzen bij een uitwerking van het competentiebegrip geven Yeates en Selman aan, namelijk of daarmee zicht wordt gehouden op de ontwikkeling op langere termijn. We komen nog op dit belangrijke punt terug (par. 1.3).

Volgens Chandler (1991) bestaat er in de psychologie geen overeenstemming over hoe de relatie tussen competentie en performance begrepen moet worden, ook niet binnen de structuralistische cognitieve ontwikkelingstheorie van Piaget, Kohlberg, e.a. Voor Piaget heeft competentie betrekking op de algemene cognitieve structuur die ten grondslag ligt aan de reacties die kinderen geven op specifieke taken. De algemene structuur (competence) wordt toegepast om de specifieke taak op te lossen (performance).

1.1.2. Onderscheid tussen competence en performance

Hoewel we geen systematische verhandeling over het onderscheid tussen competence en performance bij Kohlberg hebben teruggevonden, gebruikt hij deze veelvuldig.

Kutnick (1986, p. 118) meent dat Kohlberg zich niet op een consistente wijze over de begrippen *competence* en *performance* en hun onderlinge relatie heeft uitgelaten. Kohlberg zou met *competentie* vooral de reikwijdte van de capaciteiten van een bepaalde persoon bedoelen en met *performance* de actualisering van die *competentie* in een gegeven activiteit. Hoewel Kohlberg volgens Kutnick ook de Chomskiaanse opvatting zou hebben gebruikt, zijn wij op een tekst gestoten waarin hij expliciet deze opvatting heeft verworpen: "Kantian moral intuitionists see their task as that like Chomsky, who attempts to delimit the principles, the deep structure transformations, which define competent syntax in any language." (Kohlberg, 1981, p. 181). Immers, bij Chomsky is er sprake van één grammaticaal systeem dat intuïtief bekend is bij alle vijfjarige kinderen. Uit het onderzoek van Kohlberg blijkt daarentegen dat individuen tot op het conventioneel niveau van moreel oordelen geen notie hebben van de moraliteit van stadium 6. Dus, terwijl Chomsky een Kantiaanse aangeboren structuur voor de syntaxis kan aanvaarden, is er bij Kohlberg sprake van (morele) principes die een actieve reconstructie zijn van de ervaring. Met andere woorden, Kohlberg verbindt zijn molaire opvatting van de morele *competentie* met een interactionistische visie op ontwikkeling: de invloed van de sociale ervaring wordt erkend als constituerend voor de *competentie*.

Verder verwijst *competentie* in de Chomskiaanse benadering naar een geïdealiseerd repertoire van operaties om op een bepaald domein te handelen en de realisering ervan in een bijzonder geval (*performance*). De grammatica is een geïdealiseerde beschrijving van het linguïstisch vermogen; bij de transformationele generatieve grammatica ging het niet om een psychologisch model van de wijze waarop mensen taaluitingen produceren en begrijpen. Ieder psychologisch model van de wijze waarop dit taalvermogen werkt in het concrete taalgebruik (*taalkundige performance*) zal rekening moeten houden met factoren die de *taalkundige* in zijn definitie van het begrip grammatica welbewust buiten beschouwing laat.

Volgens deze redenering zou ook het stadium model van de morele *competentie* van Kohlberg geen psychologische betekenis moeten worden toegekend en worden hier ethische systemen over rechtvaardigheid gegeven die in toenemende mate zijn gedifferentieerd en geïntegreerd. Deze opvatting heeft Kohlberg niet verdedigd (Kohlberg, 1971/1984). Hoewel de belangstelling van Kohlberg evenals die van Piaget uitging naar de ontwikkeling van het moreel oordelen en niet naar de ontwikkeling van het individu, wat ontwikkelingspsychologen als een tekort plegen te beschouwen, is er geen sprake van dat de stadia van het moreel oordelen geen empirisch psychologische beschrijving zouden inhouden of dat de geschetste ontwikkeling slechts voldoet aan logische criteria en geen ontogenetische status zouden hebben.

Competentie verwijst bij Kohlberg naar het meest gevorderde stadium van moreel oordelen waarvan een individu blijkt geeft te beschikken. Preferentie voor de beste oplossing bepaalt mede de consistentie van dit stadium. Competentie verwijst hier naar de externe wereld en impliceert geen monadische opvatting van het individu. In feite geeft het model aan dat de persoon in de loop van de ontwikkeling in toenemende mate minder afhankelijk wordt van de omgeving.

De relatie tussen competence en performance is van het grootste belang voor het begrijpen van de ontwikkeling van de competentie: "A discussion of our position on the appropriateness of the transformational model for the development of spontaneous production in moral judgment requires reference to the distinction between competence and performance." (Colby & Kohlberg, 1987, p. 8). Dit transformationele model van de competentie wordt afgezet tegen een additief model, zoals dat van Selman voor sociale competentie, waarbij de verschillende niveaus van competentie beschikbaar en toegankelijk blijven en een laag niveau voor een adequate oplossing kan zorgen in een bepaald probleem. Het competentie model van Rest (1986) voor morele oordelen is eveneens een additief model. Voor een additief model is variatie in performance veel minder problematisch dan voor een molair model.

Ook voor de relatie tussen moreel oordeel en gedrag is het onderscheid tussen competentie en performance van groot belang. Duidelijk is dat bij Kohlberg geen sprake is van een determinisme, waarbij de competentie eenzijdig de performance bepaalt.

1.1.3. Competence - performance en typen dilemma's

Het onderscheid tussen de klassieke, abstracte, contextvrije, hypothetische dilemma's zoals het Heinz-dilemma en praktische, contextgebonden, levensechte ("real-life") dilemma's sluit aan op het competence-performance onderscheid (Leming, 1976). Het gebruik van hypothetische dilemma's is bedoeld voor een bepaling van de competentie, terwijl dat van praktische dilemma's is bedoeld voor de performance. Nu kan competentie nooit rechtstreeks worden vastgesteld (Kutnick, 1986); er is altijd sprake van een toepassing waarvan de competentie moet worden afgeleid. Het gebruik van hypothetische dilemma's betekent dat de stimuli worden gestandaardiseerd en dat externe factoren die verwarrend kunnen werken op een interpretatie van het redeneren van een subject zoveel mogelijk worden verwijderd; de complicerende omstandigheden van het leven van alledag die van invloed zijn op het morele oordeel worden zoveel mogelijk uitgesloten. Er wordt uitsluitend gevraagd naar het moreel oordeel vanuit het gezichtspunt van de protagonist in het dilemma, zoals in het dilemma van Heinz: "Wat zou Heinz volgens jou moeten doen?"

Verondersteld wordt dat hypothetische dilemma's de qua stadium hoogste prescriptieve denkwijze van een persoon blootleggen en daarmee ook het meest geschikt zijn voor het bepalen van de competentie. In vele variaties herhaalt Kohlberg dit thema (zie voor een kritiek op de daarop afgestemde scoringsregels Eisenberg, 1989, p.97). We geven enkele citaten: "We have attempted to minimize the gap between the competence and performance by using hypothetical dilemmas, by using probing questions that attempt to elicit the upper limit of the subject's thinking, and by our scoring rules according to which only the most mature expressed version of a particular moral idea is scored." (Colby & Kohlberg, 1987, p. 5). "(...) Kohlberg's current assessment instrument and interview method taps a subject's competence rather than his or her performance by providing probing questions that attempt to elicit the upper limits of a subject's thinking" (Kohlberg, Levine, & Hewer, 1983, p. 37). En: "(...) we do not assume that people always use their highest stage of moral reasoning. However, the Standard Moral Judgment Interview does attempt to tap the most advanced reasoning of which a subject is capable." (Colby & Kohlberg, 1987, p. 8). Uitspraken (Kohlberg et al., 1983, p. 152) als zou gepoogd worden om de 'optimale competentie' vast te stellen lijken ons een pleonasme.

In het algemeen blijken hypothetische dilemma's bij volwassenen het hoogste niveau te ontlocken (Walker et al., 1987), maar bij kinderen ligt dit mogelijk anders. De resultaten van de onderzoeken van Norma Haan (1975) en Gilligan en Belenky (1980) laten overheersend een omgekeerde tendens zien: hogere scores op de echte dilemma's.

Bij praktische, contextgebonden en levensechte dilemma's wordt behalve naar de prescriptieve redenering ook gevraagd naar wat de persoon zelf in de onderhavige situatie doet of zou doen en in hoeverre men zich verantwoordelijk beschouwt; nagegaan wordt of het gedrag prosociaal van aard is. "Werkelijke morele beslissingen zijn niet alleen voorschrijvend, ze vragen ook een beschrijving van de feitelijke situatie en van de behoeften en motieven van zelf en anderen geïntegreerd in een weloverwogen praktisch oordeel." (Higgins et al., 1984, p. 77).

1.1.4. Ontwikkeling van moreel klimaat als interventiestrategie

Aan de school wordt traditioneel een belangrijke rol toegeschreven in de ontwikkeling van het moreel besef en gedrag van de leerlingen (Kohlberg, 1981). Afname van de conventionele moraliteit in onze samenleving en in samenhang daarmee een verzwakking van de morele autoriteit zou verantwoordelijk zijn voor een negatief effect van het schoolklimaat.

De effecten van groepsdruk op het gedrag van individuen hebben Kohlberg en zijn medewerkers gesteund in de opvatting dat voor morele opvoeding het verbeteren van het

moreel klimaat waarin beslissingen worden genomen in veel gevallen de beste strategie is (Kohlberg, 1970, 1981, Higgins et al., 1984, Power et al., 1989). De school moet leerlingen een cultuur bieden waarin zij conform hun stadium van moreel oordelen kunnen handelen en moet er zorg voor dragen dat leerlingen het conventionele stadium voorspoedig doorlopen. Dus, in plaats van interventies te richten op het stimuleren van de ontwikkeling van het individueel moreel besef, waardoor de leerling beter zou leren onderkennen welk beroep zijn groepsgenoten op hem of haar doen, heeft Kohlberg er voor gekozen om het moreel klimaat van de groep te ontwikkelen.

De keuze voor de ontwikkeling van de groeps Moraal betekent dat de school zich niet richt op de bevordering van de morele 'competentie' van de leerlingen, maar juist op de 'morele performance' (Pollard, Kurtines, Carlo, Dancs, & Mayock, 1991). Die opvatting sluit aan bij de belangstelling van opvoeders voor het tot stand komen van pro sociaal gedrag als opvoedingsdoel.

Het bereiken van het conventionele niveau (stadium 3) beschouwen Power et al. (1989, p. 29) als een minimale basis voor morele competentie van achttienjarigen. De ontwikkeling van stadium 2 naar stadium 3 houdt de omslag in van een premoreel gezichtspunt waarin rechtvaardigheid is gebaseerd op concrete wederkerigheid, een economisch evenwicht, naar een zuiver moreel gezichtspunt waarin rechtvaardigheid is gebaseerd op de gulden regel en de sociale relatie als een doel op zich wordt beschouwd. Het conventionele niveau staat voor een internalisering van morele waarden. Als leerlingen zich onvoldoende morele waarden eigen maken zal dit ook consequenties hebben voor het gedrag buiten school.

Deze klimaatontwikkeling vindt plaats door middel van een radicaal democratiseringsproces. Scholieren worden in onze samenleving niet als moreel incompetent beschouwd. Power et al. (1989, p. 30), trekken hieruit de conclusie dat scholieren geen serieuze gebreken hebben op grond waarvan men hen niet capabel zou achten voor rationele democratische participatie. In de 'just community' scholen discussiëren en onderhandelen leerlingen over de regels op school en leren zij zelf mede verantwoordelijkheid te dragen voor de naleving van deze regels. De leerkrachten zijn deelnemer in dit beslissingsproces.

Het onderscheid tussen normale scholen en 'just community' scholen komt tot uitdrukking op de variabelen die zijn opgenomen in Kader 1 en die in dit onderzoek van belang zijn (Higgins et al., 1984). De scores op schoolgebonden morele vraagstukken voor zelf worden daarbij opgevat als operationalisatie van de morele performance, terwijl de scores voor ander opgevat worden als operationalisatie van moreel klimaat (zie par. 1.2.4). Op alle categorieën van zelf/ander scoren de just community scholen hoger dan de normale scholen. Andere, voor dit onderzoek interessante gegevens zijn: (a) er is een

discrepantie zelf - ander, waarbij zelf consistent hoger scoort dan ander; (b) de discrepantie zelf-ander varieert bij keuze van -1% (CS) tot 41% (BH) en bij actie van 10% (SWS) tot 35% (BH); (c) de verschillen tussen de 2 normale scholen zijn op keuze bijna 4 maal groter dan deze verschillen tussen de 2 community scholen. Er lijkt in dit opzicht geen reden het klimaat van alle normale scholen over een kam te scheren (Brugman et al., 1994).

Kader 1: Scores van de leerlingen van 2 normale en 2 just community scholen op praktische morele vraagstukken in de context van de school: morele performance (zelf) en moreel klimaat (ander).

	normaal		just comm.	
	BH	CH	SWS	CS
prosociaal:	N=10	N=15	N=16	N=20
keuze zelf	75%	66%	86%	83%
gedrag zelf	61%	63%	77%	74%
keuze ander	34%	45%	81%	84%
gedrag ander	26%	34%	67%	58%

1.2. De stadiumgewijze ontwikkeling van morele competentie, performance en klimaat

1.2.1. De ontwikkeling van de morele competentie

Kohlberg is bekend geworden door zijn theorie over de ontwikkeling van het moreel oordelen bij individuen, in de zin van morele competentie. Morele oordelen worden uitgelokt door aanbieding van een dilemma. In zo'n dilemma is sprake van een conflict tussen aanspraken, rechten van twee of meer personen of partijen. In het moreel oordeel wordt aangegeven welke van deze aanspraken het meest gerechtvaardigd is en wat men verplicht is te doen om deze rechten te verzekeren. Moreel oordelen vraagt dat men zich kan verplaatsen in de claims van de betrokken partijen.

Kohlberg heeft de ontwikkeling van het moreel oordelen beschreven in 6 stadia (zie Kader 2). De stadia vertegenwoordigen een toenemende adequaatheid van moraliteit. Hoe hoger, hoe moreel juist. In de loop van hun ontwikkeling gaan mensen steeds 'moreler' denken. Het eindstadium wordt gekenmerkt door universele ethische principes.

Rechtvaardigheid (justice) wordt opgevat als de kern van moraliteit. Rechtvaardigheid is een moreel principe en geen concrete regel of handeling zoals bij eerlijkheid. (Eerlijkheid betekent: bedrieg niet, steel niet en lieg niet.) Er zijn uitzonderingen op regels, maar niet op principes. Als principe en reden om te handelen is rechtvaardigheid respect voor mensen.

Kader 2: De stadia van de individuele morele competentie van moreel oordelen

Pré-conventioneel niveau

Mensen op dit niveau (voornamelijk kinderen) beschouwen regels en sociale verwachtingen als extern, als iets buiten hen zelf.

Stadium 1: Heteronome moraliteit

Het sociaal-morele perspectief op dit stadium is een perspectief van naïef moreel realisme. Morele oordelen worden niet ter discussie gesteld, ze zijn vanzelfsprekend. Normen zijn concrete regels, waar geen uitzonderingen op mogelijk zijn. Er is geen oog voor de bedoelingen achter het handelen. De autoriteit bepaalt wat goed en slecht is. Het perspectief is egocentrisch. Men gaat er van uit dat er maar één perspectief is wat door iedereen gedeeld wordt. Er is geen oog voor conflicterende belangen. Men moet moreel handelen om straf te vermijden. Men moet de autoriteit gehoorzamen. Voorbeelden: "Vader is de baas, want hij is groter", "Het redden van het leven van meer mensen is belangrijker dan het leven van een persoon, omdat een persoon maar een huis heeft, misschien wel veel meubels, maar een aantal mensen hebben samen een enorme hoeveelheid meubels", "als je je niet aan je belofte houdt, dan zal die persoon je in elkaar slaan". Kenmerken van de situatie die indruk maken op de waarneming (grootte, objecten of handelingen) wekken de aandacht van het kind en hebben invloed op z'n verbeelding.

Stadium 2: Individualistische instrumentele moraliteit

Op dit stadium is er sprake van een pragmatisch perspectief. Moreel handelen is zo handelen dat de eigen belangen gemaximaliseerd worden en de negatieve consequenties geminimaliseerd. Daarvoor is een instrumentele ruif noodzakelijk: 'Ik doe wat voor jou, opdat jij wat voor mij doet'. Men ziet in dat iedereen eigen belangen heeft en dat belangen kunnen conflicteren. Het perspectief is concreet individualistisch, er is nog geen sprake van een derde persoonsperspectief.

Conventioneel niveau

Mensen op dit niveau (voornamelijk oudere adolescenten en volwassenen) identificeren zich met, of hebben geïnternaliseerd, de regels en sociale verwachtingen van anderen, inclusief die van de autoriteiten.

Stadium 3: Wederzijdse interpersoonlijke verwachtingen, relaties en interpersoonlijk conformisme

Op dit niveau is er een derde persoonsperspectief, er zijn relaties tussen mensen die bepaald worden door wederzijds vertrouwen. Er zijn gedeelde morele normen volgens welke mensen geacht worden te leven. Op dit stadium heeft men oog voor de motieven achter het handelen. Men moet moreel handelen om het wederzijds vertrouwen te behouden en om sociale goedkeuring te krijgen. Er zijn stereotiepe rolverwachtingen: een goede zoon zijn, een goede vriend zijn en dergelijke.

Stadium 4: Sociaal systeem en geweten

Het perspectief op dit stadium is het perspectief van het systeem, bijvoorbeeld als lid van een maatschappij of als lid van een kerkgenootschap. Individuele belangen volgen is alleen acceptabel als ze niet in strijd zijn met het belang van de samenleving als geheel. Morele oordelen verwijzen naar instituties of systemen, die zowel legaal, sociaal of religieus kunnen zijn. Dat wat een bijdrage aan de samenleving als geheel levert wordt als rechtvaardig beschouwd. Burgerlijke ongehoorzaamheid is echter ook op dit niveau wel mogelijk. De wetten van de maatschappij kunnen in strijd zijn met de religieuze wetten die voor het individu van een hogere orde kunnen zijn. Dit kan tot gewetensconflicten leiden. Zo kan iemand op stadium 4 bijvoorbeeld dienstweigeraar zijn door te verwijzen naar zijn religieuze overtuiging als Jehova getuige.

Post-conventioneel niveau

Mensen op dit niveau (een kleine minderheid van volwassenen) differentiëren zichzelf van de regels en verwachtingen van anderen en denken in termen van zelf gekozen principes.

Stadium 5: Sociaal contract, utiliteit en individuele rechten

Op dit stadium wordt niet geredeneerd vanuit de bestaande maatschappij maar wordt de bestaande maatschappij getoetst aan de regels waarop de totstandkoming van de maatschappij gebaseerd is. Er zijn fundamentele rechten en waarden die beschermd moeten worden door de maatschappij. Er is een verplichting om morele keuzes te maken die in strijd kunnen zijn met de bestaande wetten of normen. Er is oog voor de rechten van minderheden. Op dit stadium is er nog geen volledig evenwicht gevonden. Men heeft nog moeite met het integreren van de verschillende perspectieven.

Stadium 6: Universele ethische principes

Op dit stadium gaat het om 'the moral point of view'. Universele ethische principes worden gehanteerd om morele conflicten op te lossen. Door zich te verplaatsen in alle deelnemers die betrokken zijn bij een conflict en afwisselend in gedachte de rol van elke deelnemer in te nemen kan men komen tot een oplossing waarmee elk rationeel persoon kan instemmen. Kohlberg noemt dit 'ideal role taking' of 'moral musical chairs'. Op zich is het sociaal-moreel perspectief op dit stadium niet anders dan op stadium 5. Het verschil is echter dat er expliciet gebruik gemaakt wordt van rechtvaardigheidsprincipes.

Rechtvaardigheid is de enige morele deugd die de relaties tussen personen in een maatschappij regelt. Behandel de claim van iedere persoon onpartijdig, ongeacht de persoon.

Het expliciet hanteren van universele, voorschrijvende principes om conflicterende claims op te lossen wordt als het hoogste ontwikkelingsstadium opgevat. Op het hoogste niveau is een moreel oordeel universeel, dat wil zeggen er is een morele verplichting voor iedereen om in deze situatie zo te handelen. Een moreel oordeel is prescriptief, een verplichting om te handelen. Het gaat niet om kunnen maar om moeten handelen.

De sterkste claims voor stadia in de sociaal-cognitieve ontwikkeling zijn naar voren gebracht door Kohlberg. Volgens het zogenaamde "harde stadium" model moeten de stadia voldoen aan vier criteria (zie Boom, 1992 voor een uitgebreide discussie over het stadiumbegrip bij Kohlberg):

- (a) De stadia geven kwalitatieve verschillen in het denken weer over morele vraagstukken.
- (b) Een stadium vormt een structureel geheel, er ligt aan de reacties van het subject een bepaalde organisatie van het denken ten grondslag. De reacties zijn dus betrekkelijk onafhankelijk van de situatie. "Insofar as we are successful in eliciting the respondent's highest level, we expect that the thinking exhibited will be quite internally consistent, and with the exception of some very slight decalage, will form a structural whole across widely varying content." (Colby & Kohlberg, 1987, p. 8).
- (c) De stadia zijn hiërarchisch geïntegreerd. De structuren vormen een ordening op grond van een toenemende differentiatie en integratie. Op de hogere stadia wordt het denken van de voorafgaande stadia gereconstrueerd. Hogere stadia integreren en vervangen de patronen van denken die behoren bij de lagere stadia.
- (d) In de ontwikkeling worden deze stadia in een onomkeerbare volgorde doorlopen. Deze ontwikkeling kan worden versneld, vertraagd of stopgezet, maar de volgorde van de stadia blijft ongewijzigd, er kan geen stadium worden overgeslagen. Ook terugval is niet mogelijk.

De meeste cognitieve ontwikkelingstheoretici volgen Kohlberg niet in deze criteria. Damon, Eisenberg, Hoffman, Kegan, Krebs, Loevinger en Rest hebben 'zachte' modellen voor sociale ontwikkeling voorgesteld. Deze theoretici menen dat ontwikkeling weliswaar gekenmerkt wordt door toenemende complexere wijzen van denken, maar de stadia worden minder exclusief gedefinieerd in onderliggende cognitieve operaties of structuren. Ook cognitieve inhouden, affectieve oriëntaties en gedragsstijlen kunnen deel uitmaken van de stadiumomschrijvingen. Ze geven prototypische, aan leeftijd gebonden cognitieve

oriëntaties in plaats van systemen die het hoogste niveau van de structurele competentie weergeven. Zulke theoretici veronderstellen dat de oude stadiumstructuren behouden kunnen blijven nadat nieuwe zijn verkregen. Zij verwachten dus minder structurele consistentie dan harde stadia theoretici over verschillende inhouden. Hoewel zachte stadia theoretici veronderstellen dat mensen normaliter de stadia doorlopen zoals ze deze hebben gespecificeerd, veronderstellen ze minder stringent dat deze volgorde invariant is (Krebs & Van Hesteren, 1994). Het begrip van "zacht stadium" vinden we eveneens door Kohlberg toegepast op de morele performance en het moreel klimaat.

Vanaf stadium 3 onderscheidt Kohlberg twee substadia, type A en type B. De conventionele stadia 3 en 4 in het oorspronkelijke scoringssysteem van Kohlberg vormen het A type. Mensen die oordelen volgens type A redeneren vanuit een heteronoom respect voor regels en autoriteit. Het type B gaat verder dan het type A omdat hier de voorschrijvende oordelen in geïdealiseerde en gegeneraliseerde vorm voorkomen. De oordelen klinken principieel. Type B wordt ook als een gewetensoriëntatie aangeduid. Mensen die oordelen volgens type B, redeneren vanuit een oriëntatie op autonomie en rechtvaardigheid. Respondenten van het type B zullen in hun oordelen meer prescriptief en universeel zijn dan die van het type A. Zij zullen dus meer nadruk leggen op de plicht voor iedereen om in soortgelijke omstandigheden zo te handelen. Uit de antwoorden van type B respondenten blijkt dat zij er meer op gericht zijn om de belangen van alle deelnemers aan het morele conflict in evenwicht te brengen, hun keuze zal rechtvaardiger zijn. Verder hebben ze meer oog voor fundamentele waarden, zoals de waarde van het leven. Stadium 3B wordt gekenmerkt door universele interpersoonlijke idealen en stadium 4B door universele maatschappelijke idealen. Volgens Kohlberg behoren ze tot het conventionele niveau omdat ze alleen impliciet of intuïtief prioriteit geven aan morele waarden boven legale en gebruikelijke waarden.

Het expliciet hanteren van universele, voorschrijvende principes om conflicterende claims op te lossen wordt als het hoogste ontwikkelingsstadium opgevat. Niet ieder mens bereikt dit stadium. Integendeel, in westerse culturen bereikten slechts 13% van de volwassenen het vijfde stadium, waarbij stadium 5 als een mixture met lagere stadia is inbegrepen. Deze personen hebben gewoonlijk een academische opleiding gevolgd.

Omdat niet kon worden aangetoond dat een enigszins substantieel percentage mensen volledig stadium 6 bereikte, heeft Kohlberg het zesde stadium als een empirisch stadium laten vallen (Colby & Kohlberg, 1987). Theoretisch blijft dit stadium van belang als eindpunt van ontwikkeling (echter, zie Puka, 1991, voor een andere opvatting), wat niet wegneemt dat ook op dit stadium nog onoplosbare problemen voorkomen. Dit is voor

Kohlberg aanleiding is geweest voor speculaties omtrent een zacht 'zevende' stadium. De voorgestelde oplossing wordt echter onbevredigend gevonden (zie o.a. Krebs & Van Hesteren, 1994, p. 126, Puka, 1991).

De competentie wordt 'gereconstrueerd' op grond van de redenen die individuen geven als zij abstracte, hypothetische morele dilemma's krijgen voorgelegd en worden gevraagd hoe de protagonist zou moeten handelen en waarom (prescriptief, deontisch oordeel). Door gebruik te maken van een gevarieerde set van hypothetische morele dilemma's waarin de drie aspecten van rechtvaardigheid aan de orde komen wordt een representatief beeld verkregen van de competentie. Dezelfde wijze van redeneren kan worden opgewekt door een andere set hypothetische dilemma's waarin overeenkomstige problemen aan de orde worden gesteld. De variatie aan hypothetische situaties in een interview of een ander instrument, is de beste garantie dat er een maat wordt verkregen voor de competentie, los van de persoonlijkheids- en situatiekenmerken die het individueel moreel oordelen in een specifieke situatie kunnen beïnvloeden (Higgins et al., 1984).

1.2.2. Een Piagetiaanse herziening van Kohlbergs stadium model door John Gibbs

Gibbs heeft intensief samengewerkt met Kohlberg aan het onderzoek van moreel oordelen. Hij (Gibbs, 1979, 1991, Gibbs, Basinger & Fuller, 1992) volgt in grote lijnen Kohlberg, maar gebruikt een andere terminologie dan Kohlberg en wijst met een beroep op Piaget de indeling van Kohlberg in een pré-conventioneel, conventioneel en post-conventioneel niveau af.

Gibbs spreekt niet van moreel oordelen, maar van sociomorele reflectie en rechtvaardiging. Niet van morele ontwikkeling, maar van sociomorele ontwikkeling omdat daarmee beter wordt aangegeven dat het bij de bestudeerde morele redeneringen gaat om relaties en transacties tussen mensen, en moraliteit dus alles te maken heeft met sociale interactie. De term "moreel" heeft misleidende solipsistische en privé connotaties (Gibbs et al., 1992, p. 20).

Onder sociomorele rechtvaardigen verstaat hij de redenen die een persoon geeft voor beslissingen of waarden die betrekking hebben op eerlijk gedrag en op gedrag dat beoogt het welzijn van anderen te bevorderen. Voorbeelden van zulke gedragingen zijn: het houden van een belofte, de waarheid vertellen, een vriend helpen, niet stelen, een leven redden.

Gibbs onderschrijft het idee van hiërarchische integratie. Natuurlijke morele ontwikkeling houdt ook voor hem een verandering in naar toenemende morele juistheid. De ontwikkeling van sociomoreel oordelen verloopt volgens hem van een oppervlakkig (onvolwassen) naar een diepgaand (volwassen) niveau van oordelen. Hij onderscheidt in

deze ontwikkeling vier stadia, die in veel opzichten identiek zijn aan Kohlbergs beschrijving van deze stadia. Het moreel redeneren op het onvolwassen niveau is oppervlakkig. Moraliteit wordt verward met fysieke kracht (stadium 1) of pragmatische transacties (stadium 2). Het moreel redeneren op het volwassen niveau wordt gekenmerkt door een interne morele oriëntatie. Er is een volwassen begrip van normen en waarden. Tussen de twee niveaus vindt een overgang plaats van instrumentele naar ideale wederkerigheid. Dit weerspiegelt een constructief proces van decentrerend, d.i. het schenken van aandacht aan en het met elkaar in verband brengen van veelzijdige kenmerken van situaties. Op stadium 3 wordt wederkerigheid begrepen als de grondslag voor interpersoonlijke relaties en op stadium 4 als grondslag voor de maatschappij. Op de stadia 3 en 4 onderscheidt Gibbs, evenals Kohlberg, twee substadia, morele typen of oriëntaties, A en B. Subjecten die gekenmerkt worden door het B-type kunnen vragen opwerpen over het morele en daarmee een meta-ethisch of existentieel gezichtspunt innemen.

Stadium 3 moraliteit geeft geen volledige morele cognitieve juistheid of volwassenheid voor individuen die in een complexere samenleving leven dan een kleine gemeenschap. Individuen in een complexe sociale wereld moeten in het algemeen begrijpen dat ze onderling van elkaar afhankelijk zijn. De omgang met anonieme individuen, het onderhouden van relaties met mensen die verschillende waarden er op nahouden en de behoefte aan een wederzijds vertrouwen vraagt een waardering voor algemeen aanvaarde, consistente standaards en eisen. Deze bevorderen samenwerking en werken als regels die ontworpen zijn om misverstand en wanorde te vermijden. Mensen moeten een balans aanvaarden tussen hun rechten of vrijheden enerzijds en hun verantwoordelijkheid om de rechten van anderen te respecteren als ook een bijdrage te leveren aan de samenleving anderzijds.

Gibbs verwerpt Kohlbergs onderscheid tussen de conventionele en postconventionele moraliteit en daarmee ook de claim dat alleen het postconventionele niveau zou getuigen van morele volwassenheid. Het post-conventionele niveau komt volgens hem zo weinig voor dat het toepassen van stadium 5 en 6 inconsistent is met de stadiatheorie van Piaget waarin gesteld wordt dat een groot gedeelte van de volwassen bevolking het hoogste stadium moet kunnen bereiken. Ook moet het hoogste stadium volgens Piaget relatief spontaan bereikt kunnen worden zonder te worden beïnvloed door culturele ideologie, d.w.z. een academische opleiding en een bepaalde intellectuele scholing. Tenslotte kan volgens Gibbs een subject als het eenmaal in staat is te reflecteren op regels en wetten elke wijze van redeneren innemen die het prefereert (zie ook Siegal, 1980, p. 290). We hebben gezien dat Kohlberg deze gedachtengang verwerpt als zijnde in strijd met de

opvatting van hiërarchische integratie en dat alleen neerwaartse druk een subject ertoe kan brengen redenen van een lager niveau aan te wenden.

Gibbs (Gibbs et al., 1992) meent dat in het model van Kohlberg het volwassen moreel oordelen wordt verward met het kunnen gebruiken van filosofische argumenten. Die respondenten die op niveau 3B en 4B oordelen, oordelen volgens Kohlberg nog steeds conventioneel omdat zij slechts impliciet of intuïtief de voorkeur geven aan morele waarden boven wetten en gewoontes. Respondenten op niveau 5 kunnen dit beredeneren. Gibbs ziet het theoretisch belang er niet van in om het expliciet gebruiken van een ethische filosofie als een hoger natuurlijk ontwikkelingsniveau te beschouwen. Bovendien, zo vraagt hij zich af, hoe kan het oordelen van iemand die de morele waarden van bijvoorbeeld leven boven de gehoorzaamheid aan wetten en autoriteit stelt, zoals type B respondenten doen, conventioneel genoemd worden? (Gibbs et al., 1992, p. 15). Voor Gibbs is stadium 5 een meta-ethische positie die volwassenen ontwikkelen in de loop van hun leven door reflectie op existentiële vraagstukken. Gibbs (1991, p. 198, 199) beschouwt de filosofisch geschoolde reflectie als positief voor de individuele zingeving. Overigens merken Basinger en Gibbs (1987, p. 145) op dat het ontbreken van stadium 5 rechtvaardigingen in hun objectieve instrumenten (zie onder) deze ongeschikt maken voor bepaalde groepen volwassenen.

Kohlberg et al. (1983, p. 152-156) accepteren Gibbs' onderscheid tussen de structurele stadia enerzijds en de meta-ethische posities anderzijds. Zij wijzen de opvatting van Gibbs van de hand als zou stadium 5 niet anders zijn dan een meta-ethische opvatting. Daarbij beroepen zij zich op crossculturele evidentie voor de aanwezigheid van stadium 5 en op het feit dat stadium 5 beter moreel handelen voorspelt dan enig ander lager stadium. De (existentiële) reflectie waarop Gibbs wijst zien zij vooral optreden na stadium 5.

1.2.3. De ontwikkeling van het moreel klimaat

Het door de leerlingen waargenomen moreel schoolklimaat heeft betrekking op de morele cultuur dimensie van het klimaat van een institutie (Andersen, 1982). Gemakshalve spreken we van 'moreel schoolklimaat'. Het moreel schoolklimaat heeft betrekking op de morele waarden en normen van de leerlingen van een school en de mate waarin deze door de leerlingen worden gedeeld. Het wordt in kaart gebracht aan de hand van de waarneming van de leerlingen over morele kwesties die zich afspelen op school. Kenmerkend voor de hier gevolgde opvatting is dat de inschatting door de individuele leerling van de sociale werkelijkheid van de school voorop staat. De leerling wordt gevraagd op te treden als een waarnemer en representant van 'de leerlingen van deze school'. Het morele klimaat van een school is dus niet gelijk aan het gemiddelde van de

waarnemingen door de individuele leerlingen, maar verwijst naar het gedeelde, gemeenschappelijk perspectief van de leerlingen. De reikwijdte van dat gemeenschappelijke perspectief varieert in de praktijk: over sommige zaken menen leerlingen slechts namens zichzelf te kunnen spreken, over andere namens de vrienden of vriendinnen, of klasgenoten. De gevallen waarin leerlingen namens de leerlingen van een school kunnen spreken zijn op gewone scholen schaars. Op de dimensies van het moreel klimaat, zoals de mate waarin de norm wordt gedeeld, komt deze reikwijdte tot uitdrukking.

Moreel klimaat is een multidimensioneel begrip. Power et al. (1989) onderscheiden zes dimensies in moreel klimaat die elk op hun beurt weer nader zijn onderverdeeld in categorieën. Deze dimensies en categorieën zijn geconstrueerd op basis van het proces dat scholen doorliepen in de ontwikkeling tot een morele gemeenschap ('just community'). 'Gemeenschap' verwijst naar een ideaaltype van de relaties tussen personen in moderne instituties, waarbij de relaties als doel op zich worden gewaardeerd en niet als instrumenteel worden beschouwd. Het moreel schoolklimaat wordt gemeten door leerlingen morele dilemma's voor te leggen of naar dilemma's te vragen die zijn gesitueerd in de school en een inschatting te vragen van de gemeenschappelijke gedragingen, motieven en redenen van leerlingen daarbij. Door aggregatie wordt vervolgens de waarneming van het moreel klimaat van een school beschreven.

Het moreel oordeel is een discreet moreel argument dat een norm en een element bevat. De norm verwijst naar een complex geheel van specifieke gedragsverwachtingen, die een gemeenschappelijke, institutionele waarde delen. Een norm als vertrouwen kan bijvoorbeeld uit de volgende verwachtingen bestaan: je kan je spullen onbeheerd op tafel laten liggen, je deelt je spullen met anderen. Aan deze verwachtingen op zich ontbreekt een belangrijk motivationeel aspect of waarde. Vertrouwen is zo'n aspect; vertrouwen omvat een verwantschap met anderen dat tot uitdrukking komt in de verwachtingen maar tegelijkertijd ook verder gaat. Het element vormt de uiteindelijke waarde die de handhaving van de norm rechtvaardigt. Voorbeelden van zulke uiteindelijke waarden zijn groepsharmonie en individueel welzijn.

Twee dimensies van het moreel schoolklimaat die isomorf zijn aan de stadia van de individuele morele competentie worden in dit onderzoek betrokken: "Het stadium van de norm" en "Het stadium van gemeenschapsgevoel". Deze twee dimensies zijn gerelateerd aan de elementen. Vier andere dimensies zijn gerelateerd aan karakteristieken van de norm: de mate van collectiviteit, de fase, het stadium en de inhoud. Al deze dimensies worden in ons moreel schoolklimaat instrumenteel geoperationaliseerd, maar niet alle zijn voor dit onderzoek van belang.

Het stadium van de collectieve norm heeft betrekking op overeengekomen manieren waarop vraagstukken met tegenstrijdige belangen, die de groepsleden of de groep als geheel aangaan, worden opgelost. Door Power et al. (1989, p. 115-127) worden drie stadia beschreven in de ontwikkeling van collectieve normen (zie Kader 3).

De vraag of leerlingen die onderling van elkaar verschillen in moreel redeneren op stadium 2, 3 of 4, een gedeelde perceptie van de collectieve normen en waarden van de school kunnen hebben, beantwoorden Power et al. bevestigend. Echter, de stadia van collectieve normen zijn niet gelijk aan Kohlbergs stadia van individueel moreel redeneren. Aan de collectieve stadia ('zachte' stadia) ligt een zwakkere toepassing van de criteria van stadium structuur ten grondslag dan bij de stadia van individueel moreel oordelen ('harde' stadia). Power et al. (1989, p. 136-142) noemen drie verschillen, die in het navolgende worden samengevat.

Ten eerste is bij de stadiumgewijze ontwikkeling van collectieve normen geen sprake van een onomkeerbare volgorde van in toenemende mate gedifferentieerde en geïntegreerde structuren. Power et al. geloven niet in een groepsbewustzijn: alleen individuen denken. Individuen ontwikkelen -als zij handelen als lid van een groep- gemeenschappelijke normen, die op hun beurt hun denken in de groep beïnvloedt. De constructie van dergelijke gemeenschappelijke normen weerspiegelt de performance van het moreel redeneren, niet de competentie. Collectieve stadia worden niet afgeleid uit het moreel redeneren van individuele groepsleden, maar uit hun onderlinge interacties of prestaties in een groepscontext. Collectieve stadia worden dus begrepen als collectieve 'performances'. Twee aannames van de claim van onomkeerbare volgorde worden daartoe aangepast: (a) Collectieve normen hoeven niet op stadium 0 of 1 te beginnen. Het beginpunt van de ontwikkeling van het collectief stadium zal gedeeltelijk afhankelijk zijn van de stadia van de individuele groepsleden. Deze stadia bepalen de bovengrens maar niet de ondergrens van het collectieve stadium. (b) Er kan regressie optreden in het collectieve stadium omdat de performance kan fluctueren en omdat er elk jaar nieuwe jongere leerlingen op school komen.

Bij de individuele stadia gaat het om structuren die verwijzen naar georganiseerde wijzen van denken over morele vraagstukken. Bij collectieve stadia gaat het om structuren die verwijzen naar overeengekomen manieren waarop vraagstukken met conflicterende belangen, die groepsleden of de groep als geheel aangaan, worden opgelost. Terwijl de stadia qua definitie van elkaar verschillen, vertonen ze inhoudelijk echter een grote overeenkomst; vergelijk de omschrijvingen in Kader 3 met die in Kader 2.

Kader 3: Stadia van de collectieve norm

Stadium 2

Er is nog geen sprake van een expliciet besef van het bestaan van collectieve normen. Wel wordt van elkaar verwacht dat concrete individuele rechten erkend worden en dat conflicten worden opgelost door concrete wederkerige uitwisseling tussen individuen. Het lid zijn van een groep betekent dat men het eigenbelang van anderen respecteert en anderen helpt als het eigenbelang daarmee gediend is.

Stadium 3

Collectieve normen refereren aan relaties tussen groepsleden. Het groepslidmaatschap impliceert het naleven van gedeelde verwachtingen. Conflicten worden opgelost door een beroep te doen op de gedeelde verwachtingen, zoals vertrouwen en zorgen voor elkaar, die definiëren hoe iemand een goed groepslid kan zijn. Deze verwachtingen benadrukken wat noodzakelijk is om de relaties tussen de groepsleden te bevorderen.

Stadium 4

Collectieve normen benadrukken de gemeenschap als geheel onderscheiden van haar individuele leden. Leden zijn verplicht te handelen uit bezorgdheid voor het welzijn en de harmonie van de groep. Conflicten worden opgelost door te refereren aan iemands (rol-) verplichtingen of verantwoordelijkheden voor de activiteiten van de groep als geheel, zoals het aannemen van collectieve verantwoordelijkheid.

Ten tweede is er het probleem van de toepasbaarheid van het criterium van 'structureel geheel' op de collectieve stadia. Het begrip 'structureel geheel' impliceert dat een stadium niet afhankelijk is van een specifiek dilemma, maar consistent is over verschillende dilemma's. Een dergelijke consistentie van collectieve stadia over verschillende schoolproblemen en collectieve normen, is volgens Power et al. wel mogelijk maar niet noodzakelijk. Zij verwachten namelijk een grotere consistentie op het individuele niveau dan op het groepsniveau. Dit wordt toegeschreven aan het onderscheid tussen individuele intelligentie en collectief bewustzijn en het onderscheid tussen competentie en 'performance'. Op het individuele vlak komt het 'structureel geheel' van de stadia namelijk voort uit een logische behoefte aan consistentie bij het oplossen van verschillende dilemma's. Bij collectieve stadia is dat echter niet het geval. Daar gaat het om het in stand houden van de groep en het bevorderen van samenhang binnen de groep. Groepstabieleit en groepsamenhang sluiten logische inconsistentie niet uit.

Het belangrijkste probleem, gegeven het criterium van structureel geheel, is de vraag of de collectieve stadia logische structuren zijn of dat het inhouden zijn die, analoog aan de structurele ontwikkeling van de individuele stadia, veranderen. Power et al. vragen zich af of het gedeeld redeneren zich met een zodanige duidelijkheid en logische coherentie ontwikkelt, dat de structuur vastgesteld kan worden volgens de scoringsmethode die wordt gebruikt voor het vaststellen van de individuele stadia. In een groep is het belangrijker dat de leden het er over eens zijn hoe ze zich zullen gedragen dan waarom ze zich zo zullen gedragen. Bezorgdheid over de inhoud, de beslissing en de normen en waarden die daaraan gerelateerd zijn, domineert de structuur, het redeneren dat de

beslissing en de normen en waarden een morele betekenis geeft. Omdat de inhoud over de structuur domineert is een meer op de inhoud georiënteerde scoringsbenadering gewenst.

De relatie tussen inhoud en structuur is probabilistisch, er is geen noodzakelijke één op één relatie tussen instemming met de norm en het zelf kunnen produceren van de norm. Dit wordt duidelijk aan de hand van het volgende voorbeeld. Tijdens een discussie op een Just Community school werd besproken in hoeverre leerlingen het er over eens waren gezamenlijk geld op te brengen voor een klasgenoot waarvan dat geld gestolen was. De scorebare redeneringen van leerlingen die voor een gezamenlijke teruggave stemden, waren meestal van het derde of vierde stadium. De collectieve teruggave norm heeft dan ook een inhoud die gewoonlijk een overgangsstadium 3/4 structuur vergt voor zijn rechtvaardiging. Dat wil echter niet zeggen dat men geen voorstander van collectieve teruggave op een lager stadium kan zijn (bijvoorbeeld stadium 2, 'omdat je dan zelf niet zoveel geld hoeft te betalen'). Uit dit voorbeeld blijkt dat het stadium van de norm op een goede manier bepaald kan worden door je te richten op de uitingen waarin de norm voor het eerst werd voorgesteld. Een voorstel doen voor een nieuwe norm is een produktietaak, waarbij de inhoud het meest waarschijnlijk direct uit de onderliggende structuur volgt.

Om het stadium van de collectieve norm vast te stellen, proberen Power et al. na te gaan in welke mate een voorstel voor gedrag een algemeen gedeelde betekenis heeft. Scoring gebeurt door probabilistische criteria op één stadium of meerdere aangrenzende stadia. Zo wordt ondersteuning van collectieve teruggave gezien als een betrokkenheid bij collectieve verantwoordelijkheid op stadium 3 en 4. Dit betekent niet dat de meeste studenten de stadium 3 of 4 redenen kunnen produceren. Het betekent volgens de onderzoekers wel dat de meeste studenten stadium 3 of 4 redeneringen zouden identificeren als een betere uitdrukking van de gedeelde betekenis van het voorstel dan stadium 2 redeneringen. Ze verwijzen hierbij naar de claim dat individuen een voorkeur hebben voor argumenten van een hoger stadium dan het stadium van de argumenten die zij zelf kunnen produceren. Door deze preferentie wordt het mogelijk om het stadium van de collectieve norm te definiëren als het stadium dat de meeste individuen zouden identificeren als de beste uiting van de betekenis van de norm voor de groep.

Een derde probleem met het concept van collectief stadium is de hiërarchische integratie van de stadia, d.w.z. dat elk hoger stadium een lager stadium integreert en vervangt. De ontwikkeling van collectieve normen op Just Community scholen wordt beschreven als een ontwikkeling van stadium 2 naar stadium 3-4. Volgens Power et al. impliceert deze beschrijving dat er inderdaad stadium integraties en vervangingen zijn opgetreden. Echter, de score voor een collectieve norm kan niet zonder meer worden afgeleid uit individuele weergaven van die norm. Gesteld wordt dat het mogelijk is om

i.p.v. de collectieve norm een stadium-score te geven, je ook de stadium distributies van de individuele weergaven van die norm zou kunnen bepalen. Door het vaststellen van verschillen in de distributies van stadium scores over de tijd, wordt afstand genomen van de claim van stadium vervanging en kan een soepeler veronderstelling worden gehanteerd, namelijk dat een toenemend aantal leerlingen na verloop van tijd geneigd is normen op een hoger stadium weer te geven. Power probeert echter de data zo goed mogelijk binnen het hiërarchische model te plaatsen en geeft, telkens als een collectieve norm wordt vastgesteld, een enkele (bijv. 2) of aangrenzende (bijv. 2/3) stadium score. Bij de analyse van de gemeenschapsbijeekkomsten bleek dat het modale stadium van de individuele weergaven van de collectieve norm overeenkwam met de klinische indrukken van het collectieve stadium, dus deze score werd gebruikt als de collectieve score.

Met *gemeenschapsgevoel* bedoelen Power et al. (1989) een gedeelde waardering voor de school als gemeenschap. De term gemeenschap heeft betrekking op sociale relaties die voornamelijk gereguleerd worden door collectieve normen van zorgen, vertrouwen, integratie, participatie, openbaarheid en verantwoordelijkheid. Een gemeenschap wordt dan ook gekenmerkt door een hoge mate van solidariteit en een sterke verbondenheid met de groep; m.a.w. het verwijst naar wat de ideale school als gemeenschap zou moeten zijn.

Het stadium van het gemeenschapsgevoel verwijst naar rechtvaardigingen voor conflictoplossingen die gebaseerd zijn op een gedeelde waardering voor de gemeenschap. Centraal staat de vraag naar het stadium waarop de betekenissen, die leerlingen aan de school als gemeenschap geven, zich bevinden. De stadia van het gemeenschapsgevoel zijn opgenomen in Kader 4.

Kader 4: Stadia van het gemeenschapsgevoel

Stadium 2

Er bestaat geen duidelijk gemeenschapsgevoel. De gemeenschap wordt als een onpersoonlijk instituut instrumenteel gewaardeerd voor de concrete voordelen die de leden elkaar kunnen geven en de vrijheid die men heeft om het eigenbelang na te streven. De gemeenschap wordt gewaardeerd in zoverre zij tegemoet komt aan de concrete behoeften van haar leden.

Stadium 3

Het gemeenschapsgevoel refereert aan de relaties tussen groepsleden. De groep wordt gewaardeerd voor de gevoelens van vriendschap van haar leden. De waarde van de groep wordt gelijkgesteld aan de waarde van haar collectieve normatieve verwachtingen.

Stadium 4

De gemeenschap wordt expliciet gewaardeerd als een entiteit onderscheiden van haar individuele leden en hun onderlinge relaties. Het lid zijn van de gemeenschap betekent het aangaan van een sociaal contract waarin de groepsnormen- en idealen gerespecteerd worden.

1.2.4. De ontwikkeling van de morele performance

Met morele 'performance' wordt het moreel denken en beslissen in actuele handelingssituaties bedoeld. Power et al. (1989, p. 134) zijn van mening dat de morele performance alleen vastgesteld kan worden tijdens het uitvoeren van de handeling. Onderzoek naar deze morele performance is schaars; Power et al. noemen 1 studie.

Om de morele performance te benaderen worden morele oordelen en soms ook gedrag van individuen onderzocht in problematische situaties die tot zijn/haar leefwereld behoren. Zowel door de persoon zelf ervaren en geformuleerde dilemma's als gestandaardiseerde dilemma's worden gebruikt (zie respectievelijk de onderzoeken van Haan, 1975, Gilligan & Belenky, 1980, en van Kohlberg, Scharf & Hickey, 1974).

Het onderscheid tussen morele performance en moreel klimaat is o.i. niet altijd duidelijk. Kohlberg et al. (1974, p. 7, 8) wijzen er op dat ze hun gegevens van gevangenen over morele dilemma's in de gevangenis vanuit 2 perspectieven hebben geanalyseerd: (a) het stadium van individueel moreel oordelen van de gevangenen op deze dilemma's, (b) de waarneming van het moreel klimaat. De morele oordelen hadden betrekking op wat de gevangenen vonden wat ze zouden moeten doen in de gevangenis. Het moreel klimaat had betrekking op wat de gevangenen meenden wat er waar was van een rechtvaardige omgang in de instelling. Power et al. geven voorbeelden van gestandaardiseerde schoolgebonden moreel dilemma's ('Practical School Dilemmas Interview'). Dit bevat zowel vragen die betrekking hebben op de waarneming van zelf als naar de waarneming van de leerlingen van de school. Strikt gesproken gaat het in het eerste geval waarbij individuele morele oordelen worden gevraagd over contextgebonden dilemma's niet om (een aspect van) de morele performance, maar om contextgebonden moreel oordelen. Higgins et al. (1984) spreken echter ook in dit geval van morele performance. De uitspraken over de verhouding tussen moreel klimaat en de individuele morele performance zijn gebaseerd op uitspraken binnen een en hetzelfde interview die vanuit deze beide perspectieven ('zelf' en 'ander') worden geanalyseerd (Kader 1).

De kenmerken van Kohlbergs stadium model hebben uitsluitend betrekking op de morele competentie en niet op de morele 'performance' (Colby & Kohlberg, 1987, p. 8). Het is echter volgens Kohlberg onjuist om te veronderstellen dat bij de morele performance individuen er niet naar zouden streven optimaal te redeneren; alleen bij neerwaartse druk is daarvan geen sprake. Hij verwerpt de gedachtengang dat men in het dagelijks leven niet verder zou gaan dan een stadium 3 redenering als deze voldoet om het probleem op te lossen: "Because of the prescriptive nature of morality, lower level responses can never be said to be more morally appropriate than higher level responses, though one might want

to argue that under some circumstances they could be more psychologically appropriate. Thus, although we do distinguish between competence and performance in moral judgment, we hold that lower levels are used only in situations with a significant downward press." (Colby & Kohlberg, p. 8).

Dat neemt niet weg dat er fluctuaties optreden in het moreel redeneren, niet alleen voor wat betreft de performance, maar ook voor wat betreft de competentie: "We assume that competence and performance may differ to some degree depending on the problem being addressed, the context, and other factors. That is, people do not always use their highest stage of moral reasoning." (o.c.)

Het onderzoeken van de effecten van situatietekenen op het moreel oordeel vindt plaats in het onderzoeksprogramma van Krebs (Krebs et al., 1991). Krebs et al. (1991) hebben aangetoond dat situationele factoren de morele oordeelsvorming beïnvloeden. Als reden voeren zij nu juist de gedachte aan die Colby & Kohlberg verwerpen op grond van het idee van intrinsieke ontwikkeling. In situaties waar hogere stadia geen nieuwe gezichtspunten op de zaak werpen ontbreekt de motivatie een hoger stadium van oordelen te ontwikkelen, zo menen zij. Overigens heeft o.i. Kohlberg zelf meer mogelijkheden voor het gebruiken van lagere stadia voorgestaan: "[it] continues to be used in concrete situations where it is adequate or when efforts at solution [by the higher stage] have failed" (Kohlberg, 1969, p. 353).

1.3. Constructen en meetmethoden

Er zijn verschillende methoden en instrumenten voor het vaststellen van de morele competentie. Gemeenschappelijk is echter dat het subject hypothetische morele dilemma's krijgt aangeboden. De uitzondering op deze regel is de SRM-SF, waarbij niet langer gebruik wordt gemaakt van dilemma's. De meting van de morele competentie met behulp van instrumenten bestaande uit open vragen wordt onderscheiden van die waarbij gebruik wordt gemaakt van instrumenten bestaande uit meerkeuzevragen.

In het eerste geval is bijna altijd sprake van een instrument dat mondeling, individueel wordt afgenomen (interview). Bij open vragen moeten respondenten oordelen geven, redeneringen produceren, vandaar dat dit type instrument bekend staat als produktietaak. De geproduceerde redeneringen worden vervolgens gecodeerd.

In het tweede geval is bijna altijd sprake van een instrument dat schriftelijk, groepsgewijs wordt afgenomen (vragenlijst). Bij meerkeuze vragen wordt van respondenten gevraagd of ze de aangeboden oordelen herkennen als overeenkomend met die zij zelf gebruiken. Vandaar dat hier gesproken wordt van een herkenningstaak. De aangeboden oordelen zijn geconstrueerd als operationalisatie van de stadia in de

ontwikkeling van het moreel oordelen en derhalve gesleuteld. Omdat de intensieve interpretatie en codeerwerkzaamheden van de produktietaak niet aan de orde zijn bij een herkenningstaak, wordt deze in de literatuur wel een 'objectieve procedure' genoemd.

De instrumenten zijn vooral ontworpen voor diverse doelgroepen en praktische toepassingen: Moral Judgment Interview (MJl), Defining Issues Test (DIT), Sociomoral Reflection Measure (SRM), Sociomoral Reflection Objective Measure (SROM); van de instrumenten zijn uitgebreide en verkorte versies beschikbaar. Het Moral Judgment Interview (MJl), een produktietaak (Colby & Kohlberg, 1987), is het 'oer-instrument' waarop de verschillende versies teruggaan. Omdat dit instrument zeer tijdrovend is in afname en voor de codering van de uitspraken van de respondenten een intensieve training vereist, wordt dikwijls gebruik gemaakt van de 'alternatieve' instrumenten. Echter, hier zitten nogal wat haken en ogen aan. De alternatieve instrumenten kunnen gepaard gaan met andere criteria voor wat onder morele competentie wordt verstaan en zijn meestal tamelijk beperkt in hun toepassingsmogelijkheden.

Objectieve instrumenten zouden niet hetzelfde construct meten als instrumenten die spontane oordelen ontlokken (Evans, 1986, Colby & Kohlberg, 1987, p. 116, Gibbs et al., 1992, p. 34). De discussie is hier vooral gevoerd aan de hand van het onderscheid tussen het MJl, een instrument om de produktie van morele oordelen te ontlokken, en de Defining Issues Test (DIT), een 'objectief' instrument waarin een dilemma wordt aangeboden en daarbij gesleutelde overwegingen. Respondenten beoordelen in hoeverre ze deze overwegingen van belang vinden bij de beslissing in het dilemma en ordenen tenslotte de vier belangrijkste overwegingen naar voorkeur voor het nemen van de beslissing. In de loop der jaren zijn de verschillen tussen deze instrumenten steeds explicieter geworden, mede op grond van de verschillende onderzoeksprogramma's waarin ze werden ingebed. Terwijl het MJl steeds werd aangepast om te voldoen aan de theoretische eisen van Kohlbergs hard stadium model, ontwikkelde Rest, ontwerper van de DIT, een theorie die consistent was met resultaten van de sterker aan inhoud gebonden redeneringen van de DIT.

Colby en Kohlberg (1987, p. 116) zijn van mening dat een transformationeel model de gegevens het beste kan verklaren ("simple stage model"). Rest is volgens hen veeleer een additief of spekkook-model toegedaan ("complex stage model", Rest, 1986, p. 199 noemt zijn model een "soft stage model"). De structuur wordt niet volledig omgevormd, maar de verschillende cognitieve structuren blijven als het ware naast elkaar bestaan en afhankelijk van omstandigheden kan een persoon een structuur gebruiken die tot zijn uitrusting behoort. Overigens gaat dit standpunt niet zover dat voor een individu de stadia gelijkwaardige modi vormen een probleem op te lossen. Integendeel, de assumptie

dat respondenten 'pieken' op een bepaald stadium en in afnemende mate gebruik zullen maken van overwegingen naarmate deze stadia verder van het modale stadium afliggen is getoetst met data verkregen met de DIT (Boom & Moolenaar, 1988, Davison, Robbins & Swanson, 1978).

Colby en Kohlberg (1987, p. 116) onderkennen de mogelijkheid dat het zich ontwikkelende individu één basisstructuur heeft voor het oplossen van morele dilemma's en het rechtvaardigen van de gekozen oplossingen. Deze structuur heeft enige samenhang voor het individu en ondergaat de transformatie door ontwikkeling die kenmerkend is voor de morele competentie zoals hierboven (par. 1.2.1) beschreven. Andere aspecten van de morele oordeelsvorming, zoals het begrijpen van oordelen van anderen, hoeven niet deze stadiumgewijze ontwikkeling te doorlopen. Dat een individu de redeneringen van alle stadia beneden het eigen stadium begrijpt en een of twee stadia boven het eigen stadium waarop het individu zich bevindt (dit komt tot uitdrukking op de scores van de DIT) vindt Kohlberg niet strijdig met zijn model. M.a.w. het additieve model van Rest zou heel goed kunnen opgaan voor het begrijpen van morele redenen.

Dat er ook sprake kan zijn van een preferentie van oordelen van lagere stadia, zoals eveneens de DIT scores uitwijzen, verklaart Kohlberg hier niet uit neerwaartse druk, maar uit het feit dat veel inzichten van lagere stadia geldig blijven hoewel ze ingebed zijn in een complexere en meer ontwikkelde positie (vgl. Levine, 1979). Deze inbedding wordt volgens Kohlberg niet zichtbaar in het instrument van Rest waarin geïsoleerde overwegingen worden aangeboden. M.a.w. deze zelfde redeneringen zouden als ze door het subject spontaan geproduceerd zouden worden deel uitmaken van een uitgebreidere redenering die van een hoger stadium afkomstig is en als zodanig wordt onderkend en gescoord. De scoringsregel bij het coderen bevoordeelt hier de redenering van het hogere niveau.

Rest is evenals Kohlberg van mening dat een individu een idee kan herkennen en discrimineren voordat het dit spontaan onder woorden kan brengen. Een produktiemaat vereist volgens hem het verbaal vermogen zich uit te kunnen drukken - wil het subject daadwerkelijk de cognitieve structuur toegekend worden. Rest (1986a, p. 198, 1986b, p. 482) beschouwt dit als een serieus probleem voor de MJI: het gevolg is immers dat het moreel oordeelsniveau van het subject met behulp van de Moral Judgment Interview onderschat wordt, juist het intuïtieve morele oordeelsniveau dat het handelen leidt. Inderdaad wordt op herkenningstaken voor morele competentie gewoonlijk een hogere score behaald dan op produktietaken voor morele competentie. Dit gegeven speelt o.a. een rol in de vraag of er een 'natuurlijke' voorkeur zou bestaan voor hogere stadia van moreel redeneren. Zo'n natuurlijke voorkeur maakt deel uit van een verklaring voor het feit dat

onder bepaalde condities 'ontwikkeling' optreedt. Het gevaar van de herkenningstaak is echter overschatting.

Afhankelijk van de mate van heterogeniteit van de populaties variëren de correlaties tussen de scores op de MJI en de DIT tussen de .30 en de .70, hetgeen als bewijs wordt opgevat dat de instrumenten *niet* hetzelfde meten. De correlaties met leeftijd en opleiding zijn echter van dezelfde orde van grootte en ook in de gevoeligheid voor het vaststellen van effecten van interventies komen ze overeen (Rest, 1983). Rest (1986, p. 198) concludeert: "Perhaps the robustness of the underlying phenomena overcomes the particular vulnerabilities of each assessment tool". Duidelijk is echter dat zowel op begripsmatige als empirische gronden de DIT niet zonder meer als een substituuut voor de MJI mag worden beschouwd.

In deze discussie is niet verwezen naar de instrumenten die Gibbs heeft ontwikkeld en die in dit onderzoek worden gebruikt. De SRM (-SF) is evenals de MJI een produktiemaat en de SROM (-SF) evenals de DIT een herkenningstaak. De instructies bij de DIT en de SROM verschillen echter van elkaar. In het geval van de DIT wordt van subjecten gevraagd 12 aangeboden morele redeneringen bij een dilemma te beoordelen naar de mate van belangrijkheid op een Likertschaal van 1-5 en vervolgens welke 4 van deze 12 redeneringen zij als het meest belangrijk prefereren voor de beslissing van de hoofdpersoon in het dilemma. Bij SROM (-SF) wordt gevraagd welke van de aangeboden morele redeneringen zij zouden gebruiken of welke het meest lijkt op de reden die zij zouden geven. De SROM (-SF) als herkenningstaak lijkt zich daarmee minder ver te verwijderen van een produktietaak als de MJI en ook van de SRM(-SF) dan de DIT.

Gibbs zelf heeft onderzoek uitgevoerd naar de convergente validiteit van de herkenningstaken met de produktietaken (Basinger & Gibbs, 1987, Gibbs et al., 1984, zie verder par. 2.1.2). De herkenningstaken hebben enkele praktische voordelen boven de produktietaken, maar hebben als belangrijkste nadelen dat ze een behoorlijke aandacht en leesvaardigheid vereisen, waarover men gewoonlijk pas in de loop van de puberteit gaat beschikken. Dezelfde onderscheidingen die gehanteerd zijn bij instrumenten voor de morele competentie (schriftelijk-mondeling, produktie-herkenning, interpretatie-constructie) zijn van toepassing op instrumenten voor het moreel klimaat.

Een valideringsstudie zoals hier wordt voorgesteld is voorzover ons bekend niet eerder uitgevoerd. Dergelijke studies zijn gewenst in het kader van de instrumentontwikkeling, maar ook voor theorie-ontwikkeling (zie Krebs, 1994). Ze werpen ook nieuwe vragen op, bijvoorbeeld of de leerlingen op de herkenningstaak van het Moreel Schoolklimaat Instrument ook hoger scoren dan op de produktietaak, en zo ja, hoe dat gegeven dan zou moeten worden geïnterpreteerd.

1.4. Probleemstelling

Het NWO project is in de laatste fase van het onderzoek. In deze fase werd een schriftelijk versie van het Moreel Schoolklimaat Instrument (MSI) afgenomen samen met andere instrumenten, waaronder een instrument (SROM-SF) ter bepaling van de morele competentie, bij leerlingen van 32 scholen van voortgezet onderwijs. De probleemstelling van het NWO project luidt: "Kan een voldoende betrouwbaar en valide, schriftelijk af te nemen instrument worden ontwikkeld voor het vaststellen van de wijze waarop leerlingen van het voortgezet onderwijs het moreel klimaat op hun school waarnemen? In hoeverre zijn verschillen in de perceptie van het moreel klimaat tussen scholen te voorspellen uit verschillen in morele competentie van de leerlingen?"

Morele competentie en moreel klimaat beogen verschillende begrippen te meten. De probleemstelling van het ANGO onderzoek luidt:

"Wat is de convergerende en discriminerende validiteit van het moreel klimaat op scholen van voortgezet onderwijs, en de morele competentie van scholieren, gemeten met behulp van een vragenlijst en gemeten met behulp van een (half-)gestructureerd interview?"

Door middel van een confirmatorische factor analyse (CFA) van een multitrait-multimethod matrix zullen de convergerende en discriminerende validiteit van de instrumenten voor morele competentie en moreel klimaat worden onderzocht. Het begrip 'morele performance' is bij nadere beschouwing aan de begrippen morele competentie en klimaat toegevoegd omdat dit begrip volgens een bepaalde opvatting eveneens met het klimaatinstrument wordt vastgesteld.

Bij de Sociomoral Reflection Measure-Short Form (SRM-SF) voor morele competentie produceren de leerlingen hun oordelen over morele vraagstukken; deze redeneringen worden vervolgens gecodeerd. Bij de Sociomoral Reflection Objective Measure-Short Form (SROM-SF) geven de leerlingen aan welke morele redenering zij herkennen als overeenkomend met die zij zelf in de betrokken situatie gebruiken; op 'objectieve' wijze wordt de morele competentie vastgesteld. In het voorgaande is aangegeven dat volgens het hier gebruikte theoretische raamwerk de morele competentie middels één factor beschreven kan worden.

Bij het Moreel Schoolklimaat Instrument-interview (MSI-i) produceren de leerlingen de oordelen van hoe de leerlingen van hun school naar hun mening oordelen in dilemma's in en rond de school. Bij de morele performance gaat het om de oordelen die ze voor zichzelf produceren over deze dilemma's. Het Moreel Schoolklimaat Instrument-

schriftelijk (MSI-s) stelt op objectieve wijze vast welke oordelen leerlingen bij hun medeleerlingen waarnemen. Tevens wordt met dit instrument op objectieve wijze de individuele morele performance vastgesteld.

De vragenlijsten zijn meerkeuze instrumenten: leerlingen beoordelen in hoeverre uitspraken die worden aangeboden bij standaard dilemma's van toepassing zijn. De uitspraken zijn voorgesleuteld naar de onderscheiden stadia. De vragenlijsten zijn afgeleid van de gegevens die zijn verzameld met behulp van mondelinge interviews.

2. OPERATIONALISERING EN INSTRUMENTALISERING VAN KERNBEGRIJPPEN

2.1. Morele competentie: de Sociomorele Reflectie

Het Moral Judgment Interview (MJI) van Kohlberg bevat minimaal 21 vragen die het redeneren van personen uitlokken bij morele dilemma's waarin de aspecten van rechtvaardigheid aan de orde komen. Er zijn negen dilemma's. Per interview worden drie dilemma's gebruikt. Het meest bekende dilemma is het Heinz-dilemma: Heinz verkeert in de situatie dat zijn vrouw zal sterven als zij niet geholpen wordt met een pas ontdekt medicijn. De plaatselijke apotheker heeft dit medicijn ontdekt en wil het alleen voor veel geld verkopen. Heinz heeft niet zoveel geld en de apotheker weigert hem het medicijn te verstrekken. Wat moet Heinz doen: de wet overtreden door het medicijn te stelen, waarmee zijn vrouw gered wordt, of de wet respecteren en toezien hoe zijn vrouw sterft?

In het interview wordt aan mensen gevraagd welke oplossing ze kiezen en waarom ze die oplossing kiezen. Om de uitspraken te beoordelen is een omvangrijke handleiding ontwikkeld. Op basis van de antwoorden van het subject wordt een 'Moral Maturity Score' (MMS) toegekend die loopt van 100 tot 500 (volledig stadium 1 tot volledig stadium 5). Stadium 6 wordt als een theoretisch stadium beschouwd omdat er nauwelijks mensen zijn die op dit stadium redeneren. Daarnaast wordt het zogenaamde 'globale stadium' bepaald met behulp van een 13-punts schaal waarbij ook overgangsstadia geïdentificeerd worden.

Het Moral Judgment Interview is een instrument dat erg tijdrovend is, voornamelijk doordat het erg lang duurt om het scoren te leren. Ook na een intensieve training is het moeilijk om betrouwbaar te scoren. Deze praktische argumenten waren voor Gibbs beslissend om te proberen alternatieve meetinstrumenten te ontwikkelen (Gibbs et al., 1984). Om diezelfde redenen ontwikkelde Rest al eerder de Defining Issues Test (DIT, Rest, 1979), die eveneens in het Nederlands is vertaald (Overwegingen bij Sociale Beslissingen, OSB, Hoeks, Dudink & Wouters, 1984). Dit instrument is echter speciaal bedoeld om vast te stellen in hoeverre mensen redenen van het postconventioneel niveau gebruiken ('herkennen').

Voor het meten van het niveau van moreel oordelen ontwikkelde Gibbs achtereenvolgens de Sociomoral Reflection Measure (SRM), de Sociomoral Reflection Objective Measure (SROM), de Sociomoral Reflection Objective Measure-Short Form (SROM-SF) en de Sociomoral Reflection Measure-Short Form (SRM-SF). De nadelen van deze instrumenten waren dat ze of te moeilijk waren voor basisschool kinderen en voor mensen met leesproblemen, of dat het niet mogelijk was om te onderscheiden tussen delinquenten en niet-delinquenten. Voor onze doelgroep, leerlingen van het voortgezet

onderwijs, gaan deze bezwaren in het algemeen gesproken niet op. Echter de schriftelijke instrumenten werden afgenomen in een sessie die (te) veel aandacht van de leerlingen vroeg, m.n. van lbo leerlingen. De afname bij deze leerlingen werd ook aangepast.

2.1.1. De Sociomoral Reflection Measure - Short Form (SRM-SF)

De SRM-SF (zie Bijlage D) is een mondeling interview voor het vaststellen van het stadium van moreel redeneren in hypothetische dilemma's. De SRM-SF is een voor kinderen bewerkte, vereenvoudigde en kortere versie van het Moral Judgment Interview (MJI, Colby & Kohlberg, 1987). Het instrument heeft als voordeel dat de onderzoeker door zelftraining de benodigde vaardigheid kan verwerven om te scoren, dit in tegenstelling tot de MJI waarvoor speciale cursussen dienen te worden gevolgd. De scoringsprocedure voor dit instrument is vertaald (Zwart-Woudstra, Meijer, Fintelman & Van IJzendoorn, 1993). Bij het bezoek van Gibbs aan Nederland is intensief met hem samengewerkt bij het coderen van een groot aantal interviews. De bereikte overeenstemming in het coderen van de uitspraken van de leerlingen voldeed ruimschoots aan alle door Gibbs et al. (1992) gestelde criteria.

De SRM-SF vereist veel minder leeswerk dan de SRM doordat er geen dilemma's in voorkomen. Het instrument bestaat uit elf vragen die kort ingeleid worden. Bijvoorbeeld vraag 1: "Hoe belangrijk is het dat mensen beloftes tegenover vrienden nakomen, als ze kunnen?", wordt ingeleid met de zin "Denk eens terug aan een keer dat u iets heeft beloofd aan een vriend(in) van u". Nadat respondenten aangekruist hebben hoe belangrijk ze het vinden, wordt gevraagd om uit te leggen waarom ze dit vinden. De vragenlijst behandelt wel dezelfde thema's als het Moral Judgment Interview: het houden van beloftes, het redden van iemands leven, je aan de wet houden, diefstal, de waarheid spreken en dergelijke.

Voor elke vraag krijgt de respondent een score op een bepaald stadium. Het gemiddelde van deze scores vormt de uiteindelijke SRMS-score. Het Globale stadium wordt bepaald aan de hand van deze SRMS score volgens een vaste normering (zie Tabel 1). Vragenlijsten moeten

Tabel 1: Normtabel stadium scores SRM-SF

SRMS-score	Stadium
1.00-1.25	Stadium 1
1.26-1.49	Overgangsstadium 1(2)
1.50-1.74	Overgangsstadium 2(1)
1.75-2.25	Stadium 2
2.26-2.49	Overgangsstadium 2(3)
2.50-2.74	Overgangsstadium 3(2)
2.75-3.25	Stadium 3
3.26-3.49	Overgangsstadium 3(4)
3.50-3.74	Overgangsstadium 4(3)
3.75-4.00	Stadium 4.

minimaal 7 scoorbare antwoorden bevatten om een Globale stadium score te krijgen.

Daarnaast wordt een 'Moreel type A' of 'Moreel type B' score toegekend. Het verschil tussen moreel type A en B is de mate waarin prescriptieve, voorschrijvende idealen

gebruikt worden bij het moreel oordelen. Moreel type B denken bevat volgens Gibbs drie componenten:

(a) Evenwicht: moreel type B denken is het meest rechtvaardig. Het gaat om de ideale wederkerigheid van interpersoonlijke of maatschappelijke verwachtingen.

(b) Fundamenteel waarden: moreel type B denken is het meest universeel, waarden gelden voor de hele mensheid en niet alleen voor de eigen relaties of maatschappij.

(c) Geweten: moreel type B denken is het meest intern, het meest prescriptief. De idealen van volwassen moraliteit worden van binnen gevoeld (Gibbs et al., 1992, p. 25-26).

Een respondent krijgt een moreel type B score toegekend wanneer hij minstens twee van de drie elementen Evenwicht, Fundamenteel Waarden en Geweten gebruikt heeft, zelfs als het maar in één vraag is.

Gibbs heeft een handleiding ontwikkeld waarmee via zelftraining geleerd kan worden om de vragenlijsten te scoren. De SRM-SF voldoet volgens Gibbs aan de betrouwbaarheids- en validiteitscriteria en is daarnaast een efficiënt en praktisch instrument omdat het groepsgewijs afgenomen kan worden, weinig tijd kost om in te vullen en zelfs bij jonge kinderen gebruikt kan worden. Verder is het redelijk snel en makkelijk te scoren (Gibbs et al., 1992).

Intercodeurbetrouwbaarheid

De SRM-SF interviews werden door twee personen gescoord die zich zelf getraind hebben met behulp van de handleiding van Gibbs. Na het volgen van een 'advanced workshop' bij Gibbs hebben zij de handleiding in het Nederlands vertaald. Deze Nederlandse handleiding is vervolgens gebruikt om de redeneringen, die ontlokt waren met behulp van de SRM-SF, te scoren van een groep moeders en een groep VWO-scholieren. Voor het meten van de intercodeurbetrouwbaarheid is 10% van de interviews door beide personen gescoord. Bij deze eerste set interviews bleek de intercodeurovereenstemming nog te laag ($r = .75$). Na het scoren van een set trainingsprotocollen is een nieuwe set van 10% van de interviews door beiden gescoord. De uiteindelijke intercodeurovereenstemmings-statistieken staan vermeld in Tabel 2.

Tabel 2: Intercodeurovereenstemming SRMS-SF

statistiek	waarde	minimum vereiste
SRMS correlatie r	.94	.80
Gemiddelde absolute SRMS discrepantie (in punten)	0.09	0.20
Globale Stadium overeenkomst binnen 1 interval	100 %	80%
Exacte Globale Stadium overeenkomst	68 %	50%

Nadat aldus een bevredigende intercodeursovereenstemming was verkregen zijn de overige interviews gescoord. De 24 interviews die door beide personen gescoord waren, hebben een 'intercodeur-agreementscore' gekregen. Twee interviews hebben geen SRMS-score gekregen omdat er meer dan vier niet scorebare antwoorden waren.

2.1.2. Sociomoral Reflection Objective Measure - Short Form (SROM-SF)

De SROM-SF (zie Bijlage E) is een schriftelijk instrument voor het vaststellen van het stadium van moreel redeneren in hypothetische dilemma's. De SROM (Gibbs et al., 1984, De Mey, 1994, Van IJzendoorn, 1986) heeft betrekking op de herkenning of evaluatie van aangeboden morele redeneringen. Hiervan is een herziene, verkorte versie beschikbaar (SROM-Short Form, Basinger & Gibbs, 1987).

Basinger en Gibbs (1987) rapporteren een onderzoek naar de psychometrische kwaliteit van de SROM-SF bij drie steekproeven: kinderen van het 6-de leerjaar (12 jarigen), delinquente jongeren (16-17 jarigen) en leerlingen van het 11-de leerjaar (16-17 jarigen). Voor het vaststellen van de concurrente validiteit van de SROM-SF werd de SRM gebruikt, een schriftelijke produktietaak die evenals de SRM-SF groepsgewijs afgenomen werd. De tijdsduur tussen de afnames was 2 weken. De SROM-SF had een interne consistentie van .76 en een hertest betrouwbaarheid van .75. Voor de delinquente groep was de test-hertest betrouwbaarheid aanzienlijk lager (.46). De samenhangen tussen de SROM-SF scores met de SRM scores waren matig voor de zestien-zeventienjarigen (.48), en voor de twaalfjarigen slecht (.21).

Voor het vaststellen van de constructvaliditeit werden een intelligentiemaat en een sociale wenselijkheidsschaal gebruikt. De onderzoekers rapporteren een positieve samenhang met verbale intelligentie en een negatieve samenhang met sociale wenselijkheid voor zowel de SROM-SF als de SRM. Delinquente jongeren scoren lager dan de niet-delinquente jongeren; als gecontroleerd wordt voor intelligentie is het verschil echter niet langer significant. Gezien deze en andere goede eigenschappen (korte afnameduur, weinig uitval) van de SROM-SF bevelen Basinger en Gibbs in grote onderzoeken het gebruik van de SROM-SF aan boven de SROM als meting van het niveau van de sociomorele reflectie bij zestienjarige leerlingen van het voortgezet onderwijs. Gebruik bij twaalfjarigen wordt afgeraden.

In een vooronderzoek vond Stams (1994, p. 75) bij een steekproef van 48 scholieren in het voortgezet onderwijs een correlatie van .77 tussen de Nederlandstalige versies van de SRM-SF en de SROM-SF. Dat is aanzienlijk hoger dan de hierboven vermelde .48 die Basinger en Gibbs vonden tussen de SROM-SF en een schriftelijke versie van de SRM. Dat Stams de verkorte versie gebruikte van de SRM (zonder dilemma's) die mondeling is afgenomen maakt het resultaat nog opmerkelijker. Uit het onderzoek van Stams bleek tevens dat SRM-SF en SROM-SF hetzelfde correlatiepatroon lieten zien met verbale intelligentie en sociale wenselijkheid als in het onderzoek van Basinger en Gibbs: een matig positieve correlatie met verbale intelligentie (respectievelijk .44 en .37) en een laag negatieve correlatie met een sociale wenselijkheid (respectievelijk -.23 en -.15).

2.2. Moreel klimaat: Zelf en Ander

Het *Moreel schoolklimaat* is geoperationaliseerd in twee instrumenten: een vragenlijst (MSI-s) en een interview (MSI-i). Beide instrumenten zijn zoveel mogelijk analoog aan elkaar ontworpen. De verschillende aard van de instrumenten (een productie- versus een herkenningstaak) maakte echter een volledige overeenkomst onmogelijk. Daarnaast worden in de minder belastende schriftelijke taak een aantal extra concepten gemeten, die niet in het mondelinge interview konden worden opgenomen. Daarvan is alleen het construct *Stadium van gemeenschapsgevoel* in het onderhavige valideringsonderzoek betrokken.

Zoals eerder vermeld kan het moreel klimaat in kaart worden gebracht aan de hand van de waarneming van leerlingen van morele kwesties op school. Het moreel klimaat verwijst daarbij naar het gedeelde gemeenschappelijk perspectief van leerlingen ten aanzien van de socio-morele werkelijkheid op school. De reikwijdte van dat gemeenschappelijke perspectief varieert van de perceptie door een leerling van zichzelf tot de perceptie van medeleerlingen: over sommige zaken menen leerlingen slechts namens zichzelf te kunnen spreken, over andere namens de vrienden of vriendinnen, of klasgenoten. De instrumenten zijn zo ontworpen dat de respondent steeds systematisch dit onderscheid kon maken tussen zelf-perceptie en perceptie-van-de-ander (zijn medeleerlingen). De eerste categorie variabelen zal in het vervolg met *Zelf* worden aangeduid, de tweede categorie met *Ander*.

Enige discussie is mogelijk over de vraag of de eerste categorie variabelen een operationalisering vormt van -een onderdeel van- het moreel klimaat of van een vorm van morele performance. Power et al. (1989, p. 134) stellen echter dat morele 'performance' betrekking heeft op het moreel denken en beslissen in *actuele* handelingssituaties. Zij zijn

van mening dat de morele performance alleen vastgesteld kan worden *tijdens het uitvoeren van de handeling*. Volgens deze opvatting betreft de meting van de zelf-perceptie ten aanzien van morele kwesties op school dus geen meting van morele performance, aangezien het hier hypothetische, zij het realistische, dilemma's betreft.

2.2.1. Moreel schoolklimaat instrument-interview-produktie (MSI-i)

De pogingen het moreel schoolklimaat te meten zijn van meer recente datum. Er zijn ons inmiddels 3 schriftelijke varianten bekend om de dimensies van het moreel schoolklimaat te meten: de Moral Atmosphere Questionnaire van Power en Higgins (niet gepubliceerd), de Moralische Atmosphäre Fragenbogen van Georg Lind en ons Moreel Schoolklimaat Instrument.

Daarnaast bestaat er een interview versie voor het vaststellen van het moreel schoolklimaat van Higgins en Power en van ons. Dezelfde dimensies en categorieën van het moreel schoolklimaat die ten grondslag liggen aan de vragenlijst zijn ook gebruikt voor het coderen van de uitspraken in het interview. Het MSI-i is een gestandaardiseerd interview dat, analoog aan het 'Practical School Dilemma's Interview' (Power et al., 1989, p. 242-248), is ontwikkeld voor leerlingen van het voortgezet onderwijs.

Het interview bevat vier morele dilemma's, die in de school zijn gesitueerd: twee dilemma's hebben betrekking op een antisociale gebeurtenis (diefstal en pesten) en twee op een prosociale gebeurtenis (helpen en participatie in schoolse activiteiten). Aan de hand van de dilemma's worden ongeveer dertig vragen gesteld, die erop gericht zijn te achterhalen: (a) hoe leerlingen zelf redeneren en welke beslissingen zij nemen (*zelf*), en (b) hoe leerlingen denken dat (het merendeel van) hun medeleerlingen redeneren en beslissingen nemen (*ander*). Ter illustratie volgt hieronder het helpen-dilemma met twee vragen:

"Soms zitten er leerlingen in de klas die niet bij de groep horen, er een beetje buiten vallen of niet populair zijn. Bert is zo'n onpopulaire leerling. Hij presteert slecht in een vak waar de meeste leerlingen goed in zijn, en blijft misschien wel zitten. Na het gezamenlijk gedeelte van de les kunnen jullie de rest van de tijd zelf werken. Bert vraagt of iemand hem wil helpen."

- Wat zou jij in een dergelijke situatie doen? Waarom? (*zelf*)
- Wat denk je (verwacht je) dat je klasgenoten in een dergelijke situatie zullen doen? Hoeveel van je klasgenoten zullen Bert in deze situatie helpen? Waarom? (*ander*).

De vraag naar de gedragskeuze heeft betrekking op "de inhoud van de norm" (zie paragraaf 2.2.2.), de waaromvraag lukt een redenering uit waarmee het stadium van de norm vastgesteld wordt.

Scoring.

Voor de scoring van de uitspraken van leerlingen op het stadium van de collectieve norm moest worden volstaan met een eenvoudige handleiding. Hierin wordt aangegeven aan welke kenmerken een uitspraak moet voldoen om in een bepaald stadium van de norm geplaatst te worden. Deze kenmerken komen in grote lijnen overeen met de kenmerken van de individuele stadia. Het verschil is dat de gedragskeuzes ofwel de inhoud van de beslissingen meestal niet relevant zijn voor de bepaling van de individuele stadia, terwijl ze wel belangrijk zijn voor de scoring van de stadia van de collectieve norm. Het gaat dan vooral om het 3e en 4e stadium van de collectieve norm. Op deze stadia worden sociale relaties en gedeelde verwachtingen, zoals vertrouwen en zorgen voor elkaar belangrijk. Toegepast op bijvoorbeeld het helpen-dilemma, betekent dit dat wanneer leerlingen stadium 3 of 4 redeneringen produceren, deze gekoppeld zijn aan de keuze om medeleerlingen te helpen. Het is zeer onwaarschijnlijk en in dit onderzoek ook niet voorgekomen dat leerlingen stadium 3 of 4 redenen kunnen produceren om een medeleerling niet te helpen. Een dergelijke, meer inhoudsgerichte scoring, is ook door Power uitgevoerd.

Deze wijze van scoring van het stadium van de collectieve norm is zowel op de zelf-perceptie als op de ander-perceptie toegepast.

Intercodeurbetrouwbaarheid

In het vooronderzoek werden voor de verschillende dilemma's Cohen kappa's vastgesteld variërend van .70 tot .80. Voor dit onderzoek is, in verband met de beschikbare middelen, één dilemma (helpen) volledig uitgewerkt.

De Moreel Schoolkimaat interviews werden door een persoon gescoord die van begins af aan betrokken is geweest bij de ontwikkeling van het instrument. De mate van overeenstemming met een ander ingevoerde persoon werd vastgesteld over 10 interviews afkomstig van twee scholen. Voor het dilemma helpen varieerde kappa van .75 tot .92 voor de interviews afzonderlijk met een gemiddelde van .83. Voor het scoren van de redenen, dat afzonderlijk heeft plaatsgevonden, was kappa .90

2.2.2. Moreel schoolklimaat instrument -herkenning (MSI-s)

Het schriftelijke moreel schoolklimaat instrument (MSI-s) is ontwikkeld op basis van twee interview-rondes onder leerlingen van het voortgezet onderwijs. Het tracht de waarneming van het moreel klimaat te meten van de leerlingen van scholen met opleidingsniveaus variërend van VBO t/m VWO.

De MSI-s bestaat uit twee gestandaardiseerde schooldilemma's (Helpen en Diefstal, zie Bijlage B) en een vragenlijst ("Vragen over jou en de school") met bijbehorende meerkeuze items die - onder andere - een operationalisering zijn van de categorieën van de dimensies van moreel schoolklimaat. De MSI vraagt de leerlingen naar hun eigen waarneming en naar de waarneming van de meerderheid van de leerlingen van hun klas of school. Tijdens het vooronderzoek bleek dat leerlingen meestal geen uitspraak waagden over de meeste medeleerlingen maar wel iets konden zeggen over de meningen en gedragingen van de klasgenoten, daarom is gekozen om vooral naar de klasgenoten te vragen.

De volgende zes constructen van Power et al. worden gemeten:

1. Het niveau van de institutionele waarde
2. Het stadium van het gemeenschapsgevoel
3. De inhoud van de norm
4. De mate van collectiviteit van de norm
5. De fase van de norm
6. Het stadium van de norm

In de vragenlijst "Vragen over jou en de school" worden de eerste twee constructen gemeten, terwijl middels de schooldilemma's de laatste vier constructen worden gemeten. Daarnaast zijn een aantal schalen in de vragenlijst opgenomen die wij op grond van literatuurstudie en onze eerste interview-ronde van belang achten voor het beoordelen van het schoolklimaat. Het betreft: 1) Gedrag op school (overtredingen/kleine criminaliteit); 2) Democratie/ invloed van de leerlingen op school; 3) Band met de school; 4) Buitenschoolse activiteiten (activiteiten op school in de vrije tijd); 5) Relatie met en mening over de leraren.

De validatiestudie betreft alleen het stadium van het gemeenschapsgevoel, de inhoud van de norm en het stadium van de norm. De operationalisering van deze constructen wordt hieronder uitgewerkt.

Het stadium van het gemeenschapsgevoel.

In de MSI-s is een vragenlijst opgenomen die het stadium van het gemeenschapsgevoel tracht te meten (zie Bijlage B). Hoewel in beginsel het stadium van gemeenschapsgevoel

(de betekenissen die leerlingen aan de school als gemeenschap geven) in grote mate af te leiden is uit het stadium van de norm (de gedeelde betekenissen van de normen), is getracht beiden toch afzonderlijk te meten. In een volgende paragraaf zal uitgebreid worden ingegaan op de samenhang tussen de twee constructen.

De vier stadia van het gemeenschapsgevoel worden gemeten middels vier schalen, met in totaal 21 items:

1. Machtsgebruik (of -misbruik) (5 items)
2. Concrete wederkerigheid (6 items)
3. Relaties en delen/ vriendschap (5 items)
4. Sociaal contract (5 items)

De items zijn geformuleerd als uitspraken over gedrag op school of als stellingen die wij typerend vinden voor de vier stadia. De leerlingen moeten aangeven in hoeverre deze uitspraken waar zijn voor hen zelf en voor hun klasgenoten.

Het stadium van de norm.

In het vooronderzoek is met behulp van de MSI-i (zie de vorige paragraaf) een groot aantal redeneringen van de leerlingen ontlokt over een viertal praktische schooldilemma's. Deze zijn voor zover mogelijk gescoord volgens de vier stadia van Power et al. (zie Tabel 2). In de MSI-s zijn een aantal van deze door de leerlingen geproduceerde scoorbare redenen opgenomen. Om daarbij voor alle stadia steeds een antwoordcategorie te hebben waren enkele aanvullingen nodig.

De leerlingen moeten per dilemma steeds van iedere reden aangeven of deze veel, een beetje of niet lijkt op een reden die zij zouden geven (zie het voorbeeld in Bijlage B). Na een blok van vier redenen (vier stadia) geven ze aan welke van de vier hun voorkeur heeft. Ook als leerlingen in het voorbeeld hebben aangegeven niet te willen helpen moeten zij een reden om wel te helpen kiezen.

Het 'stadium van de norm' wordt berekend als de gemiddelde stadiumscore op vijf vragen: De redenen die de voorkeur van de leerlingen hebben om een onpopulaire klasgenoot te helpen (twee scores), de reden om niet te stelen van een klasgenoot, de reden om in te grijpen als je een klasgenoot ziet stelen in de klas en de reden om tegen de leraar zeggen wie gestolen heeft.

De inhoud van de norm

De inhoud van "de norm" wordt in dit onderzoek beperkt tot prosociaal gedragskeuze in een aantal situaties. Wij proberen niet na te gaan welke normen überhaupt op school gelden maar in hoeverre een aantal specifieke normen worden gedeeld.

In het helpen-dilemma - een onpopulaire klasgenoot helpen - gaat het om de norm 'elkaar helpen'/'zorg voor elkaar'. Er wordt in 5 situaties gevraagd naar het gedrag van de leerling zelf en dat van zijn klasgenoten/medeleerlingen, zie het voorbeeld in Bijlage B, vraag 1B en 1C. In de eerste situatie is helpen nog 'makkelijk', in de volgende twee situaties worden er kosten aan verbonden, je wordt gepest als je helpt of je mist je favoriete televisieprogramma. De vierde situatie betreft het geval dat de leraar vraagt of de leerling wil helpen en in de laatste situatie gaat het niet meer om een klasgenoot, daar wordt de bereidheid om een anonieme medeleerling te helpen onderzocht.

De relatie tussen het stadium van de collectieve normen en het stadium van het gemeenschapsgevoel.

Hoewel er een duidelijk verschil is tussen de stadia van de collectieve normen en de stadia van het gemeenschapsgevoel was het niet mogelijk ze onafhankelijk van elkaar te coderen bij de door ons verzamelde data.

Collectieve normen veronderstellen een zeker stadium van gemeenschapsgevoel. Anderzijds is handhaving van stadia van gemeenschapsgevoel niet mogelijk zonder dat er collectieve normen ontwikkeld zijn of worden. Zo kan een collectieve norm van vertrouwen in een klas alleen ontstaan als leerlingen om elkaar geven en het belangrijk vinden om op een goede manier met elkaar om te gaan. En andersom, als er een gemeenschapsgevoel bestaat dan impliceert dit dat er een collectieve norm van bijvoorbeeld vertrouwen bestaat.

Power bepaalde het stadium van het gemeenschapsgevoel op de Just Community scholen. Hij scoorde hiertoe redeneringen van leerlingen, verkregen met o.a. het 'Practical School Dilemma's Interview' (Power et al., 1989, p. 242-248). Dezelfde redeneringen leveren daarbij vaak zowel een score op voor het stadium van de collectieve norm als een score voor het stadium van het gemeenschapsgevoel. Een voorbeeld hiervan zou de volgende uitspraak kunnen zijn van een leerling op het helpen-dilemma: "Leerlingen op deze school zouden een medeleerling helpen omdat we als school een gemeenschap zijn, dat je elkaar helpt is hier vanzelfsprekend. Leerlingen vertrouwen erop dat we elkaar helpen, als medeleerlingen hoor je toch bij elkaar." Aan de hand van deze uitspraak wordt het stadium van de collectieve norm helpen op stadium 3 gescoord (er wordt verwezen naar sociale relaties tussen de leerlingen, vertrouwen en een onderlinge verbondenheid tussen leerlingen komt tot uitdrukking), maar tevens is er sprake van een gemeenschapsgevoel op stadium 3 (relaties tussen groepsleden en vriendschap zijn belangrijk).

Het gemeenschapsgevoel wordt door Power dus bepaald aan de hand van de collectieve normen. Nu is het niet verwonderlijk dat leerlingen op de Just Community scholen vaak collectieve normen naar voren brengen die gerechtvaardigd worden door een gemeenschapsgevoel. Het doel van de interventies op deze scholen is immers het creëren van een rechtvaardige gemeenschap, waar leerlingen en leraren om elkaar geven. De bij het hier beschreven onderzoek betrokken leerlingen brachten bij het beantwoorden van de dilemma's soms wel gemeenschapsnormen (zoals vertrouwen) naar voren, maar rechtvaardigingen die expliciet op een (stadium van het) gemeenschapsgevoel duiden ontbreken. Nu kan men inbrengen dat het ontbreken van dergelijke rechtvaardigingen er op wijst dat de school niet als gemeenschap gewaardeerd wordt en dat het stadium van het gemeenschapsgevoel dan ook niet bepaald kan worden. Gezien de hierboven beschreven correspondentie tussen de collectieve normen en het gemeenschapsgevoel kan dit probleem echter ook opgelost worden door te stellen dat wanneer er collectieve normen op stadium 3 of 4 gevonden worden, het gemeenschapsgevoel zich ook op stadium 3 of 4 bevindt (vanaf het derde stadium wordt de school als gemeenschap gewaardeerd en zijn leerlingen zich bewust van collectieve normen die aan relaties tussen groepsleden refereren). Voor deze laatste mogelijkheid is gekozen.

2.3. Concurrerende constructen

Met het oogmerk een vorm van discriminante validiteit van de constructen Morele Competentie, Morele Performance en Moreel Klimaat te kunnen vaststellen is bij de respondenten tevens een aantal instrumenten afgenomen die concurrerende constructen meten. Het betreft hier een intelligentiemeting met behulp van de Groninger Intelligentie Test en een sociale wenselijkheidmeting met behulp van een instrument uit een onderzoek van Brugman (1988). De operationalisatie van de constructen in deze instrumenten wordt in de volgende twee paragrafen beschreven.

2.3.1. Groninger Intelligentie Test

Om de verbale intelligentie van de proefpersonen vast te stellen is gekozen voor een onderdeel uit de schriftelijke verkorte vorm van de Groninger Intelligentie Test (svv-GIT), nl. de zogenaamde *Woordmatrijs* (Kooreman & Luteijn, 1987). Dit onderdeel bevat 20 meerkeuze-items waarbij een logisch verband tussen twee paren van woorden ontdekt moet worden: het gaat hierbij om analoog redeneren met verbaal materiaal. De items lopen geleidelijk op in moeilijkheidsgraad.

Een belangrijk uitgangspunt bij de keuze voor dit instrument was de overweging dat de test weinig tijd in beslag neemt. De totale test-batterij vergt los van de GIT al veel tijd

en aandacht van de respondent. Bij de keuze voor alleen de woordmatrijs moet wel rekening gehouden worden met een mogelijk geringere criteriumvaliditeit t.o.v. de gehele svv-GIT. Kooreman en Luteijn rapporteren een correlatie van 0.77 (n=539) tussen de woordmatrijs-score en de totale svv-GIT-score. Bovendien blijkt uit onderzoek van Van Dongen en anderen (1972) dat de svv-GIT een ander soort intelligentie meet dan de WAIS (Wechsler Adult Intelligence Scale) of Raven's Progressive Matrices. Ons inziens hoeft dit geen bezwaar te zijn voor het gebruik van de Woordmatrijs in het onderhavige onderzoek omdat de woordmatrijs-score gebruikt zal worden om de discriminante validiteit tussen verbale intelligentie en de verschillende morele constructen vast te stellen.

Het aantal goede antwoorden op de woordmatrijs noemen we in het vervolg de ruwe woordmatrijs-score. Deze ruwe score kan worden omgezet in een genormeerde GIT-score op basis van een normtabel uit het normerings-onderzoek van Kooreman en Luteijn. In het onderhavige onderzoek is dit echter om een aantal redenen niet gebeurd. Ten eerste verschilt de normgroep qua samenstelling duidelijk van de onderzochte groep respondenten. Kooreman en Luteijn normeerden de svv-GIT bij een landelijk steekproef van 539 AAW/WAO uitkeringsgerechtigden met een gemiddelde leeftijd van 45 jaar en een LBO opleidingsniveau, die voor het grootste deel behoren tot de beroeps categorieën 'ambacht' en 'industrie'. Uit verschillend onderzoek bleek (zie Luteijn & Van de Ploeg, 1983) dat met name de svv-GIT onder- of overschattingen geeft voor bepaalde groepen respondenten, met andere woorden: een zekere steekproef-gebondenheid is te verwachten. Ten tweede staan de ruwe scores een realistischer vergelijking met andere constructen toe, aangezien de overige constructen in het onderhavige onderzoek evenmin genormeerd worden.

Een beschrijving van de betrouwbaarheid en validiteit van de GIT test is o.a. te vinden in de test handleiding van Kooreman en Luteijn (1983). Een volledige beschrijving van de psychometrische gegevens van de ssv-GIT is te vinden in de test-handleiding van Kooreman en Luteijn (1987). In paragraaf 1.6.1 en verder zullen deze gegevens vergeleken worden met de belangrijkste resultaten van de afname in het onderhavige onderzoek.

2.3.2. Sociale wenselijkheidsschaal

De Sociale wenselijkheidsschaal bestaat uit 11 items. De schaal is oorspronkelijk ontleend aan de PMT-K en is eerder gebruikt in onderzoeken van het project waardenontwikkeling (Brugman, 1988, p.178 e.v.). De items bevatten een vraag naar gedrag dat iedereen wel eens vertoont, terwijl de sociale standaard vereist dat het niet wordt vertoond. Indien de respondent de vraag hoe vaak het gedrag bij hemzelf voorkomt

beantwoordt met dat hiervan nooit sprake is, dan wordt dat antwoord als onwaarschijnlijk opgevat.

Sociale wenselijkheid kan worden opgevat als een pure antwoordtendentie, maar sterk staat deze opvatting niet. Nunnally (1978, p. 558, p. 661) noemt drie componenten die naar alle waarschijnlijkheid deel uitmaken van sociale wenselijkheid: (a) de mate van actuele aanpassing van individuen, (b) de kennis van individuen over hun eigenschappen, en (c) de oprechtheid van individuen in hun zelfbeoordeling; met een sociaal wenselijk antwoord wil de respondent zich beter voordoen dan hij is.

De elf items van de schaal zijn:

Eerlijk ben ik ...	soms/altijd*
Ik beledig andere mensen..	nooit*/soms
Opscheppen doe ik ..	nooit*/wel eens
Ik ben tegen anderen	soms onvriendelijk/altijd vriendelijk*
Tegen volwassenen ben ik	wel eens brutaal/altijd beleefd*
Als ik een fout maak, geef ik die	altijd toe*/niet altijd toe
Iets minder goed te kunnen dan een ander vind ik	nooit*/soms vervelend
Een hekel aan iemand heb ik	bijna nooit*/wel een enkele keer/vaak
Meer krijgen dan een ander probeer ik	nooit*/wel eens
Mijn zin doordrijven doe ik	vaak/bijna nooit*
Iets kwaads vertellen over een ander doe ik	soms/zelden/nooit*

De met een * aangemerkte antwoorden worden als sociaal wenselijk aangemerkt. De uiteindelijke sociale wenselijkheids-score bestaat uit het totaal aantal sociaal wenselijke antwoorden op de 11 items. Tot slot zij opgemerkt dat de items tussen andere items van de MSI-s zijn opgenomen, in tegenstelling tot het onderzoek van Stams waarin de items gezamenlijk op een bladzijde van een instrument waren opgenomen.

2.4. Omschrijving van de onderzoeksgroep en van de selectieprocedure

Aan het onderzoek van het NWO project namen 32 scholen (zie Tabel 3) deel die alle binnen de Randstad zijn gevestigd. Voor deelname aan het onderzoek werden categoriale scholen voor voortgezet onderwijs (VBO, MAVO, VWO) en scholengemeenschappen (HAVO/VWO of MAVO/HAVO/VWO) geselecteerd. Sommige scholen voor VBO en MAVO maken deel uit van een groter scholengemeenschap, maar functioneren als tamelijk zelfstandige scholen, de locatie is in ieder geval apart van de rest van de scholengemeenschap.

Van elke school werden leerlingen uit het tweede en derde leerjaar geselecteerd, van scholen met HAVO/VWO bovendien HAVO/VWO-leerlingen uit het vierde leerjaar. Per leerjaar werden twee klassen (indien mogelijk aselekt) gekozen en uit elke klas acht leerlingen aselekt getrokken - bij voorkeur vier meisjes en vier jongens -in totaal 1536 leerlingen. Vooral op scholen voor VBO bleek het vaak onmogelijk om evenveel meisjes en

jongens uit een klas te trekken. Enkele scholen hebben een populatie van (bijna) alleen jongens of meisjes. Op de gemengde scholen komen vanaf het derde leerjaar veel klassen met alleen meisjes of jongens voor omdat de leerlingen dan hun vakrichting hebben gekozen.

HAVO en VWO worden in dit onderzoek als hetzelfde opleidingsniveau beschouwd, op scholen die beide opleidingen hebben zijn per leerjaar leerlingen uit één HAVO-klas en

Tabel 3: Verdeling van de steekproef van het NWO project naar Schooltype, school, leerjaar en klas

Schooltype	N scholen	N leerjaren	N klassen	N leerlingen
VBO	8	16	32	259
MAVO	8	16	32	257
HAVO en/of VWO	8	24	48	384
MAVO/HAVO/VWO	8	40	80	653
Totaal	32	96	192	1553

één VWO-klas getrokken.

Aan de validatiestudie namen van deze grote groep 120 derdeklassers deel (zie Tabel 4). Uit de beschikbare groep zijn de leerlingen per klas aselekt getrokken; 59 van de 120 leerlingen waren meisjes. Er waren 48 MAVO leerlingen, waarvan 24 op een

Tabel 4: Verdeling van de steekproef van de validiteitsstudie naar Schooltype, school, en klas.

Schooltype	N scholen	N klassen	N leerlingen
VBO	4	8	24
MAVO	4	8	24
HAVO en/of VWO	4	8	24
MAVO/HAVO/VWO	4	16	48
Totaal	16	40	120

scholengemeenschap zaten, en 48 HAVO/VWO leerlingen.

Eén leerling is verwijderd uit de steekproef omdat er onzekerheid was omtrent de juiste koppeling van gegevens. Bij de resterende 119 leerlingen waren 58 meisjes.

De leeftijd varieert van 13,8 jaar tot 17,3 jaar, 67 leerlingen zijn 14 jaar op het tijdstip van afname. 7 leerlingen zijn in het buitenland geboren, deze woonden van 5 tot 15 jaar in Nederland. Van 21 leerlingen is de vader in het buitenland geboren en van 18 leerlingen de moeder.

2.5. Afname

2.5.1. Afname procedure schriftelijke instrumenten

De schriftelijk vragenlijsten zijn op de 16 bij dit onderzoek betrokken scholen afgenomen in de periode van eind september tot medio november 1995. De afnames

werden verzorgd door vier stagiaires die vooraf op drie middagen een introductie en training hadden gevolgd. Daarnaast ontvingen de proefleidsters een schriftelijk draaiboek voor de afname. De vragenlijsten werden meestal per leerjaar afgenomen, dus groepsgewijs, met 16 leerlingen afkomstig uit twee klassen.

Behalve de onder 2.2.2. genoemde vragenlijsten en de SROM-SF werden ook een sociale wenselijkheidsschaal, een verbale intelligentietest en persoonlijke gegevens van de leerling afgenomen. De leerlingen kregen een vragenboekje met vijf bijbehorende antwoordbladen plus nog 32 direct in te vullen bladzijden (de twee dilemma's, de SROM-SF en de persoonlijke gegevens). Voor het invullen van dit omvangrijke pakket waren drie lesuren beschikbaar, eerst een blok van twee lesuren en later op dezelfde dag één lesuur. De vragenlijsten zijn in twee verschillende condities (volgordes) afgenomen, waarbij de SROM-sf, die de moeilijkste is, en het helpen-dilemma altijd aan het begin van een sessie werden afgenomen. De afnames zijn meestal volgens plan verlopen. Als de leerlingen tijd tekort kwamen waren ze meestal bereid om de resterende vragen in de pauze of de volgende lesuur te beantwoorden. Vooral voor VBO-leerlingen was de hoeveelheid vragen erg groot, van deze leerlingen is het veel gevraagd om op één dag drie à vier uur geconcentreerd met teksten te werken. Wij hebben de procedure voor deze leerlingen voor zover mogelijk aangepast zodat de afname in drie afzonderlijke lesuren, met mogelijkheid tot uitloop, plaatsvond. Ook werden de vragenlijsten vaak in groepjes van 8 i.p.v. 16 leerlingen afgenomen. Op enkele scholen konden de vragenlijsten niet op één dag worden afgenomen maar werd de afname verspreid over twee of drie dagen.

De SROM-SF en de persoonlijke gegevens zijn manueel ingevoerd, de resterende data zijn gescand met 'optical mark reading'.

2.5.2. Afnameprocedure mondelinge instrumenten

Bij de 120 leerlingen is de MSI-i (helpen- en diefstaldilemma) en de SRM-SF mondeling afgenomen. De interviews werden vastgelegd met behulp van een bandrecorder. Bij de helft van de leerlingen werd eerst de SRM-SF afgenomen, bij de andere helft is met dit instrument geëindigd. De leerlingen hadden al eerder de schriftelijke vragenlijsten beantwoord. De tijdsduur tussen de schriftelijke afname en het mondeling interview was minimaal 3 dagen. Voor de afname van de interviews waren twee personen beschikbaar. De afname duurde ongeveer 1 uur. De interviews zijn uitgetypt en vervolgens gescoord. In verband met de beschikbare middelen kon het diefstal-dilemma niet op tijd geheel worden uitgetypt en gescoord, dit wordt alsnog gedaan in een scriptie-onderzoek.

3. BESCHRIJVING RESULTATEN PER INSTRUMENT

3.1. Inleiding

De eerste stap in het validatieproces wordt gevormd door het beoordelen van de univariate eigenschappen van de geoperationaliseerde constructen en de eigenschappen van de schalen die deze operationalisaties constitueren. Centraal staat daarbij de stabiliteitsvraag: geven herhaalde metingen van een subject onder verschillende omstandigheden steeds dezelfde resultaten. Is deze stabiliteit aanwezig, dan spreken we van betrouwbare metingen. Daarnaast kan de betrouwbaarheid uitgedrukt worden in de mate van interne consistentie van de schaal waarmee het construct gemeten wordt. In dat geval wordt de mate van representativiteit van de afzonderlijke items uit de schaal beoordeeld. De interne consistentie geeft daarmee een indicatie voor de mate van generaliseerbaarheid van het construct naar een andere verzameling van items uit hetzelfde domein.

Naast de betrouwbaarheid van de constructen is van belang of de constructen voldoende discrimineren tussen de subjecten uit de onderzoekspopulatie, of het instrument van een voldoende grote groep van respondenten een meting geeft en of de metingen binnen de grenzen van redelijkheid liggen, zoals de in de eerdere hoofdstukken beschreven theorie die stellen. Samen met de inhoud van de afzonderlijke items vormen deze gegevens een eerste aanwijzing voor de inhouds-validiteit van een construct. Bovendien vormen zij een noodzakelijke conditie voor een succesvolle afronding van de overige stappen in het validatieproces.

Gegeven het ontwerp van dit onderzoek kan aan de hand van de huidige meetresultaten alleen betrouwbaarheid in de vorm van interne consistentie en intercodeurovereenstemming bepaald worden. Ten behoeve van het hoofdonderzoek wordt in een aparte studie de test-hertest betrouwbaarheid van het schriftelijk Moreel Schoonklimaat Instrument onderzocht. Daarnaast is in een eerdere studie de betrouwbaarheid en validiteit van de vertalingen van de SRM-SF en SROM-SF onderzocht (Stams, 1994).

In dit hoofdstuk wordt een descriptie gegeven van de metingen per instrument, die het mogelijk maakt bovenstaande eerste stap in het validatie proces te maken. De descriptie omvat een beschrijving van de omvang en aard van de missing data en van de distributieve eigenschappen van de metingen, waaronder het vóórkomen van zogenaamde 'outliers'. Daarnaast zal middels een item-analyse worden ingegaan op de interne consistentie van de verschillende schalen die aan de constructen ten grondslag liggen. De

resultaten zullen worden vergeleken met eerdere resultaten, voor zover zij voorhanden zijn.

3.2. Resultaten Moreel Schoolklimaat Instrument - Interview (MSI-i)

3.2.1. Uitval

Tijdens het MSI-Interview wordt de respondent naar aanleiding van een dilemma steeds voor een serie gedrags- of morele keuzen gesteld (*Inhoud van de norm*), waarbij een argumentatie voor iedere keuze moet worden gegeven, die door de interviewer volgens een protocol worden gescoord (*Stadium van de norm*). De verschillen tussen de morele keuzen die gemaakt moeten worden dienen om een grotere diversiteit in de argumentatie te bewerkstelligen. Uitgangspunt is dat zoveel mogelijk verschillende argumenten aan de respondent ontlokt worden. Daarbij is het goed mogelijk dat de respondent niet met iedere nuancering raad weet. Deze 'probing' techniek heeft dan ook tot gevolg dat lang niet elke respondent op iedere vraag een scoorbare argumentatie heeft gegeven: het aantal missings op item niveau is hoog.

In Tabel 5 en 6 staat het aantal scoorbare argumenten per item. Bij ieder item konden zowel redenen voor een prosociale als een asociale morele keuze worden gegeven. Indien er zowel voor de prosociale- als asociale keuze argumenten werden gegeven, is één van de twee geselecteerd voor verdere verwerking. In eerste instantie is dit gebeurd overeenkomstig de prosociale of de asociale richting van de keuze. Was deze keuze onduidelijk, dan is indien mogelijk gekozen voor de prosociale argumentatie, omdat prosociale argumenten in de praktijk een grotere differentiatie in het *Stadium van de norm* te zien geven. De laatste kolom van beide tabellen bevat de aantallen op deze wijze geselecteerde argumenten.

Tabel 5: Aantal scoorbare norm antwoorden MSI-i voor Zelf (n=119)

Item	inhoud van de norm	stadium van prosociale norm	stadium van asociale norm	stadium van de norm
3. DOEN	119	93	11	102
6. AFKEURING	112	50	4	53
8. VERANTW.	117	3	60	63
11. DOEN	107	26	47	73
14. AFKEURING	111	20	10	30
17. DOEN	109	16	13	29
20. AFKEURING	107	19	2	21
23. LERAAR	117	34	1	35
26. AFKEURING	111	20		20
28. DOEN	118	42	19	61

Tabel 6: Aantal scoorbare norm antwoorden MSI-i voor Ander (n=119)

item	inhoud van de norm	stadium van prosociale norm	stadium van asociale norm	stadium van de norm
4. DOEN	115	60	67	99
5. PRESC.	109	53		53
7. AFKEURING	99	25	14	34
9. VERANTW.	91	4	52	55
12. DOEN	115	24	68	82
13. PRESC.	108	29	7	36
15. AFKEURING	106	11	11	22
18. DOEN	114	23	69	81
19. PRESC.	101	20	1	20
21. AFKEURING	106	11	8	17
24. LERAAR	111	30	3	33
25. PRESC.	106	23		23
27. AFKEURING	109	11	2	13
29. DOEN LLN	114	26	41	54
30. PRESC.SCH.	109	23	3	26

We merken op dat :

- Het aantal antwoorden op de gedragskeuzen (Inhoud van de norm) steeds voldoende is.
- Er veel minder argumenten dan morele keuzen zijn gescoord.
- Het aantal gescoorde argumentaties per vraag sterk wisselt.
- Het aantal gescoorde argumentaties bij de verantwoording van de primaire morele keuze (wat zou je / zouden je klasgenoten DOEN) het hoogst is.
- Het aantal gescoorde argumentaties na een DOEN vraag afneemt.

Het wisselende aantal argumentaties bij de verschillende items zou er op kunnen wijzen dat de verschillende soorten morele keuzen niet uit eenzelfde universum afkomstig zijn. Of dit een consistente meting van het stadium van de norm in de weg staat is niet op voorhand duidelijk. Enerzijds kan - analoog aan het construct *Morele competentie* - verondersteld worden dat hoewel de vragen inhoudelijk kunnen verschillen de argumentatie toch steeds tot hetzelfde stadium gerekend zou kunnen worden. Anderzijds geeft een dergelijke spreiding te denken over de mate van generaliseerbaarheid naar een andere set items. In het bijzonder staat de vergelijkbaarheid met het schriftelijk instrument hierdoor ter discussie, aangezien in het schriftelijk instrument niet steeds de argumentatie van voor de niet-primaire morele keuzen wordt gevraagd. We zullen in een volgende paragraaf de mate van samenhang tussen de verschillende items nader bestuderen.

Dat het aantal missings op item-niveau hoog is, verontrust ons op zich niet, gegeven de opzet van het instrument. Uitgangspunt bij de constructie van het MSI-Interview was

dat zowel voor de *Zelf*- als voor de *Ander*-meting van *Stadium van de norm* tenminste 7 scorebare argumentaties zouden voorkomen, waarna een construct bepaald zou kunnen worden door het gemiddelde stadium over de gescoorde argumentaties te berekenen (vergelijk Gibbs, Widaman & Colby, 1982). In Tabel 7 zien we een verdeling van het aantal scorebare argumentaties.

Opgemerkt moet worden dat de norm die Gibbs hanteert voor de SRM door ons was vastgesteld voor het hele MSI-interview, terwijl in het onderhavige onderzoek slechts de gegevens van het *Helpen* - dilemma zijn verwerkt. We constateren dan ook dat de norm van tenminste 7 scorebare argumentaties voor het dilemma leidt tot een uitval van respectievelijk 95% en 68% van alle respondenten. Besloten is de norm voorlopig te bepalen op *tenminste* 2 scorebare argumentaties.

Tabel 7: Verdeling aantallen valide metingen MSI-i

Aantal valide items	inhoud van de norm		stadium van de norm	
	Zelf	Ander	Zelf	Ander
0			2	
1		1	2	2
2	1		21	12
3	2	5	20	14
4	27	22	29	19
5	89	91	20	15
6			20	19
7			2	13
8			3	14
9			1	7
10				2
11				3
Maximaal	5	5	9	11

Hoewel de hoge uitval op item-niveau niet verontrustend is, is de systematiek in de uitval wél een probleem. Naast het oproepen van een vraag naar de representativiteit van de items kan zij ook een niet te verwaarlozen invloed betekenen op de aggregatie van het construct *Stadium van de norm*. Vast staat wel dat er in dit verband geen directe relatie met enkele achtergrond variabelen bestaat: het aantal missings op item niveau bleek niet significant afhankelijk van sexe, verbale intelligentie en etniciteit. Vooralsnog is dan ook een verklaring voor het voorkomen van missings te vinden in een vorm van respondent-vermoeidheid: we constateerden eerder dat het aantal missings per subvraag bij iedere situatie toeneemt. Op grond van de hoge uitval en de onevenwichtige verdeling van de uitval over de verschillende items is besloten tot aggregatie van de *Stadium van de norm* scores d.m.v. een dubbele vorm van middelen: eerst wordt per context gemiddeld over alle beschikbare argumenten, daarna wordt als geaggregeerde score het gemiddelde over deze context gemiddelden berekend. De procedure beoogt het evenwicht tussen de verschillende contexten te bewaren.

Van een hoge uitval op item-niveau is geen sprake bij het construct *Inhoud van de norm*. In Tabel 7 kunnen we constateren dat een norm van tenminste 3 valide items voor het construct *inhoud van de norm* tot een uitval van één respondent bij de *Zelf*-meting, en één respondent bij de *Ander*-meting leidt.

Samenvattend kunnen we stellen dat de uitval op het construct *Inhoud van de norm* bevredigend laag is, evenals de uitval op geaggregeerde construct *Stadium van de norm*. Het laatste construct kent echter een dermate grote en onregelmatige uitval op item-niveau, dat vraagtekens gesteld moeten worden bij de vergelijkbaarheid van de scores tussen respondenten onderling en tussen het schriftelijk en het mondeling instrument.

3.2.2. Descriptie op item niveau.

In Tabel 8 en Tabel 9 vinden we voor de verschillende contexten uit het 'helpen dilemma' de verdeling van de antwoorden op de primaire handelings vraag: 'Zou jij/zouden je klasgenoten Bert helpen?' Het betreft hier de *Zelf* - en de *Ander* - meting van het construct *Inhoud van de norm*. We zien dat 'als de leraar het vraagt' Bert het meeste geholpen zal worden en wanneer het 'tijd kost' Bert het minste geholpen zal worden. Voorts merken we op dat de perceptie van de bereidheid van de klasgenoten om te helpen systematisch lager ligt dan de zelf-perceptie. De volgorde in mate van hulp-bereidheid bij de diverse items is echter dezelfde.

Tabel 8: Verdeling van antwoorden per item, *Inhoud van de norm - Zelf*

Handelings keuze	Bert helpen		als het tijd kost		als gevolg is gepest worden		als leraar het vraagt		als het geen klasgenoot is	
helpen	86	72.3%	38	31.9%	80	67.2%	95	79.8%	71	59.7%
onder voorwaarden	24	20.2%	25	21.0%	11	9.2%	19	16.0%	15	12.6%
niet helpen	8	6.7%	44	37.0%	13	10.9%	3	2.5%	29	24.4%
uitval	1	.8%	12	10.1%	15	12.6%	2	1.7%	4	3.4%

Tabel 9: Verdeling van antwoorden per item, *Inhoud van de norm - Ander*

Handelings keuze	Bert helpen		als het tijd kost		als gevolg is gepest worden		als leraar het vraagt		als het geen klasgenoot is	
bijna iedereen helpt	55	46.2%	16	13.4%	24	20.2%	92	77.3%	39	32.8%
de helft helpt	15	12.6%	25	21.0%	19	16.0%	7	5.9%	41	34.5%
sommigen helpen	39	32.8%	45	37.8%	58	48.7%	10	8.4%	10	8.4%
niemand helpt	4	3.4%	26	21.8%	13	10.9%	1	.8%	2	1.7%
uitval	6	5.0%	7	5.9%	5	4.2%	9	7.6%	18	15.1%

Tabel 10: Gescoorde stadia per context, stadium van de norm - Zelf (Multiple respons)

Niveau van argumentatie	Bert helpen		als het tijd kost		als gevolg is gepest worden		als leraar het vraagt		als het geen klasgenoot is	
	n	%	n	%	n	%	n	%	n	%
stadium 2	60	38.7%	61	59.2%	7	14.0%	6	10.9%	37	60.7%
stadium 2.5	4	2.6%	1	1.0%	14	28.0%	6	10.9%	1	1.6%
stadium 2.75	64	41.3%	34	33.0%	20	40.0%	31	56.4%	19	31.1%
stadium 3	25	16.1%	6	5.8%	9	18.0%	12	21.8%	4	6.6%
stadium 4	2	1.3%	1	1.0%	0	0.0%	0	0.0%	0	0.0%
n	108		84		42		42		61	

Tabel 11: Gescoorde stadia per context, Stadium van de norm - Ander (Multiple respons)

Niveau van argumentatie	Bert helpen		als het tijd kost		als gevolg is gepest worden		als leraar het vraagt		als het geen klasgenoot is	
	n	%	n	%	n	%	n	%	n	%
stadium 1	0	0.0%	0	0.0%	0	0.0%	18	26.1%	0	0.0%
stadium 2	78	41.9%	90	64.3%	21	17.8%	6	8.7%	55	68.8%
stadium 2.5	19	10.2%	1	.7%	60	50.8%	1	1.4%	0	0.0%
stadium 2.75	46	24.7%	33	23.6%	27	22.9%	28	40.6%	18	22.5%
stadium 3	41	22.0%	13	9.3%	9	7.6%	15	21.7%	6	7.5%
stadium 4	2	1.1%	3	2.1%	1	.8%	1	1.4%	1	1.3%
n	109		97		88		50		71	

In de Tabellen 10 en 11 staan 'multiple respons' frequentie verdelingen van de gescoorde *stadia van de norm* per item weergegeven: d.w.z. per context worden alle gescoorde stadia weergegeven over alle subvragen, die bij die context horen. Onderaan iedere kolom staan de aantallen respondenten die voor de betrokken context één of meer argumentaties hebben gegeven. Wanneer we stadium 2 en stadium 2.5 samen nemen zijn de context verschillen gering. Opvallend is dat bij de context 'als de leraar het vraagt' 26% stadium 1 antwoorden gegeven worden (Tabel 11), een stadium dat verder nergens gescoord is. Het betreft hier de argumentatie dat er een vorm van repercussie zal volgen

Tabel 12: Item totaal correlaties *Inhoud van de norm* en *Stadium van de norm*

Hulp context	inhoud vd norm		stadium vd norm			
	Zelf (n=89)	Ander(n=91)	Zelf	n	Ander	n
Bert helpen	.38	.57	.05	108	.30	109
als het tijd kost	.54	.61	.15	84	.31	97
als gevolg is gepest worden	.38	.47	.17	42	.27	88
als leraar het vraagt	.48	.40	.10	42	-.15	50
als het geen klasgenoot is	.45	.25	.11	61	.20	71

indien de wens van de leraar niet gevolgd wordt.

Ten aanzien van de representativiteit van de verschillende contexten constateren we dat bij het construct *Inhoud van de norm* alle items redelijk tot goed representatief genoemd kunnen worden, met als zwakste item 'als het geen klasgenoot is' bij de andere-meting (Tabel 12). De representativiteit van de *Stadium van de Norm*- items kon niet met een standaard item-analyse onderzocht worden als gevolg van het grote aantal ontbrekende waarden. In Tabel 12 staan gecorrigeerde item-totaal correlaties vermeld die

zijn berekend over alle niet-ontbrekende waarnemingen, d.w.z. met per item een wisselend aantal waarnemingen. De item-totaal correlaties zijn overwegend zwak tot zeer zwak te noemen. De item-totaal correlatie voor het item 'als de leraar het vraagt' is zelfs negatief. Gegeven het grote aantal ontbrekende waarnemingen en de daardoor sterk wisselende samenstelling van de inhoud van de argumentatie - steeds bestaat de argumentatie per context uit antwoorden op een verschillende verzameling van subvragen - is dit resultaat niet onverwacht. De constructie van het MSI-interview behoeft dan ook enige aanpassing.

3.2.3. Schaalgegevens

De schaalgegevens voor de constructen *Inhoud van de norm* en *Stadium van de norm* staan weergegeven in Tabel 13.

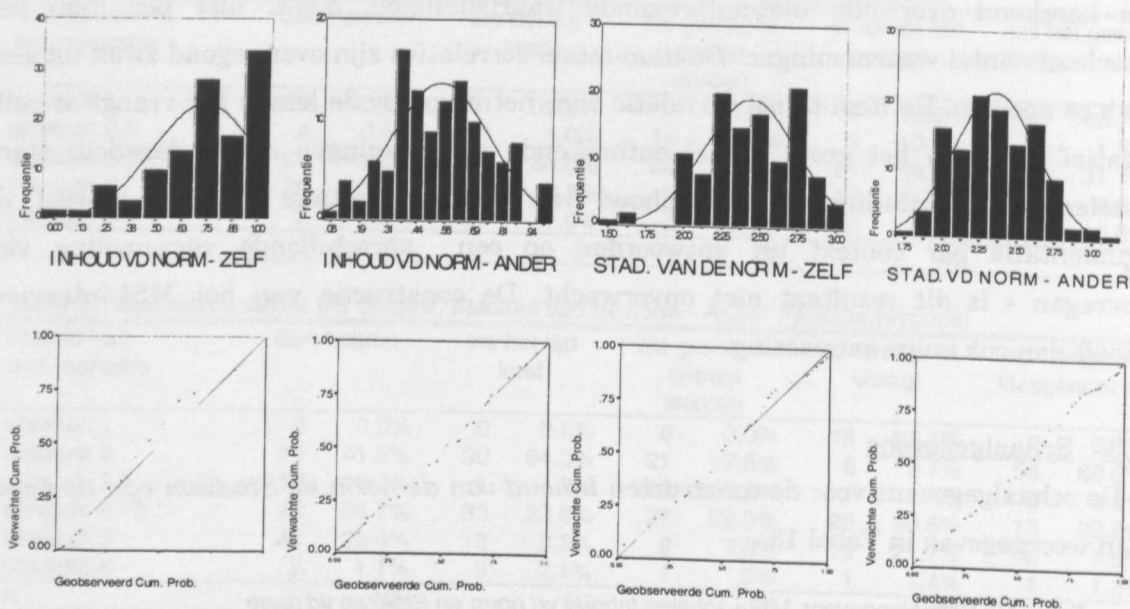
Tabel 13: Schaalgegevens MSI-i schalen *Inhoud vd norm* en *Stadium vd norm*

statistiek	inhoud vd norm		stadium vd norm	
	Zelf	Ander	Zelf	Ander
gemiddelde	.74	.54	2.44	2.35
standaard deviatie	0.25	0.21	0.31	0.28
skewness	-0.92	0.11	-0.33	0.25
kurtosis	0.22	-0.48	-0.13	-0.47
Cronbach's α	0.67	0.70	?	?
n	118	118	117	119

We zien dat het gemiddelde van de *Zelf*-meting van *Inhoud van de Norm* aanzienlijk verschilt van de *Ander* - meting: De zelfperceptie te zullen helpen ligt aanmerkelijk hoger dan de perceptie dat klasgenoten zullen helpen. Het verschil in gemiddeld *Stadium van de norm* is echter marginaal. (Beide verschillen zijn significant op 5% niveau.)

De standaarddeviatie van de *Stadium van de norm* schalen is relatief gering. Hierin zou deels de verklaring kunnen liggen voor de geringe consistentie van de schalen, zoals die uit de item-totaal blijkt (alpha kon bij gebrek aan valide waarnemingen niet berekend worden, zie vorige paragraaf). De interne consistentie van het construct *Inhoud van de Norm* is voor beide metingen als voldoende te kenschetsen.

De *Zelf* - meting van *Inhoud van de norm* is sterk scheef naar rechts verdeeld (Skewness -0.91). De verdeling van de *Ander* - metingen van beide constructen zijn te plat (Kurtosis +/- 0.48). In de figuren 1 t/m 4 is een en ander visueel te inspecteren



Figuur 1 t/m 4: Histogrammen en normal-probability-plots van *Inhoud van de norm* en *Stadium van de norm*

De aanname dat de *Zelf - Inhoud* meting normaal verdeeld is moeten we op grond van de *Kolmogorov-Smirnov toets* verwerpen ($p < 0.01$). Voor deze schaal kon geen getransformeerde schaal berekend worden die de K-S toets wel doorstaat. De schalen kennen daarnaast slechts in één geval - nl. de *Inhoud - Ander* meting - een univariate uitbijter.

Samenvattend stellen we dat de mate van uitval op *Inhoud van de norm* acceptabel is. De representativiteit van de items en de interne consistentie zijn eveneens bevredigend. De uitval op item-niveau op *Stadium van de norm* is echter te groot. Bovendien is de representativiteit van de verschillende items c.q. de interne consistentie zeer gering. Daarnaast discrimineert het construct in de onderzochte populatie slecht. De constructie van de MSI-i *Stadium van de norm* constructen behoeft dan ook enige aanpassing, enerzijds om de mate van uitval terug te dringen, anderzijds om de variantie van de meting te vergroten. De univariate verdelings-eigenschappen van de constructen zijn bevredigend, uitgezonderd het construct *Inhoud van de norm-Zelf* dat een scheve, niet-normale verdeling kent. Een getransformeerde schaal met een normale verdeling kon niet geconstrueerd worden.

3.3. Moreel Schoolklimaat Instrument - Schriftelijk (MSI-s)

3.3.1. Uitval

De uitval voor de metingen van *Inhoud van de norm* en *Stadium van de norm* beschouwen we aan de hand van Tabel 14. In de tabel staan de verdelingen van het aantal scoorbare items per respondent voor de verschillende *norm* - constructen weergegeven.

Tabel 14: Aantallen valide MSI-s *norm* - scores (n=119)

aantal valide items	inhoud van de norm		stadium van de norm			
	Zelf	Ander	Close	Closest	Close	Closest
1			4	1	5	1
2			5	2	6	2
3			9	0	14	0
4	3	4	24	3	40	7
5	116	115	77	113	54	109

Er is geen uitval op *Inhoud van de norm* bij een minimum van 4 valide items. De situatie voor *Stadium van de norm* is minder gunstig: toepassen van dezelfde norm zou tot behoorlijke uitval leiden. Opvalt dat de uitval op de Close scores duidelijk hoger is dan op de Closest scores. Dit is het gevolg van het toepassen van een fout-protocol dat oorspronkelijk is ontworpen voor gebruik bij de SROM-SF test. Volgens deze procedure zijn alle Close items ongeldig waarbij de redenering ontbreekt die in het bijbehorende Closest item wel is gegeven. Dit leidt tot gemiddeld tussen de 10% en de 15% afgekeurde antwoorden bij de verschillende items. Volgens vergelijkbare normen als die bij gebruik van de SROM-SF worden voorgeschreven zouden bovendien 9 respondenten bij de *Zelf*-meting en 11 respondenten bij de *Ander*-meting uitvallen. Om geen al te grote vermindering van het voor de validiteitsanalyse beschikbare aantal respondenten te bewerkstelligen is besloten alle respondenten te handhaven voor deze analyses.

Naast metingen van de constructen *Inhoud van de norm* en *Stadium van de norm* worden in de MSI-S ook twee metingen van het construct *Stadium van gemeenschapsgevoel* gedaan, één ten aanzien van de respondent zelf en één ten aanzien van haar perceptie van haar

Tabel 15: Aantal scoorbare items stadium van gemeenschapsgevoel (SG)

Aantal Scores	Zelf	Ander
19	1	1
20	5	9
21	113	109

klasgenoten. In Tabel 15 zien we de uitval voor deze metingen weergegeven in de vorm van het aantal wél scoorbare items op deze twee schalen, die ieder 21 items bevatten met een 5 punts-schaal (helemaal waar <-> helemaal niet waar). We zien dat een gering aantal respondenten geen score voor 1 of 2 items heeft. Dit betekent dat voor elke respondent

voldoende items zijn om de twee schalen te berekenen: er is geen uitval op deze twee schalen.

Samenvattend beschouwen we de mate van uitval - gegeven de gemaakte keuzen - op alle constructen als bevredigend gering.

3.3.2. Descriptie op item niveau

Het construct *Inhoud van de norm* wordt in MSI-s gelijk het mondelinge instrument gemeten over de 5 items rond het helpen dilemma. In Tabel 16 en Tabel 17 staan de handelingskeuzen (wel of niet helpen) weergegeven die respectievelijk de respondenten

Tabel 16: Verdeling van antwoorden *Inhoud vd norm - ZELF*

Handelings keuze	Bert helpen		als het tijd kost		als gevolg is gepest worden		als leraar het vraagt		als het geen klasgenoot is	
helpen	39	32.8%	32	26.9%	39	32.8%	70	58.8%	47	39.8%
onder voorwaarden	64	53.8%	53	44.5%	67	56.3%	43	36.1%	49	41.5%
niet helpen	16	13.4%	34	28.6%	13	10.9%	6	5.0%	20	16.9%
uitval									2	1.7%

Tabel 17: Verdeling van antwoorden *Inhoud vd norm - ANDER*

Handelings keuze	Bert helpen		als het tijd kost		als gevolg is gepest worden		als leraar het vraagt		als het geen klasgenoot is	
bijna iedereen helpt	6	5.0%	2	1.7%	5	4.2%	23	19.3%	4	3.4%
de helft helpt	22	18.5%	10	8.4%	9	7.6%	28	23.5%	30	25.4%
sommigen helpen	81	68.1%	80	67.2%	101	84.9%	65	54.6%	78	66.1%
niemand helpt	10	8.4%	27	22.7%	3	2.5%	3	2.5%	4	3.4%
uitval					1	.8%			2	1.7%

zelf zouden maken of die hun klasgenoten - in hun perceptie - zouden maken.

We zien hier evenals bij het mondeling instrument dat 'als de leraar het vraagt' Bert het meeste geholpen zal worden. Bij de overige contexten is de hulpbereidheid bij de *Zelf* - meting ongeveer even groot. De *Ander* - meting geeft iets meer differentiatie over de verschillende contexten. Merk op dat 'als het geen klasgenoot maar een gewone leerling is' bij dit instrument relatief prosocialer scoort voor de *Zelf* - meting dan bij het mondeling instrument.

In Tabel 18 staan de item-totaal correlaties van de *Inhoud van de norm* items weergegeven. Voor zowel de *Zelf* - als de *Ander* - meting zien we dat het helpen van een leerling die geen klasgenoot is,

Tabel 18: Item totaal correlaties voor *Inhoud van de norm*

Helpen context	Zelf	Ander
Bert helpen	.48	.27
als het tijd kost	.42	.45
als gevolg is gepest worden	.45	.46
als leraar het vraagt	.51	.26
als het geen klasgenoot is	.28	.22
	n=116	n=115

minder representatief voor de schaal genoemd kan worden. Uit de item-analyse blijkt

daarnaast dat de *Ander* - schaal gedomineerd wordt door de twee items die het kosten-aspect meten, de overige items hebben een zwakke item-totaal correlatie. Gegeven het geringe aantal items en de wens de verschillende vormen van het *Inhoud van de norm* construct zoveel mogelijk vergelijkbaar te houden is er voor gekozen alle items te

Tabel 19: Verdeling stadia voor *Stadium van de norm*

Stadium	Zelf				Ander			
	Close		Closest		Close		Closest	
1	175	12.3%	31	5.3%	251	17.3%	76	13.1%
2	501	35.2%	321	55.2%	489	33.7%	281	48.6%
3	374	26.3%	132	22.7%	362	24.9%	128	22.1%
4	373	26.2%	98	16.8%	351	24.2%	93	16.1%

handhaven.

De *Stadium van de norm* schalen bestonden uit twee soorten items, te weten items waarbij moest worden aangegeven van iedere reden of hij overeenkwam met een reden die ze zelf zouden geven (*Close*-schaal), en items waarbij de meest overkomende reden moest worden aangegeven (*Closest* - schaal). In Tabel 19 zien we dat het meeste stadium 2 redenen zijn aangegeven, waarna in aflopende volgorde stadium 3, stadium 4 en stadium 1 volgen. We merken daarnaast op dat de dominantie van stadium 2 redenen het sterkst naar voren komt bij de *Closest*-schalen. Stadium 2 redenen worden dus niet alleen het meest als overeenkomend met de argumentatie van de respondenten aangegeven, maar ook als *de meest overeenkomende* reden.

In Tabel 20 vinden we de item-totaal correlaties tussen de scores op de afzonderlijke items en de geaggregeerde *Stadium van de norm* schalen. We zien dat de twee 'helpen' items uit de *Zelf*-meting een geringe item-totaal-correlatie hebben op respectievelijk de *Close* en de

Tabel 20: Item-totaal correlaties *Stadium vd norm*

Item	Zelf		Ander	
	Close	Closest	Close	Closest
Stelen	.53	.28	.49	.41
Ingrijpen	.48	.49	.49	.54
Klikken	.37	.49	.44	.33
Helpen	.13	.33	.39	.39
Helpen	.49	.15	.43	.33
	n=42		n=54	

Closest schaal. De overige items kunnen we als voldoende representatief beschouwen voor de verschillende metingen van het construct .

Tabel 21: Factoren *Stadium van gemeenschapsgevoel (Zelf)* n=119

	SG1	SG2	SG3	SG4
helemaal niet waar	42.0%	31.0%	18.2%	19.0%
een beetje niet waar	17.5%	17.7%	19.6%	20.5%
niet waar en niet onwaar	17.2%	23.0%	26.9%	24.5%
een beetje waar	14.2%	17.8%	21.5%	18.7%
helemaal waar	9.1%	10.4%	13.3%	17.3%

Tabel 22: Factoren *Stadium van gemeenschapsgevoel (Ander)* n=119

	SG1	SG2	SG3	SG4
helemaal niet waar	27.7%	16.0%	16.2%	15.9%
een beetje niet waar	23.6%	22.1%	22.4%	19.1%
niet waar en niet onwaar	26.9%	32.2%	31.7%	33.3%
een beetje waar	14.4%	19.4%	20.6%	20.9%
helemaal waar	7.4%	10.3%	8.4%	10.8%

In Tabel 21 en Tabel 22 vinden we een beschrijving van verdeling van scores over de 5 punten van de 21 items voor de schalen van *Stadium van gemeenschapsgevoel* (SG), onderverdeeld in de 4 factoren: Machtsgebruik (SG1), Concrete Wederkerigheid (SG2), Relaties/Vriendschap (SG3) en Sociaal contract (SG4). In het algemeen kunnen we opmerken dat de 5 punten alle een behoorlijk percentage responses oproepen. Extreme vormen van respondent bias kunnen we niet waarnemen.

Om te komen tot een geaggregeerde score voor het *Stadium van gemeenschapsgevoel* ligt het voor de hand het gewogen gemiddelde stadium te berekenen, d.w.z. de som van de producten van subschaal scores en hun respectievelijk stadia (1 t/m 4). Een item-analyse leert echter dat een dergelijke schaal een relatief geringe interne consistentie heeft gezien het betrekkelijk grote aantal items (respectievelijk 0.63 voor de *Zelf*-meting en 0.53 voor de *Ander*-meting, 21 items). Bovendien blijkt bij beschouwing van de item-totaal correlaties dat de items van de SG1 en de SG2 niet - of zelfs negatief correleren met de totale schaal.

Tabel 23 Samenhang subschalen
SG-Zelf, interne consistentie

	SG1	SG2	SG3	SG4
SG1	1.00	.35	-.13	-.02
SG2		1.00	-.23	-.24
SG3			1.00	.50
SG4				1.00
alpha	0.23	0.61	0.70	0.68

Tabel 24 Samenhang subschalen
SG-Ander, interne consistentie

	SG1	SG2	SG3	SG4
SG1	1.00	.25	-.10	.01
SG2		1.00	-.25	-.29
SG3			1.00	.45
SG4				1.00
alpha	0.34	0.68	0.64	0.58

Wanneer we de samenhang tussen de onderlinge schalen bestuderen (zie Tabel 23 en Tabel 24) dan zien we dat de SG1 en de SG2 schaal, als mede de SG3 en de SG4 schaal onderling zwak correleren, maar met elkaar niet of zwak negatief. Gegeven deze resultaten lijkt de constructie van een totale schaal voor het *Stadium van*

gemeenschapsgevoel niet te verdedigen. Daarentegen is het hanteren van maar 4 verschillende schalen voor zowel de *Zelf* als de *Ander* - meting evenmin aantrekkelijk, hoewel de matig tot voldoende hoge interne consistentie van de hoogste 3 stadia dit niet uitsluit. Besloten is het construct onder te verdelen in twee componenten, enerzijds een samenvoeging van schaal 1 en schaal 2 tot een schaal die we in het vervolg *Persoonlijk rendement* zullen noemen, anderzijds een samenvoeging van schaal 3 en 4 tot een schaal die we in het vervolg met *Sociale relaties* zullen aanduiden. De items van de verschillende schalen tellen voor de combinatie schalen alle even zwaar mee, d.w.z. het absolute stadium niveau is daarbij verlaten.

In de modellen in het volgende hoofdstuk worden de relaties tussen het *Stadium van de norm* enerzijds en deze twee samengestelde *Stadium van gemeenschapsgevoel* schalen anderzijds zichtbaar. Aanvullende, in dit rapport niet beschreven, analyses waarin de relaties tussen *Stadium van de norm* en alle afzonderlijke *Stadium van gemeenschapsgevoel* werden onderzocht, gaven geen wezenlijk andere resultaten, zodat we concluderen dat voor ons onderzoeksdoel geen opmerkelijk informatie-verlies optreedt

Tabel 25 Item totaal correlaties
Persoonlijk rendement

	item	Zelf	Ander
SG1	6	.31	.29
	15	.17	.16
	24	.21	.21
	31	-.02	.07
	38	.28	.18
SG2	7	.28	.19
	16	.20	.28
	25	.41	.53
	26	.46	.42
	32	.39	.49
	39	.32	.35

Zie Bijlage C voor de item omschrijvingen.

Tabel 26: Item totaal correlaties
Sociale relaties

	item	Zelf	Ander
SG3	8	.56	.52
	17	.36	.25
	33	.52	.44
	40	.51	.40
	44	.34	.33
SG4	9	.39	.26
	18	.49	.33
	27	.44	.42
	34	.51	.42
	41	.38	.30

wanneer de schalen zoals bovenvermeld worden samengenomen.

De representativiteit van de items van de SG2, SG3 en SG4 schaal is voldoende (Tabel 25 en 26). De representativiteit van de SG1 items varieert van zeer zwak tot matig, waarbij het vierde item (nr. 31) er in negatieve zin uitspringt. Gegeven de eerdere resultaten is dit niet verwonderlijk. Wellicht dat dit samenhangt met de geringe mate waarin respondenten stadium-1 norm argumenten kozen (zie Tabel 19 en 21): het zou kunnen dat de respondenten zich onvoldoende herkennen in stadium-1 uitspraken.

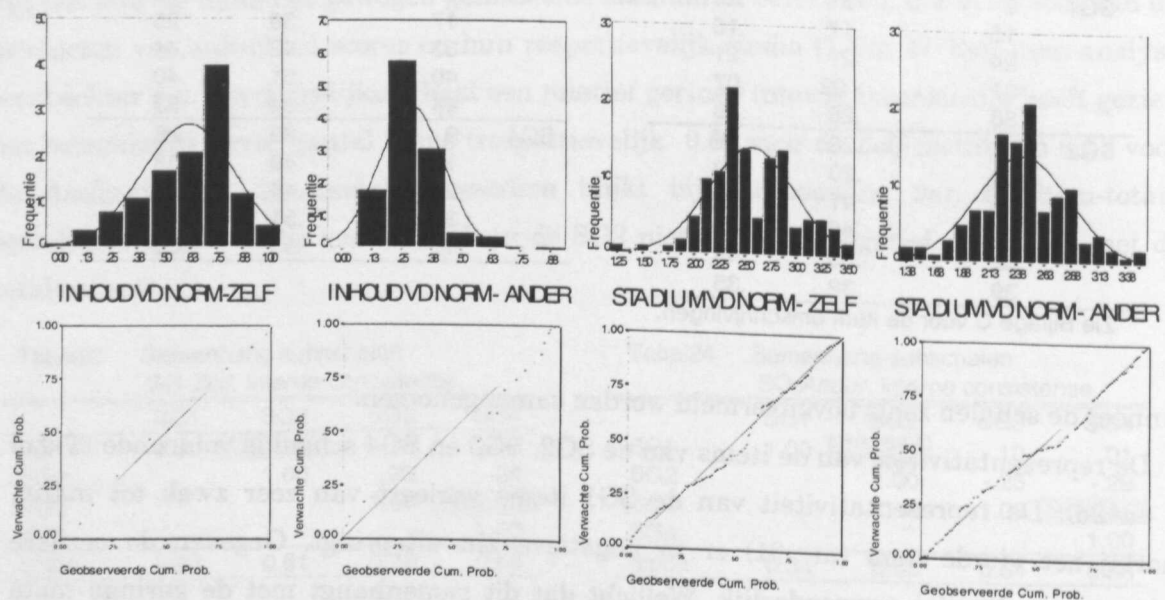
3.3.3. Schaalgegevens

De schaalgegevens van *Inhoud van de norm* en *Stadium van de norm* schalen staan weergegeven in Tabel 27. De geaggregeerde *Inhoud van de norm* scores zijn berekend door alle antwoord categorieën te coderen van 0 (niet helpen/niemand helpt) tot 1 (wel helpen/iedereen helpt). Uit de tabel blijkt dat de respondenten beduidend meer denken zelf te zullen helpen, dan ze verwachten dat hun klasgenoten dat zullen doen. Het gemiddelde stadium van de argumentatie is echter voor beide metingen vrijwel hetzelfde, tussen stadium 2 en 3 in.

Tabel 27: Schaalgegevens *Inhoud vd norm* en *Stadium vd norm*

statistiek	Inhoud vd norm		Stadium vd norm	
	Zelf	Ander	Zelf	Ander
gemiddelde	0.62	0.29	2.56	2.43
standaard deviatie	0.22	0.14	0.42	0.42
skewness	-0.55	1.61	0.04	0.02
kurtosis	0.05	3.90	0.19	0.37
Cronbach's α	0.67	0.54	0.68	0.72

De interne consistentie van de *Ander* - meting van *Inhoud van de norm* is matig, die van de overige schalen voldoende. In de vorige paragraaf bleken op één na alle items voldoende representatief.



Figuren 5 t/m 8: Histogrammen en normal-probability-plots van *Inhoud van de norm*, *Stadium van de norm*

Daarnaast zien we dat de *Stadium van de norm* schalen een bevredigende verdelingskarakteristiek hebben in termen van *skewness* en *kurtosis*. De *Inhoud van de norm* schalen zijn echter sterk scheef verdeeld, de *Zelf*-meting naar rechts, de *Ander* -

meting naar links. In de figuren 5 t/m 8 kunnen we dat ook visueel constateren. Volgens de Kolmogorov-Smirnov toets kunnen we de aanname dat de *Inhoud van de norm Zelf*-meting normaal verdeeld is accepteren en moeten we deze aanname voor de *~Ander*-meting verwerpen (respectievelijk $p > 0.20$, $p < 0.01$). De *Zelf*-meting van *Stadium van de norm* doorstaat de KS toets evenmin op 5% niveau, maar wel op 2.5% niveau, de *Ander*-meting doorstaat de toets wel ($p = 0.03$, $p = 0.19$). Dit resultaat voor de *Zelf*-meting wordt niet ondersteund door een afwijkende *skewness* of *kurtosis* zodat we er geen consequenties aan verbinden. De *Zelf*-meting van het construct *Stadium van de norm* kent 1 univariate uitbijter met een standaard score van -3.14. De *Ander*-meting van *Inhoud van de norm* kent twee univariate uitbijters met een standaard score groter dan 3.

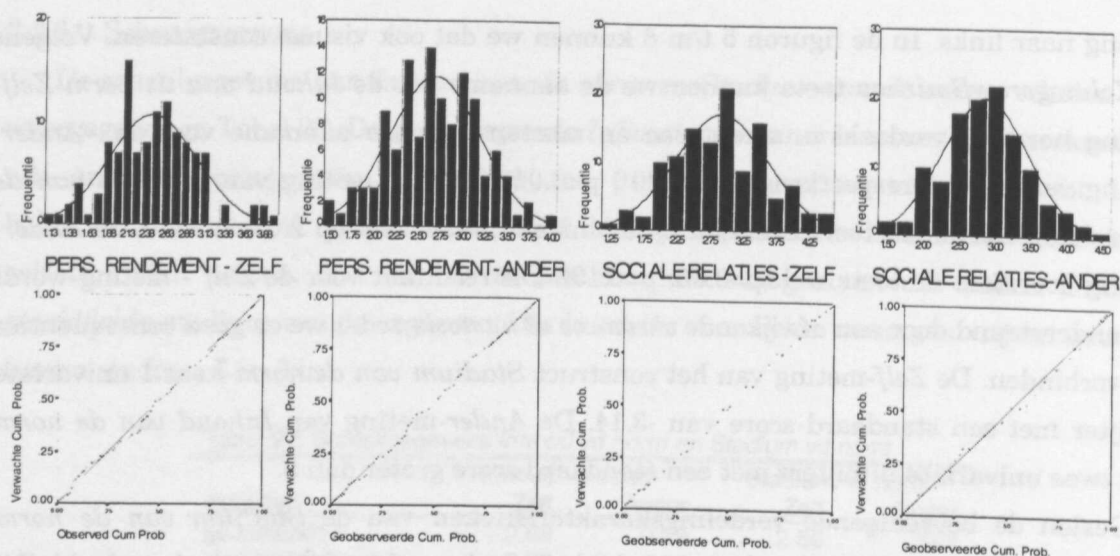
Gezien de bevredigende verdelingskarakteristieken van de *Stadium van de norm* schalen beschouwen we beide schalen als bij benadering voldoende normaal verdeeld. Dit geldt niet voor de *Inhoud van de norm* schalen. Van deze schalen konden geen getransformeerde schalen worden berekend ten behoeve van analyses die een normale verdeling van de constructen vereisen.

Tabel 28: Schaalgegevens MSI-S *Stadium v. Gemeenschapsgevoel*

statistiek	Persoonlijk rendement		Sociale relaties	
	Zelf	Ander	Zelf	Ander
gemiddelde	2.45	2.68	2.92	2.86
standaard deviatie	0.59	0.51	0.70	0.59
skewness	0.30	-0.01	0.08	0.25
kurtosis	0.37	-0.41	-0.41	0.22
Cronbach's α	0.61	0.63	0.78	0.71

Gegeven de volledig verschillende constructie van *Stadium van de norm* en *Stadium van gemeenschapsgevoel* is een interpretatie van het verschil in gemiddelden tussen beide soorten constructen niet zinvol. De standaardafwijking op deze schalen bedraagt iets meer dan een half punt (van de vijf-punts-schaal). Ook dit construct discrimineert bij de onderhavige populatie niet al te sterk.

Daarnaast zien we dat de *Stadium van de norm* en *Stadium van gemeenschapsgevoel* schalen een bevredigende verdelingskarakteristiek hebben in termen van *skewness* en *kurtosis*. Zij kennen bovendien geen univariate uitbijters. De verdeling van de schalen doorstaat dan ook de KS-toets. Eén en ander kan visueel geïnspecteerd worden in de figuren 9 t/m 12.



Figuren 9 t/m 12: Histogrammen en normal-probability-plots van *Stadium van gemeenschapsgevoel*

Samenvattend kunnen we de eigenschappen van de schalen voor het meten van het construct *Inhoud van de norm* bevredigend noemen. De eigenschappen van de schalen voor het meten van het construct *Stadium van de norm* zijn minder bevredigend. De uitval op deze schalen is iets te hoog. Om deze reden is er voor gekozen ook respondenten met slechts één scorbaar item op deze schalen toch een totaal-score te geven. De interne consistentie is wel voldoende, maar de schalen discrimineren in de onderhavige populatie echter in beperkte mate. Tot slot kunnen we de eigenschappen van de schalen die het construct *Stadium van gemeenschapsgevoel* meten eveneens bevredigend noemen. Ook dit construct kent echter een geringe discriminatoire waarde voor de onderhavige populatie.

3.4. Sociomoral Reflection Measure - Short Form (SRM-SF)

3.4.1. Uitval

Voor de SRM-SF geldt de norm dat bij een respondent tenminste 7 items gescoord moeten kunnen worden. De frequentie van het per respondent gescoorde aantal antwoorden op de SRM-SF kunt u vinden in Tabel 29.

We zien dat één persoon de norm van tenminste 7 items niet haalt. Voor deze persoon is de uiteindelijke schaal waarde niet berekend. In Tabel 30 staat de uitval per item weergegeven. Zoals uit de opzet van het interview is te verwachten (het uitlokken van zoveel mogelijk scorebare antwoorden met verschillende vragen) is het aantal missings wat hoger dan bij de schriftelijke instrumenten. We merken op dat het aantal missings bij de items 'Contract iemand' en 'Leven zelf' relatief bovengemiddeld is (27% en 39%). De overige items hebben een uitval van gemiddeld 13%. Het aantal niet gescoorde items per

persoon hangt niet significant samen met de sekse, het geboorteland (als indicator van mogelijk linguïstische of culturele verschillen) en de woordmatrijs score (zie paragraaf 3.6).

Tabel 29: Uitval SRM-SF (mondeling)

aantal gescoorde items	aantal respondenten	% respondenten
5	1	.8
7	12	10.1
8	17	14.3
9	27	22.7
10	41	34.5
11	21	17.6

3.4.2. Beschrijving op item-niveau

In Tabel 30 staan item gegevens van de 11 items uit de SRM-SF vermeld. We zien dat stadium 4, 1 en 1.5 slechts zelden gescoord zijn. Verder valt op dat de items 'Verwantschap vrienden' en 'Leven vreemden' bij deze groep respondenten relatief veel stadium drie antwoorden uitlokken. Het item 'Wettelijk recht' krijgt veel stadium 2

Tabel 30: Item gegevens SRM-SF

Item	Moreel stadium							item-totaal correlatie	uitval
	1	1.5	2	2.5	3	3.5	4		
Contract vrienden	2		13	47	32	6	1	.58	18
Contract iemand	1		25	43	15	3		.21	32
Contract kinderen	2	1	12	35	25	25	1	.54	18
Waarheid	2		18	49	29	6	1	.40	14
Verwantschap ouders	2		19	55	14	12		.37	17
Verwantschap vrienden	1		6	36	65	4		.38	7
Leven vreemden	1		9	33	50	4		.12	22
Leven Zelf			20	19	11	23		.35	46
Eigendom	1	10	24	34	37	1		.31	12
Wet		5	2	63	19	21	2	.37	7
Wettelijk recht	1	9	53	29	12	8		.39	7

antwoorden.

Aangezien slechts bij 21 leerlingen op geen enkel item een antwoord ontbreekt, blijven er voor een beoordeling van de interne consistentie met behulp van een betrouwbaarheidsanalyse slechts 21 leerlingen over. Om deze reden is de item-totaal correlatie steeds per item afzonderlijk berekend over alle respondenten met een scoorbaar antwoord. We zien dat de items 'Contract iemand' en 'Leven vreemden' relatief geringe item-totaal correlaties hebben. Stams (1994) rapporteerde in zijn studie daarentegen een relatief geringe item-totaal correlatie voor het item 'Leven zelf'. Gegeven deze discrepantie veronderstellen we voorlopig dat deze resultaten het gevolg zijn van steekproef fluctuaties. We handhaven dan ook alle items voor het berekenen van de SRM-SF totaal-score.

3.4.3. Schaalgegevens

De schaalgegevens van de SRM-SF staan in Tabel 31. We merken op dat de leerlingen gemiddeld een morele competentie hebben tussen stadium 2 en 3, met een geringe standaardafwijking van 0.26. Het instrument discrimineert daarmee niet sterk op absoluut niveau bij de onderzochte groep respondenten.

Visuele inspectie van de verdeling van SRM-SF scores laat een licht scheve verdeling zien (Fig. 13). De *Skewness* bedraagt -0.41. Één respondent heeft een standaardscore van -3.1 en is daarmee als een univariate uitbuitte te beschouwen. De

normal-probability plot (Fig. 14) toont echter dat de afwijkingen minimaal zijn. De verdeling doorstaat dan ook de *Kolmogorov-Smirnov test* op een normale verdeling ($p=0.36$).

De betrouwbaarheid van de SRM-SF hangt allereerst af van van de intercodeursbetrouwbaarheid. In Tabel 32 vinden we nogmaals de intercodeurovereenstemmings-statistieken. We zien dat deze, na de in Paragraaf 2.1.1

Tabel 31: Schaalgegevens SRM-SF (n=119)

statistiek	waarde
gemiddelde	2.6
standaard deviatie	0.26
skewness	-0.41
kurtosis	0.28
alpha	0.68

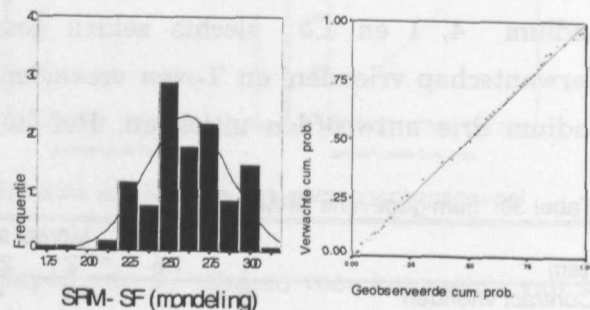


Fig. 13, 14: Histogram en normal-probability-plot SRM-SF

Tabel 32: Intercodeurovereenstemming SRMS-SF (15% van de interviews)

statistiek	waarde	minimum vereiste
SRMS correlatie r	.94	.80
Gemiddelde absolute SRMS discrepantie (in punten)	0.09	0.20
Globale Stadium overeenkomst binnen 1 interval	100 %	80%
Exacte Globale Stadium overeenkomst	68 %	50%

beschreven procedure, van een voldoende hoog niveau zijn.

Daarnaast kunnen we de betrouwbaarheid in de vorm van interne consistentie beoordelen aan de hand van Cronbach's α . Voor de 21 leerlingen met op alle items een scoorbaar antwoord bedraagt deze 0.71. Voor een schatting van Cronbach's α over meer leerlingen is eveneens een Cronbach's α berekend over alle leerlingen, waarbij voor ieder ontbrekend antwoord het schaal gemiddelde 2.66 is ingevuld. (Dit geeft mogelijk een overschatting van de α voor de gehele onderzochte populatie leerlingen.) De op deze wijze geschatte α bedraagt .68. Stams (1994) rapporteerde een aanzienlijk hogere interne

consistentie ($\alpha = 0.91$) bij gebruik van dezelfde vertaalde versie van de SRM-SF. Gibbs et al. (1992) rapporteerden eveneens een hogere interne consistentie ($\alpha = .92$). Het verschil tussen de resultaten in dit onderzoek en dat van Stams zou het gevolg kunnen zijn van de relatief homogener groep respondenten in dit onderzoek. Stams onderzocht drie leerjaren, deze studie omvat slechts één leerjaar.

Samenvattend concluderen we dat de uitval- en verdelingskarakteristieken van de SRM-SF metingen bevredigend zijn. De intercodeovereenstemming is goed. De interne consistentie is matig en blijft achter bij de interne consistentie die andere validatiestudies rapporteren. Voor gegevens ten aanzien van de test-hertest betrouwbaarheid kunnen we verwijzen naar andere studies (Gibbs et al., 1992). Het instrument discrimineert slechts in geringe mate in deze onderzoekspopulatie. De verschillen tussen de eigenschappen van het instrument in deze studie en in andere studies zijn vooreerst te interpreteren in het licht van een relatief homogene onderzoekspopulatie.

3.5. Sociomoral Reflection Objective Measure - Short Form (SROM-SF).

3.5.1. Uitval

Volgens de regels voor het uitsluiten van SROM-protocollen dienen respondenten uitgesloten te worden wanneer zij minder dan 7 'close stage' responses, minder dan 7 'closest stage' responses of meer dan 5 foutieve responses gegeven hebben op de 12 items. Van foutieve responses wordt gesproken als de 'close stage' responses ontbreken of niet overeenkomen met de 'closest stage' responses.

In de Tabellen 33, 34 en 35 staan de aantallen respondenten weergegeven met de responses op respectievelijk de Close Stage, Closest Stage en het aantal responses op de 12 items rekening houdend met de fouten.

Geen enkele respondent heeft te weinig scorebare 'Close Stage' responses. Eén respondent heeft geen enkele 'Closest

Tabel 33: Responses 'Close Stage' (n=119)

aantal gescoorde items	aantal respondenten	% respondenten
9	4	3.4
10	6	5.0
11	19	16.0
12	90	75.6

Tabel 34: Responses 'Closest Stage' (n=119)

aantal gescoorde items	aantal respondenten	% respondenten
0	1	3.4
10	2	1.7
11	6	5.0
12	110	92.4

Tabel 35: (geen) fouten (n=119)

aantal goede items	aantal respondenten	% respondenten
4	1	.8
8	1	.8
9	9	7.6
10	11	9.2
11	25	21.0
12	72	60.5

Stage' response. Een andere respondent heeft teveel fouten. In totaal vallen als gevolg van deze criteria dus 2 respondenten uit (1.7%). Stams rapporteerde een totaal uitval percentage van 5.6%. De mate van uitval is acceptabel.

3.5.2. Descriptie op item-niveau

In Tabel 36 en Tabel 37 zien we het totale aantal responses per stadium op alle respectievelijk Close Stage en Closest Stage items.

We merken op dat stadium 3 antwoorden het meest gescoord zijn. Daarnaast is het aantal stadium 4 scores eveneens zeer aanzienlijk. In dit opzicht verschilt de meting van de SROM-SF aanzienlijk van de SRM-SF: bij het laatste instrument zijn zeer weinig stadium 4 redeneringen gescoord (maar wel veel stadium 3.5 redeneringen). Voorts valt op dat het aantal stadium 3 Close scores slechts weinig hoger is dan het aantal stadium 4 scores. Het aantal stadium 3 Closest scores is echter wel aanmerkelijk hoger dan het aantal stadium 4

scores. Dit zou kunnen betekenen dat de Close scores een lichte neiging tot bias in de richting van een hoger stadium vertonen in verhouding tot het stadium waar de respondenten rapporteren zich het meest mee te kunnen identificeren. Deze observatie is tevens congruent met het veel scoren van stadium 3.5 i.p.v. stadium 4 bij de SRM-SF (mondeling interview) door de interviewer/codeur.

De item gegevens van een analyse van de interne consistentie van de twee afzonderlijke schalen en de SROM-SF totaal schaal vinden we in Tabel 38. We merken op dat de items 8 en 12 het minst samenhangen met de verschillende schalen. De effecten van het verwijderen van deze items op de totale betrouwbaarheid zijn echter gering. Daarom is ervoor gekozen de schalen in hun oorspronkelijk opzet, d.w.z. met *alle* items, te gebruiken.

Tabel 36: Closest Stage (n=119)

Stadium	Aantal responses	% responses
1	91	6.5
2	232	16.5
3	648	46.1
4	435	30.9

Tabel 37: Close Stage (n=119)

Stadium	Aantal responses	% responses
1	378	11.2%
2	807	23.9
3	1151	34.0
4	1045	30.9

Tabel 38: Item totaal correlaties SROM-SF

item	x Close (n=90)	x Closest (n=110)	Close x SROM (n=84)	Closest x SROM
1	.33	.16	.20	.17
2	.26	.37	.25	.45
3	.34	.17	.36	.24
4	.41	.09	.43	.17
5	.41	.36	.34	.42
6	.21	.19	.28	.19
7	.22	.14	.19	.22
8	.16	-.03	.16	.10
9	.21	.30	.32	.36
10	.42	.43	.45	.44
11	.31	.30	.30	.32
12	.08	.14	.14	.14

3.5.3. Schaalgegevens

In Tabel 39 vinden we de schaalgegevens van de geaggregeerde Close Stage en Closest Stage schalen, alsmede van de geaggregeerde SROM-SF schaal, de gewogen som van beide

Tabel 39: Schaalgegevens SROM-SF

statistiek	Close Stage	Closest Stage	Totaal
gemiddelde	2.91	3.01	2.98
standaard deviatie	0.23	0.34	0.26
skewness	-0.66	-0.48	-0.66
kurtosis	0.47	-0.15	0.11
alpha	0.62	0.53	0.71

(naar Gibbs et al., 1992, de Closest Score telt 2x zo zwaar). We merken op dat het gemiddelde stadium van de respondenten 3 bedraagt. Daarnaast zien we dat de totaal score scheef naar rechts verdeeld is. In de figuren 15 en 16 zien we dit eveneens aan de hand van een histogram en een normal-probability plot van de SROM-SF totaal score. Gegeven het feit dat de gemiddelde stadium score ongeveer 3 is, de maximum stadium score 4 en het aantal items relatief gering, is dit een plausibel gegeven. De verdeling van SROM-SF scores kent geen univariate uitbijters met een standaard-score van meer dan 3 of minder dan -3.

Op grond van de *Kolmogorov-Smirnov* test moeten we de aanname dat de SROM-SF scores

normaal verdeeld zijn verwerpen ($p=0.005$). Voor het uitvoeren van analysetechnieken waarvoor de aanname van een normaal verdeling van de SROM-SF scores een voorwaarde is, is een getransformeerde score berekend die deze toets wel doorstaat. In figuur 17 ziet u de samenhang tussen de oorspronkelijk SROM-SF score en deze 2e graads functie van de SROM-SF score.

Een beoordeling van de betrouwbaarheid van de SROM-SF kan alleen geschieden op basis van gegevens ten aanzien van de interne consistentie. Een eerste indicatie voor de interne consistentie van de SROM-SF is de correlatie tussen de Close schaal en Closest schaal, deze bedraagt 0.61. De Close schaal kent daarbij een wat minder matige interne consistentie ($\alpha = 0.62$) als de Closest schaal ($\alpha = 0.53$). De totale schaal kent

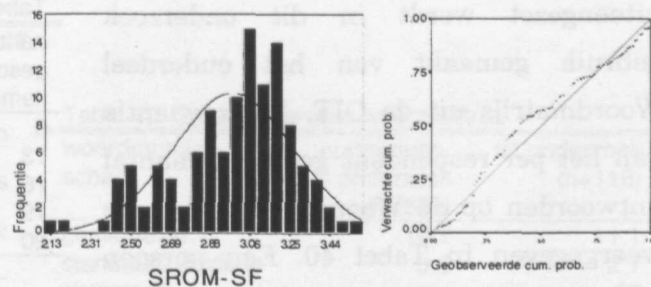


Fig. 15, 16: Histogram en normal-probability-plot SROM-SF

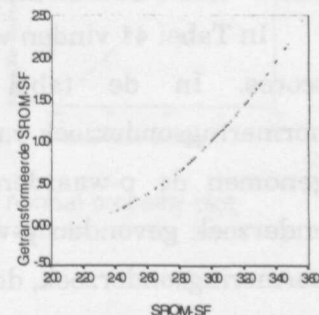


Fig. 17: Transformatie plot SROM-SF

daarnaast een α van 0.71. Deze gegevens zijn iets hoger dan de interne consistentie die Stams in zijn validatiestudie rapporteert (respectievelijk 0.56, 0.46 en 0.69), ondanks het feit dat de onderzoekspopulatie in dit onderzoek homogener is. De verschillen zijn vermoedelijk te wijten aan een andere wijze van berekenen van de interne consistentie.

Samenvattend concluderen we dat de uitval en de verdelingskarakteristieken van de SROM-SF schaal bevredigend zijn. De SROM-SF is wel scheef verdeeld, maar zulks is door een eenvoudige transformatie met een 2e graads functie te verhelpen. Ten aanzien van de betrouwbaarheid kunnen we vooralsnog alleen rapporteren over de interne consistentie. Deze is voldoende te noemen. Tien van de twaalf items zijn daarbij voldoende representatief.

3.6. Groninger Intelligentie Test (GIT)

3.6.1. Uitval

Zoals in Paragraaf 2.3.1 werd uiteengezet wordt in dit onderzoek gebruik gemaakt van het onderdeel Woordmatrijs uit de GIT. De frequentie van het per respondent gescoorde aantal antwoorden op de Woordmatrijs zien we weergegeven in Tabel 40. Eén persoon heeft geen enkel scorbaar antwoord op

Tabel 40: Uitval Woordmatrijs

aantal gescoorde items	aantal respondenten	% respondenten
0	1	.8
13	1	.8
16	2	1.7
19	3	2.5
20	112	94.1

de items van de Woordmatrijs; 112 respondenten hebben alle items beantwoord. Van de 6 overige respondenten zijn de niet te scoren antwoorden fout gerekend. Opgemerkt kan hierbij worden dat het (geringe) aantal niet te scoren antwoorden toeneemt met de moeilijkheidsgraad. We concluderen dat de uitval gering en acceptabel is.

3.6.2. Descriptie en schaalgegevens

In Tabel 41 vinden we de p-waarden en de item-totaal correlaties van de woordmatrijs scores. In de tabel zijn bovendien de oorspronkelijke p-waarden uit het normeringsonderzoek van Kooreman en Luteijn opgenomen. We zien dat over het geheel genomen de p-waarden inderdaad geleidelijk afnemen. Het patroon van de bij dit onderzoek gevonden p-waarden komt in hoge mate overeen met de p-waarden uit het normeringsonderzoek, de correlatie is (spearman rho) 0.94.

Omdat de items sterk in moeilijkheidsgraad uiteenlopen is de interne consistentie van de woordmatrijs-schaal beoordeeld met het Guttman Split-half model. Daarbij zijn de even met de oneven items vergeleken. De schaalgegevens staan in Tabel 42. De onderzochte groep kent een iets lagere schaal variantie dan de norm groep. Aangezien de leerlingen één klasniveau vertegenwoordigen is dat te verwachten. Deze lagere schaalvariantie kan mede een rol spelen bij het gegeven dat de split-half coëfficiënt (0.67) bij het onderhavige onderzoek wat lager uitvalt dan bij het normerings-onderzoek (split-half = 0.78). De interne consistentie is daarmee matig te noemen.

De woordmatrijs scores zijn daarnaast licht scheef verdeeld: de iets beter als gemiddeld scorende respondenten zijn oververtegenwoordigd. In figuur 18 en 19 zien we dit aan de hand van een *histogram* van de woordmatrijs scores en een *normal-probability plot* weergegeven.

Volgens de *Kolmogorov - Smirnov goodness of fit test* hoeven we de nulhypothese dat de woordmatrijs-scores normaal verdeeld zijn echter niet te verwerpen op 5% niveau ($p > 0.20$). Een onderzoek naar univariate uitbijters leert dat één respondent een z-score heeft van -3.1 op de woordmatrijs, de overige vallen

Tabel 41: Item gegevens Woordmatrijs

Item	norm-groep p-waarde (n=539)	p-waarde (n=118)	item-totaal correlatie (n=118)
1	.94	.97	.36
2	.93	.95	.28
3	.76	.86	.29
4	.90	.87	.40
5	.88	.97	.22
6	.76	.84	.33
7	.76	.82	.45
8	.67	.69	.22
9	.62	.47	.27
10	.35	.35	.22
11	.69	.69	.51
12	.55	.69	.22
13	.41	.41	.30
14	.32	.27	.20
15	.29	.28	.26
16	.27	.30	.24
17	.22	.20	.25
18	.13	.14	.16
19	.25	.21	.01
20	.14	.07	.11

Tabel 42: Schaalgegevens woordmatrijs

woordmatrijs schaal	normerings- onderzoek (n=539)	dit onderzoek (n=118)
gemiddelde	10.7	11.1
standaard deviatie	3.2	2.7
skewness		-.27
kurtosis		-.21
Guttman split half	.78	.67

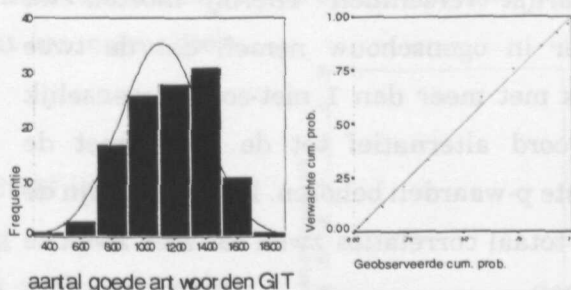


Fig. 18, 19 Histogram en normal-probability-plot
Woordmatrijs scores

ruimschoots binnen -3.0 en 3.0. Aangezien de steekproef klein is en de woordmatrijsscores enigszins scheef verdeeld zijn valt dit binnen aanvaardbare normen.

Samenvattend concluderen we dat de met de woordmatrijs verzamelde gegevens een bevredigend geringe uitval hebben. De verdelings-karakteristieken zijn eveneens bevredigend. De interne consistentie van de test is in de onderhavige steekproef matig, maar het patroon van representativiteit van de items komt in hoge mate overeen met het normeringsonderzoek van Kooreman en Luteijn. Voor overige testgegevens wordt naar dit onderzoek verwezen.

3.7. Sociale wenselijkheidsschaal

3.7.1. Uitval

In Tabel 43 vindt u de frequenties van het aantal scoorbare items bij de 119 respondenten. Een respondent valt volledig uit, één respondent mist 3 items en één respondent mist 1 item. De uitval is gering te noemen. Om listgewijs grote uitval te voorkomen in combinatie

met andere schalen is er voor gekozen om voor de 118 respondenten met scoorbare items het gemiddelde aantal sociaal wenselijke antwoorden te berekenen als sociale wenselijkheids-score.

Tabel 43: Uitval Soc. Wens.

Aantal valide items	Aantal respondenten
0	1
8	1
10	1
11	116

3.7.2. Descriptie op item-niveau

In Tabel 44 staan de item gegevens voor de sociale wenselijkheidsschaal. We zien allereerst dat de p-waarden voor alle items behoorlijk verschillen. Hierbij moeten we echter in ogenschouw nemen dat de twee items met meer dan 1 niet-sociaal-wenselijk antwoord alternatief tot de items met de laagste p-waarden behoren. Daarnaast zijn de item-totaal correlaties zwak tot zeer zwak te noemen.

Tabel 44: Item gegevens sociale wenselijkheid

Item	p-waarde	item-totaal correlatie
eerlijk zijn	.24	.08
beledigen	.18	.38
opscheppen	.34	.12
vriendelijk zijn	.35	.31
brutaal zijn	.41	.30
fout erkennen	.28	.26
zwakker erkennen	.34	.12
antipathie	.19	.32
meer krijgen	.45	.32
zin doordrijven	.65	.13
kwaadspreken	.09	.24

De interne consistentie van de sociale wenselijkheidsschaal is dan ook eveneens zwak (Cronbach's $\alpha=0.54$, $n=116$). Het verwijderen van een item uit de schaal geeft hierin geen substantiële verbetering. Stams (1994, p. 82) gebruikte dezelfde 11 items in zijn valideringsonderzoek van SROM-SF. Hij rapporteerde een beduidend hogere interne

consistentie ($\alpha=0.73$, $n=198$). De items van de sociale wenselijkheidsschaal werden door hem echter op één en dezelfde bladzijde in het instrument opgenomen. Voor dit onderzoek zijn de items verspreid tussen andere items in de MSI-s opgenomen en is de actuele formulering van de items in de sociale wenselijkheidsschaal van de PMTK gebruikt.

3.7.3. Schaalgegevens

In Tabel 45 zijn de schaalgegevens van de Sociale Wenselijkheidsschaal opgenomen. We zien dat de respondenten in gemiddeld 32% van de gevallen een sociaal wenselijk antwoord gaven. Daarnaast blijkt de sociale wenselijkheidsschaal scheef verdeeld te zijn

Tabel 45: Schaalgegevens sociale wenselijkheid (n=118)

statistiek	waarde
gemiddelde	.32
standaard deviatie	.19
skewness	.77
kurtosis	.27
alpha	.54

met een oververtegenwoordiging van respondenten met een minder dan gemiddeld aantal sociaal-wenselijke antwoorden (skewness = 0.77).

In Fig. 20 zien we dit in de vorm van een *histogram* weergegeven. De *normal probability plot* (Fig. 21) laat deze scheve verdeling eveneens zien. De verdeling doorstaat de *Kolmogorov - Smirnov test* dan ook niet ($p \leq 0.001$).

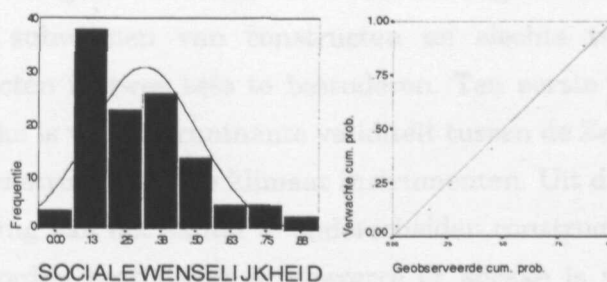


Fig 20, 21. Histogram en normal-probability plot Sociale Wenselijkheid

De visuele weergave doet echter vermoeden dat een transformatie van

de schaal wel normaal verdeeld zou kunnen zijn. Ten behoeve van Stochastische Multivariate technieken is dan ook een getransformeerde schaal geconstrueerd. De samenhang tussen de getransformeerde en oorspronkelijke schaal kunt u vinden in Fig. 22. De getransformeerde schaal is een logaritmische functie van de oorspronkelijke schaal. Ook deze getransformeerde schaal doorstaat een normaliteitstoets niet.

Samenvattend kunnen we stellen dat de uitval in de Sociale Wenselijkheidsschaal zeer gering is. De verdelingskarakteristieken zijn echter niet zoals we zouden wensen en een schaaltransformatie kan dat slechts ten dele verbeteren. De interne consistentie van de schaal is bovendien zwak. Eerder gebruik van de schaal in een onderzoek van Brugman (1988) gaf in dit opzicht betere resultaten. Er is daarom geen

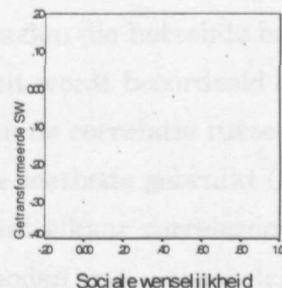
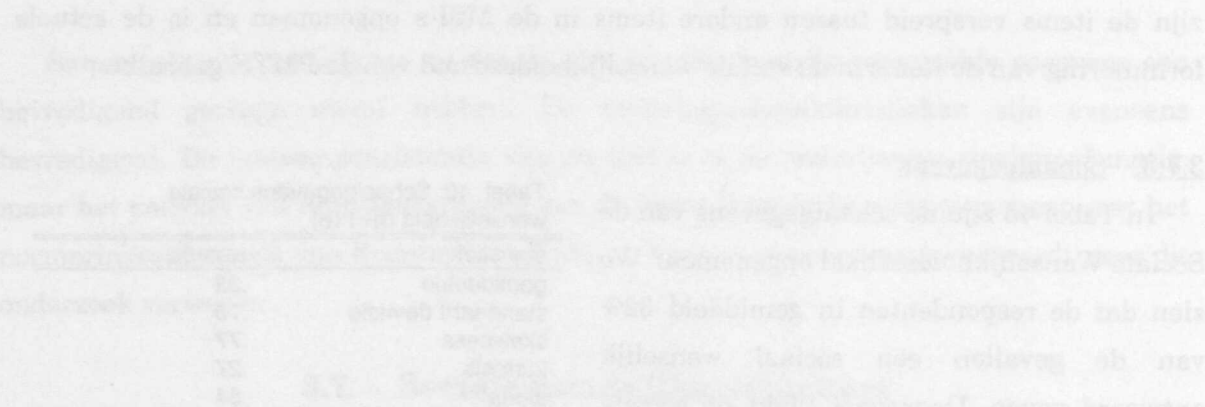


Fig. 22: Transformatie S W

reden het construct niet te handhaven voor een validatie-onderzoek van de SRM en MSI-constructen.



De eerste weergave van de SRM-constructen is te zien in Fig. 30. Het is duidelijk dat de SRM-constructen een hoge betrouwbaarheid hebben.

De tweede weergave van de SRM-constructen is te zien in Fig. 31. Het is duidelijk dat de SRM-constructen een hoge betrouwbaarheid hebben.

De derde weergave van de SRM-constructen is te zien in Fig. 32. Het is duidelijk dat de SRM-constructen een hoge betrouwbaarheid hebben.

De vierde weergave van de SRM-constructen is te zien in Fig. 33. Het is duidelijk dat de SRM-constructen een hoge betrouwbaarheid hebben.

De vijfde weergave van de SRM-constructen is te zien in Fig. 34. Het is duidelijk dat de SRM-constructen een hoge betrouwbaarheid hebben.

4. VALIDITEIT VAN DE CONSTRUCTEN

4.1. Inleiding

In dit hoofdstuk zal de onderzoeksvraag van deze studie beantwoord worden aan de hand van een analyse van de samenhang van de metingen van de verschillende constructen, voortbouwend op de resultaten van het voorafgaande hoofdstuk. De onderzoeksvraag luidde:

"Wat is de convergente en discriminante validiteit van het moreel klimaat op scholen van voortgezet onderwijs, en de morele competentie van scholieren, gemeten met behulp van een vragenlijst en gemeten met behulp van een (half) gestructureerd interview?"

In dit hoofdstuk zal deze onderzoeksvraag aan de hand van twee subvragen worden bestudeerd: gegeven het grote aantal subvormen van constructen en slechts twee methoden is er voor gekozen de constructen in twee sets te bestuderen. Ten eerste zal worden bestudeerd in welke mate er sprake is van discriminante validiteit tussen de *Zelf*- en de *Ander*-metingen van de nieuwe constructen uit de klimaat instrumenten. Uit deze analyse zal blijken dat een vereenvoudiging van het aantal te onderscheiden constructen te rechtvaardigen is. Ten tweede zal worden bestudeerd in hoeverre er sprake is van convergente en discriminante validiteit van de nieuwe constructen in samenhang met morele competentie en concurrerende constructen.

Het validatie proces wordt vaak weergegeven in een 'Multi-Trait-Multi-Method-matrix' (Campbell & Fiske, 1959). In Tabel 46 vinden we een schematische weergave van zo'n MTMM-matrix voor een aantal constructen uit het onderhavige onderzoek. Volgens de kwalitatieve criteria van Campbell en Fiske wordt convergente validiteit verkregen als de waarden van de correlaties tussen de twee verschillende methoden die hetzelfde begrip meten hoog zijn (r_{A1A2} , r_{B1B2} , Tabel 46). De discriminante validiteit wordt beoordeeld op of (a) de validiteitscoëfficiënt voor een maat (r_{A1A2} , r_{B1B2}) hoger is dan de correlatie tussen die maat en een andere maat die noch dit begrip meet noch dezelfde methode gebruikt (r_{A1B2} , r_{B1A2}), (b) verschillende metingen van hetzelfde begrip hoger met elkaar correleren dan metingen van verschillende begrippen, maar met dezelfde methoden (r_{A1B1} , r_{A2B2}), (c) het patroon van correlaties tussen maten hetzelfde is voor dezelfde en verschillende methoden. De coëfficiënten op de hoofddiagonaal (r_{A1A1}) worden doorgaans als betrouwbaarheidscoëfficiënten opgevat en berekend aan de hand van bijvoorbeeld test-

hertest gegevens. In het onderhavige onderzoek zijn deze coëfficiënten niet voorhanden. Voor het hoofdonderzoek is echter wel in een stabiliteitsmeting voorzien voor de nieuwe klimaat constructen.

Tabel 46. Multitrait-multimethod-matrix voor moreel schoolklimaat (inhoud van de norm en stadium van de norm) en morele competentie, gemeten met behulp van een semi-gestructureerd interview en een schriftelijke vragenlijst

		Mondeling			Schriftelijk		
		A1	B1	C1	A2	B2	C2
<i>Mondeling</i>							
Morele Competentie	A1	r_{A1A1}					
Moreel Schoolklimaat - inhoud	B1	r_{A1B1}	r_{B1B1}				
Morele Schoolklimaat - stadium	C1	r_{A1C1}	r_{B1C1}	r_{C1C1}			
<i>Schriftelijk</i>							
Morele Competentie	A2	r_{A1A2}	r_{B1A2}	r_{C1A2}	r_{A2A2}		
Moreel Schoolklimaat - inhoud	B2	r_{A1B2}	r_{B1B2}	r_{C1B2}	r_{A2B2}	r_{B2B2}	
Moreel Schoolklimaat - stadium	C2	r_{A1C2}	r_{B1C2}	r_{C1C2}	r_{A2C2}	r_{B2C2}	r_{C2B2}

Volgens Schmitt en Stults (1986) beperkt meer dan de helft van de door hen gereviewde MTMM studies zich tot 'kwalitatieve toetsing' van MTMM-data aan de Campbell-Fiske criteria. Een dergelijke kwalitatieve analyse is echter niet meer dan een eerste stap. De boven voorgestelde criteria lijken intuïtief zeer plausibel, maar opgemerkt moet worden dat de conclusies die er volgens de theorie aan verbonden worden gemakkelijk tot misinterpretaties kunnen leiden indien de MTMM - data niet duidelijk aan de geëiste structuur beantwoorden: Wanneer bijvoorbeeld voor sommige constructen de criteria wel voldoen maar voor andere niet, of wanneer de toets voor sommige criteria wel en voor andere niet doorstaan wordt, kan de Campbell-Fiske interpretatie van de resultaten niet valide zijn. Dit komt mede doordat het onderscheid tussen methode - en construct variantie alleen op theoretische gronden verondersteld kan worden, maar niet kwalitatief uit de MTMM - matrix kan worden afgeleid.

Van de in de literatuur voorgestelde alternatieve analysemethoden van MTMM-data is de hiërarchische Confirmatieve Factor Analyse (CFA) duidelijk favoriet (Keeves, 1988, Schmitt & Stults, 1986, Widaman, 1985). Schmitt & Stults (1986) noemen als voordelen boven variantie-analyse: de mogelijkheid verschillende alternatieve MTMM-modellen te toetsen; het gebruik van onderliggende in tegenstelling tot geobserveerde variabelen, waardoor het betrouwbaarheidsprobleem omzeild wordt; de mogelijkheid om trek-, methode- en unieke variantie in elke meting te scheiden. Deze methode zal dan ook na een kwalitatieve toetsing van de Campbell en Fiske criteria worden gebruikt om de

beantwoording van de onderzoeksvraag verder te onderbouwen. Het gebruik van deze techniek wordt helaas slechts ten dele ondersteund door de stochastische eigenschappen van de verschillende constructen (zie Hoofdstuk 3). Waar mogelijk zullen getransformeerde variabelen worden gebruikt om dit probleem te omzeilen. Overigens zijn ook aan de CFA methode voor MTMM-data nog aanzienlijke problemen verbonden die meestal genegeerd worden. Problemen met misinterpretaties kunnen bij CFA ook gemakkelijk optreden. Naast de mogelijkheid van conceptueel verschillende modellen met een vergelijkbare fit, is het ook goed mogelijk dat delen van sommige modellen niet goed geschat worden. Empirische identificatie problemen treden met name op wanneer het minimale aantal van drie methoden en drie constructen niet gehaald wordt, zoals in het onderhavige onderzoek met slechts twee methoden het geval is. Brown (1984) stelt bovendien dat indien er zowel substantiële methode - als construct variantie is, het additieve model in de CFA methode niet tegelijkertijd beide goed kan schatten. Het door hem voorgestelde (alternatieve) multiplicatieve model kan in het onderhavige onderzoek bij ontstentenis van de noodzakelijke programmatuur helaas niet gebruikt worden. Eventuele problemen met de CFA aanpak kunnen echter gedetecteerd worden door de kwalitatieve en kwantitatieve analyse met elkaar te vergelijken: ze mogen elkaar niet te veel ontlopen.

Zoals eerder vermeld wordt de CFA aanpak superieur geacht aan de variantie-analytische aanpak van MTMM- data. Daarnaast kan opgemerkt worden dat de vorm van convergente en discriminante validiteit die bij een variantie analyse wordt getoetst een heel andere is dan die in de Campbell en Fiske terminologie wordt verondersteld. Deze validiteit kan echter in zichzelf van belang zijn voor een onderzoek. In de regel wordt bij de variantie-analytische aanpak van MTMM - data alleen naar de interactie-effecten gekeken. In het onderhavige onderzoek zijn ook de hoofdeffecten van methoden en constructen interessant, omdat verschillende constructen op een soort absoluut niveau worden gemeten, zodat eventuele verschillen in gemiddeld niveau interpreteerbaar zouden moeten zijn. Waar zinvol zal dan ook langs deze weg de validiteit worden bekeken.

4.2. Het onderscheid tussen zelfperceptie en perceptie van medeleerlingen.

In het mondelinge- en het schriftelijke klimaatinstrument wordt bij het meten van de klimaat-componenten *Inhoud van de norm*, *Stadium van de norm* en *Stadium van gemeenschapsgevoel* een systematisch onderscheid gemaakt tussen de zelfperceptie van het individu (de *Zelf* - meting) en zijn perceptie van zijn klasgenoten/medeleerlingen (de *Ander* - meting). Theoretisch is het daarbij een punt van discussie of de *Zelf*- perceptie

opgevat kan worden als een vorm van morele performance, of veeleer als een andere component van de perceptie van het moreel klimaat. Op deze plaats kunnen we ingaan op de empirische implicaties van deze discussie voor het onderhavig onderzoek. Zou de *Zelf*-perceptie als een van de *Ander*-meting onderscheiden component gezien kunnen worden, of zelfs als een vorm van morele performance, dan moeten de *Zelf*- en de *Ander*-meting respectievelijk over voldoende of zelfs sterke discriminante validiteit beschikken.

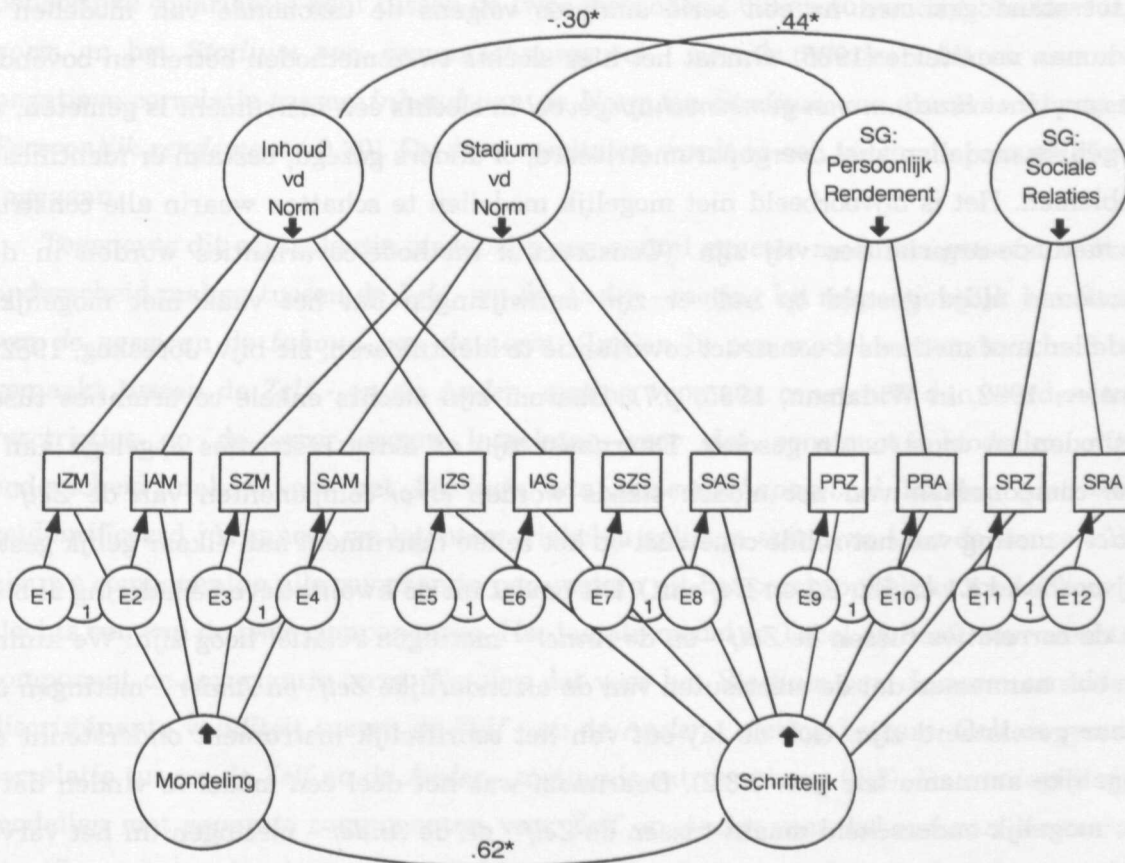
In Tabel 47 wordt een MTMM - matrix gegeven die de samenhang tussen de drie componenten van moreel klimaat beschrijft, onderverdeeld in de *Zelf*- en de *Ander*-metingen. De samenhang wordt beschreven met behulp van Pearson correlatie-coëfficiënten. In de tabel zijn de convergente validiteitscoëfficiënten (zelfde construct, verschillende methode) donker gekleurd weergegeven. Deze coëfficiënten dienen tenminste hoger te zijn dan de licht gekleurd weergegeven coëfficiënten die de samenhang beschrijven tussen de verschillend geachte *Zelf* en *Ander* metingen. (We zijn op deze plaats vooreerst geïnteresseerd in de mate van discriminante validiteit van de *Zelf*- en de *Ander*-meting, de samenhang tussen de constructen onderling komt in een volgende paragraaf aan bod.)

Tabel 47. De discriminante validiteit van de *Zelf* en de *Ander*-meting van Moreel Schoolklimaat (n=115).

		Mondeling				Schriftelijk						
		Inhoud		Stadium		Inhoud		Stadium		Stadium v. Gem.		
		Z	A	Z	A	Z	A	Z	A	Z	A	
<i>Mondeling</i>												
Inhoud vd Norm	Zelf	?										
	Ander	.51	?									
Stadium vd Norm	Zelf	.51	.30	?								
	Ander	.29	.41	.43	?							
<i>Schriftelijk</i>												
Inhoud vd Norm	Zelf	.54	.36	.26	.17	?						
	Ander	.22	.42	.04	.09	.45	?					
Stadium vd Norm	Zelf	.22	.16	.29	.28	.20	-.01	?				
	Ander	.10	.20	.19	.26	.09	.12	.61	?			
Stadium van Gem.	Zelf	-.30	-.24	-.12	-.14	-.41	-.21	-.13	.00	?		
	Ander	-.23	-.35	-.07	-.19	-.29	-.25	-.06	-.03	.81	?	
	Zelf	.33	.20	.24	.25	.39	.09	.40	.29	-.24	-.19	?
	Ander	.29	.24	.26	.31	.28	.13	.36	.32	-.30	-.26	.80
Significantie r	eenzijdig:	r>.30 ⇒ p≤.001,		r>.21 ⇒ p≤.01,		r>.16 ⇒ p≤.05						
	tweezijdig:	r>.30 ⇒ p≤.001,		r>.24 ⇒ p≤.01,		r>.20 ⇒ p≤.05						

Een kwalitatieve beschouwing van de tabel leert ons dat in het algemeen de samenhang tussen de *Zelf*- en de *Ander*-meting matig tot zeer hoog is. Met name de *Stadium van gemeenschapsgevoel* constructen vertonen een zeer sterke samenhang tussen de *Zelf*- en de *Ander*-meting (0.81 en 0.80). Bovendien zijn de *Zelf*-*Ander* correlatie coëfficiënten, uitgezonderd de *Inhoud van de norm Zelf*-meting, hoger dan de

convergente validiteitscoëfficiënten, of enige andere coëfficiënten (in de tabel steeds de onderstaande twee *Zelf* en *Ander* kolommen per construct). We constateren dat de *Zelf* - en de *Ander* - metingen een gebrek aan discriminante validiteit bezitten. De enige uitzondering daarop zouden de metingen van het construct *Inhoud van de norm* kunnen



Figuur 23. Equivalentie model voor de nieuwe klimaat constructen. Geen afzonderlijke componenten voor de *Zelf* en de *Ander* - metingen.

vormen.

In een aanvullende CFA analyse zijn twee soorten modellen met elkaar vergeleken: modellen met voor elk construct precies één latente variabele en modellen met voor één construct latente variabelen voor zowel de *Zelf*- als de *Ander* - meting. De modellen zijn - in weerwil van enkele geschonden normaliteitseisen - geschat met behulp van Maximum Likelihood schatters. Helaas bleek het in de praktijk niet mogelijk schatters op basis van een arbitraire distributie-theorie te gebruiken: de data waren daarvoor niet goed geconditioneerd. De analyses zijn met EQS (Bentler, 1989) verricht. Om de vergelijkbaarheid met een kwalitatieve MTMM - analyse zo groot mogelijk te houden is

steeds de correlatie - en niet de covariantie- matrix geanalyseerd. Parameters worden dan ook steeds voor de gestandaardiseerde vergelijkingen weergegeven, met correlaties i.p.v. covarianties waar dat toepasbaar is. De identificatie van varianties in de modellen is steeds gebeurd door de varianties van de *latente variabelen* te fixeren op 1.

Een variant van de eerste soort modellen staat weergegeven in Figuur 23. Dit model is tot stand gekomen na een serie analyses volgens de taxonomie van modellen die Widaman voorstelde (1985). Omdat het hier slechts twee methoden betreft en bovendien het construct *Stadium van gemeenschapsgevoel* in slechts één instrument is gemeten, zijn dergelijke modellen snel overgeparametriseerd, of anders gezegd, bestaan er identificatieproblemen. Het is bijvoorbeeld niet mogelijk modellen te schatten waarin alle construct- en methode-covarianties vrij zijn. (Construct x methode-covarianties worden in deze taxonomie altijd gesteld op nul: er zijn aanwijzingen dat het vaak niet mogelijk is modellen met methode x construct covariantie te identificeren, zie bijv. Jöreskog, 1982, of Bentler, 1982, in Widaman, 1985, p.7). Daarom zijn slechts enkele covarianties tussen methoden en constructen geschat. Daarnaast zijn er extra restricties opgelegd aan de *error*-componenten van het model: steeds worden *error*-componenten van de *Zelf*- en *Ander* - meting van het zelfde construct op het zelfde instrument aan elkaar gelijk gesteld (Bijvoorbeeld E1 en E2, E3 en E4, enz.). Dit omdat uit de kwalitatieve benadering al bleek dat de correlaties tussen de *Zelf* - en de *Ander* - metingen relatief hoog zijn: We kunnen dan ook aannemen dat de meetfouten van de afzonderlijke *Zelf*- en *Ander* - metingen aan elkaar gerelateerd zijn. Ook de lay-out van het schriftelijk instrument ondersteunt een dergelijke aanname (zie par. 3.3.2). Daarnaast was het doel een model te vinden dat zo min mogelijk onderscheid maakt tussen de *Zelf* - en de *Ander* - metingen (in het vervolg het equivalentie-model genoemd) en toch zo goed mogelijk fit, om de vergelijking te kunnen maken met geneste modellen waarin dit onderscheid juist niet wordt gemaakt.

Volgens de taxonomie van modellen van Widaman komen we uiteindelijk uit bij het equivalentie model in Figuur 23. Modellen met slechts één methode- of construct-factor bleken niet te voldoen. Gezien de geringe convergente validiteit in Tabel 47 - de

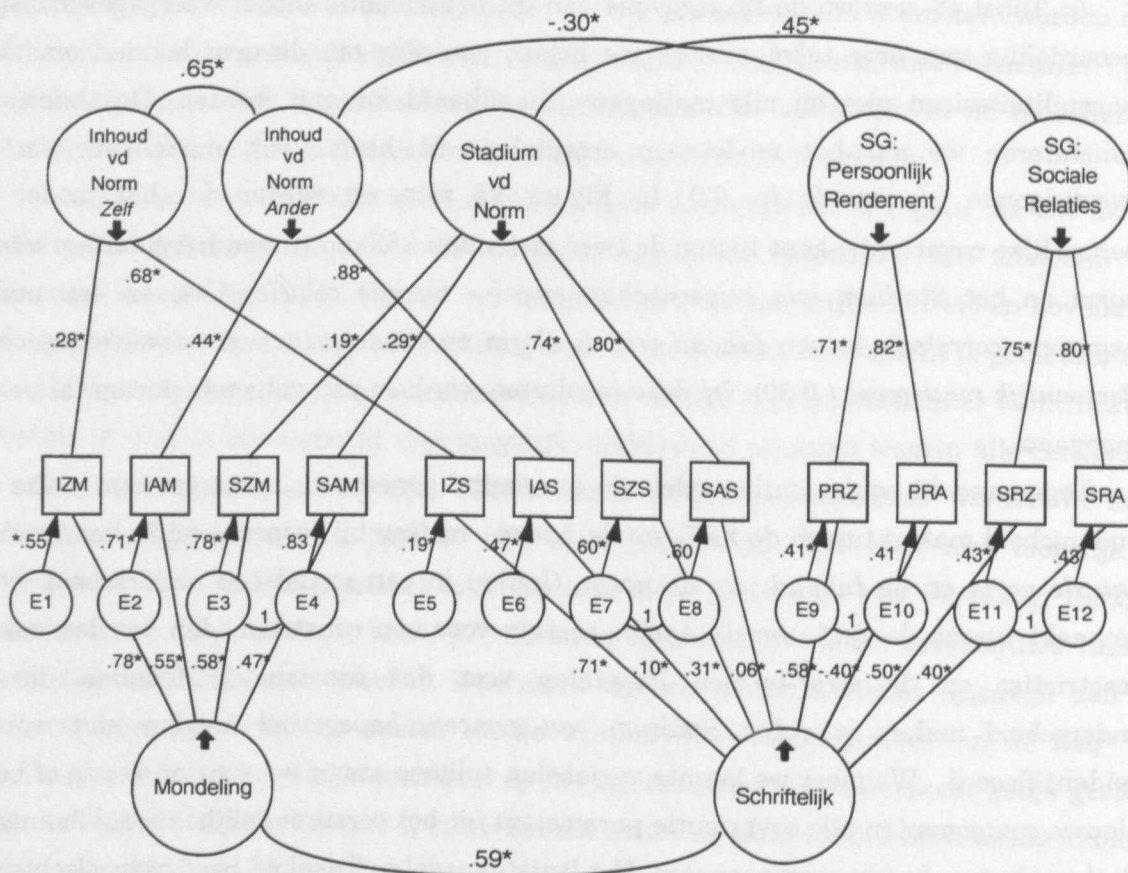
Tabel 48. Verschillen tussen modellen met- en zonder *Zelf* - *Ander* onderscheid

Model	χ^2	df	p	Fit	N-Fit	C-Fit	Z-A corr.	Verschil met equivalentie		
								χ^2	df	p
equivalentie	59.55	45	.07	.90	.96	.97	-	-	-	-
stadium:Z+A	59.12	41	.03	.90	.95	.97	.99	.43	2	>.20
inhoud:Z+A	51.78	41	.12	.92	.97	.98	.66	7.8	4	.12
Idem, - IZ*PR	51.86	42	.14	.91	.97	.98	.65	7.7	3	.05

gemiddelde correlatie is matig - is dit niet verwonderlijk.

In Tabel 48 zien we de fit-gegevens van dit equivalentie model weergegeven. Bij de beoordeling van deze tabel moeten we helaas een slag om de arm houden omdat de normaliteitseisen niet bij alle metingen gehandhaafd kunnen worden. Desalniettemin constateren we dat het model een acceptabele fit heeft (.96), maar een niet erg overtuigende χ^2 waarde ($p=.07$). In Figuur 23 zien we verder dat het model een behoorlijke covariantie kent tussen de twee methoden (.62) en tussen het *Stadium van de norm* en het *Stadium van gemeenschapsgevoel - Sociale relaties* (.44) en een matige negatieve correlatie tussen *Inhoud van de Norm* en *Stadium van gemeenschapsgevoel - Persoonlijk rendement* (-0.30). Op deze resultaten wordt in een volgende paragraaf verder ingegaan.

Tegenover dit equivalentie model zijn een aantal geneste modellen getoetst welke een onderscheid maken tussen de *Zelf* - en de *Ander* - meting bij respectievelijk het *Stadium van de norm* en de *Inhoud van de norm*. (Indien in een model een onderscheid wordt gemaakt tussen de *Zelf* - en de *Ander*- meting voor een construct, dan worden ook de restricties op de error-termen losgelaten voor dat construct.) Modellen die dit onderscheid maken voor het *Stadium van gemeenschapsgevoel* konden niet worden geïdentificeerd. Wanneer we latente variabelen splitsen staan we voor de keuze of beide nieuwe componenten alle covariantie parameters uit het oorspronkelijke model kennen of slechts één van de twee componenten. Het laatste model in Tabel 48 heeft voor slechts één component de covariantie-term. We zien dat voor het *Stadium van de norm* onvoldoende discriminante validiteit tussen de *Zelf* - en de *Ander* - meting bestaat. Ook de geschatte correlatie tussen de *Zelf* en de *Ander* - meting is extreem hoog (.99). Daarentegen zijn de modellen met separate componenten voor *Zelf* en *Ander* voor *Inhoud van de norm* wel significant beter dan het equivalentie model. Het laatste model uit de tabel, zonder de covariantie-parameter tussen de *Zelf* -meting en *Persoonlijk rendement* verdient de voorkeur. De fit-indexen van dit model overstijgen de vereiste .90. De χ^2 waarde ($p=.14$) is aan de hoge kant. Een beschouwing van de verdeling van residuen geeft een symmetrisch beeld, met een gemiddelde boven nul: er is sprake van een systematische onderschatting van de covarianties, maar niet van een onacceptabele grootte. In Figuur 24 wordt dit model volledig weergegeven.



Figuur 24. De discriminante validiteit voor de *Zelf* en de *Ander* metingen in componenten van Moreel Schoolklimaat. Structureel model met gestandaardiseerde vergelijkingen (lijnen) en correlaties (bogen). Parameters met een * zijn vrij te schatten.

In de figuur zien we dat in het algemeen de methode component van de mondelinge metingen vrij sterk is. De matige correlatie (-.27) tussen *Inhoud van de norm - Ander* en *Persoonlijk rendement* en het overbodig zijn van deze parameter voor de *Zelf* meting is verder een ondersteuning voor de validiteit van de constructen: het is logisch dat het verwachte wederkerigheids aspect van *Persoonlijk rendement* samenhangt met de verwachte handeling van de *Ander* en niet met het eigen handelen.

In het algemeen kunnen we na deze analyses stellen dat de discriminante validiteit van het onderscheid tussen de *Zelf* en de *Ander meting* onvoldoende is, behalve voor het construct *Inhoud van de norm*. Verschillende redenen kunnen ter verklaring worden aangevoerd. Allereerst is het visuele onderscheid tussen de *Zelf*- en de *Ander* - metingen in het schriftelijk instrument ineffectief te noemen: de vragen staan steeds naast elkaar. Daarnaast is het plausibel dat de leerlingen zich wel (uit ervaring) in de handelings bereidheid van de *Ander* kunnen verplaatsen, maar niet in staat zijn zich in het redeneren van de *Ander* te verplaatsen. Dit laatste is des te meer te verwachten wanneer het niveau

van eigen morele competentie en dat van de Anderen sterk uiteenloopt. In het algemeen is het de vraag in hoeverre een mens zich in de gedachtenwereld van een ander kan verplaatsen, zonder dat op het specifieke terrein intensieve communicatie plaatsvindt. Onderzoek naar communicatie geeft dit steeds weer aan. Het is te verwachten dat zonder een specifiek les-programma deze communicatie niet veel plaats vindt. Tot slot is het ook mogelijk dat een vorm van klimaat druk tot het beschreven effect leidde.

In voorgaande analyses is vooral ingegaan op het vóórkomen of ontbreken van covariantie tussen de *Zelf* en de *Ander* metingen. De klimaat constructen kunnen echter ook op absoluut niveau met elkaar vergeleken worden. In Tabel 49 vinden we de

Tabel 49. Gemiddeld niveau van *Zelf* en *Ander* metingen

Construct	Mondeling		Schriftelijk	
	Zelf	Ander	Zelf	Ander
Stadium vd norm	2.45	2.31	2.55	2.39
Inhoud vd norm	.74	.55	.62	.29
Stadium v gem. gevoel			2.61	2.56

verschillen in gemiddelden.

Met enige terughoudendheid - omdat de *Zelf* meting doorgaans iets andere categorieën kende als de *Ander* meting - constateren we dat voor alle drie de constructen de gemiddelde *Zelf*-score hoger is dan de respectievelijke gemiddelde *Ander*-score (significant, $p=0.05$). Enerzijds is er dus sprake van een afhankelijkheidsrelatie tussen de *Zelf*- en de *Ander* - metingen in termen van covariantie, anderzijds is er sprake van een systematische verschuiving van de *Zelf*- scores t.o.v. de *Ander*- scores. De richting van de verschuiving - in het voordeel van de *Zelf*-perceptie - komt overeen met te verwachten respondent-*Bias* als gevolg van sociale wenselijkheid en de theoretische onderbouwing van het concept *Moreel Klimaat*. Voor een verder onderzoek van dit verschijnsel dienen we echter allereerst te beschikken over volledig parallelle metingen.

Op grond van deze resultaten zullen in de volgende paragraaf de metingen met onvoldoende discriminante validiteit samen genomen worden, ten einde de analyses overzichtelijker te maken.

4.3. Kwalitatieve analyse van de convergente en discriminante validiteit

Tabel 50. MTMM matrix van alle klimaat - en concurrerende constructen (n=113)

	Mondeling			Schriftelijk			PR	SR	VI	SW		
	MC	Inhoud	Sta	MC	Inhoud	Sta						
		Z	A		Z	A						
<i>Mondeling</i>												
Morele competentie										.07		
Inhoud vd norm	Zelf	.11			.02					.31		
	Ander	.24	.51		.19					.10		
Stadium vd norm		.13	.47	.42	.04					.30		
<i>Schriftelijk</i>												
Morele competentie		.43	.01	.18	.04	.19	.17	.13	-.14	.15	.29	.05
Inhoud vd norm	Zelf	.22	.54	.35	.20	.20						.34
	Ander	.13	.23	.40	.10	.20	.46					.14
Stadium vd norm		.16	.17	.22	.27	.16	.16	.06				.18
SG1/2: Pers.rendement		-.19	-.26	-.29	-.15	-.15	-.36	-.24	-.06			-.30
SG3/4: Sociale relaties		.08	.33	.25	.29	.17	.35	.09	.43	-.28		.38
Verbale intelligentie		.10	-.06	.06	.09	.28	-.14	-.08	-.14	-.09	-.10	-.11
Sociale wenselijkheid		.06	.30	.10	.31	.04	.31	.12	.18	-.29	.38	-.11
Significantie r	eenzijdig: $r > .29 \Rightarrow p \leq .001$, $r > .22 \Rightarrow p \leq .01$, $r > .16 \Rightarrow p \leq .05$ tweezijdig: $r > .30 \Rightarrow p \leq .001$, $r > .24 \Rightarrow p \leq .01$, $r > .18 \Rightarrow p \leq .05$											

De beantwoording van de hoofdvraag van het onderzoek naar de convergente en discriminante validiteit van *Morele competentie* en *Moreel Schoolklimaat* zal nu plaats vinden aan de hand van de MTMM-matrix in Tabel 50. In de tabel staan onder de diagonaal alle correlaties voor ongetransformeerde scores, en boven de diagonaal enkele correlaties voor twee getransformeerde scores.

Een goed startpunt van een eventuele stochastische toets van de Campbell en Fiske criteria voor convergente en discriminante validiteit vormt een kwalitatieve analyse van deze MTMM-matrix. In de matrix zijn naast de te onderzoeken constructen ook twee concurrerende constructen opgenomen die geen grote samenhang met de overige constructen zouden mogen vertonen, te weten de GIT verbale intelligentiescore en de Sociale Wenselijkheidsscore. In de tabel worden de correlaties met deze constructen met een licht grijze achtergrond weergegeven.

4.3.1. Convergente validiteit

Het eerste criterium van Campbell en Fiske vereist dat alle convergente validiteitscoëfficiënten statistisch significant zijn en voldoende groot (bij voorkeur groter dan .55). Deze coëfficiënten bedragen voor *Morele competentie*, *Inhoud van de norm -Zelf en Ander* en *Stadium van de norm* respectievelijk 0.43, 0.54, 0.40 en 0.27. Alle vier de

coëfficiënten zijn significant ($p < 0.01$). De convergente validiteits - coëfficiënten van *Morele competentie* en *Inhoud van de norm* zijn matig, die van *Stadium van de norm* is zwak. De convergente validiteitscoëfficiënt van *Morele competentie* is lager als gerapporteerd in het onderzoek van Stams (0.77) en het onderzoek van Gibbs (0.48) .

Deze vorm van convergente validiteit is volgens de Campbell-Fiske criteria een voorwaarde voor nader onderzoek van de validiteit. We constateren dat de constructen in matige tot zwakke mate aan dit criterium voldoen. In de discussie paragraaf zal nader op het vereiste correlatie niveau voor dit criterium worden ingegaan.

4.3.2. Discriminante validiteit

Voor het vaststellen van discriminante validiteit dienen de convergente validiteitscoëfficiënten groter te zijn dan de coëfficiënten tussen het construct en andere constructen gemeten met verschillende methoden. Evenzo dienen zij groter te zijn dan de coëfficiënten tussen het construct en andere constructen gemeten met dezelfde methoden.

In dit onderzoek voldoet het construct *Morele competentie* aan dit criterium. De convergente coëfficiënt is groter dan alle correlaties tussen *Morele competentie* en andere constructen. *Morele competentie* vertoont een zwakke samenhang met het construct *Inhoud van de norm -Zelf* uit het mondelinge instrument en met de constructen *Inhoud van de norm*, *Persoonlijk rendement* en *Verbale intelligentie* uit de schriftelijke instrumenten.

Het construct *Inhoud van de norm* voldoet eveneens aan dit criterium, zij het dat de samenhang tussen de *Zelf*- en de *Ander* - metingen vrij sterk is. In de vorige paragraaf hebben we echter gezien dat de discriminante validiteit te handhaven is. Dit construct hangt verder in zwakke mate samen met de overige constructen en de *Zelf* - meting hangt samen met *Sociale wenselijkheid*. Het ontbreken van een relatie tussen *Sociale wenselijkheid* en de *Ander* - meting vormt verder een bevestiging van discriminante validiteit: volgens de theorie zou je inderdaad een vorm van sociale wenselijkheid verwachten voor de *Zelf* - meting, en niet voor de *Ander* - meting. Er is geen samenhang met *Verbale intelligentie*.

Het construct *Stadium van de norm* voldoet echter niet aan dit criterium. De correlaties tussen dit construct en *Sociale relaties* (.29 en .43) zijn hoger dan de convergente validiteitscoëfficiënt (.27). In de eerste plaats moeten we daarbij aantekenen dat de geringe convergente validiteit hieraan mede debet is. Opvallend is daarnaast de relatief hoge correlatie met *Sociale wenselijkheid*. Voorts kunnen we een zwakke correlatie met *Inhoud van de norm* vermelden. Hoewel de inhoud van de norm en het stadium van de norm zowel conceptueel als in de operationalisaties duidelijk zijn

gescheiden, komt deze relatie empirisch wel voor, onder meer omdat asociale antwoorden met een hoog moreel niveau lastig zijn te vinden. In het onderhavige onderzoek speelde deze problematiek bij het ontwerp van het schriftelijk instrument al een rol: niet bij alle asociale handelingskeuzen konden ook plausibele redeneringen van een niveau hoger dan 2 worden geformuleerd. De gevonden relatie kan dus deels het gevolg zijn van de constructie van het schriftelijk instrument.

Tot slot merken we op dat *Stadium van gemeenschapsgevoel* in relatief grote mate samenhangt met *Sociale wenselijkheid*. Aangezien een coëfficiënt voor convergente validiteit ontbreekt, kunnen we deze coëfficiënt niet aan de bovengenoemde criteria toetsen. Gezien de relatief sterke univariate eigenschappen, de hoge samenhang tussen de *Zelf-* en de *Ander* - metingen en de geringe samenhang met de overige schoolklimaat constructen kunnen we echter enige mate van discriminante validiteit verwachten.

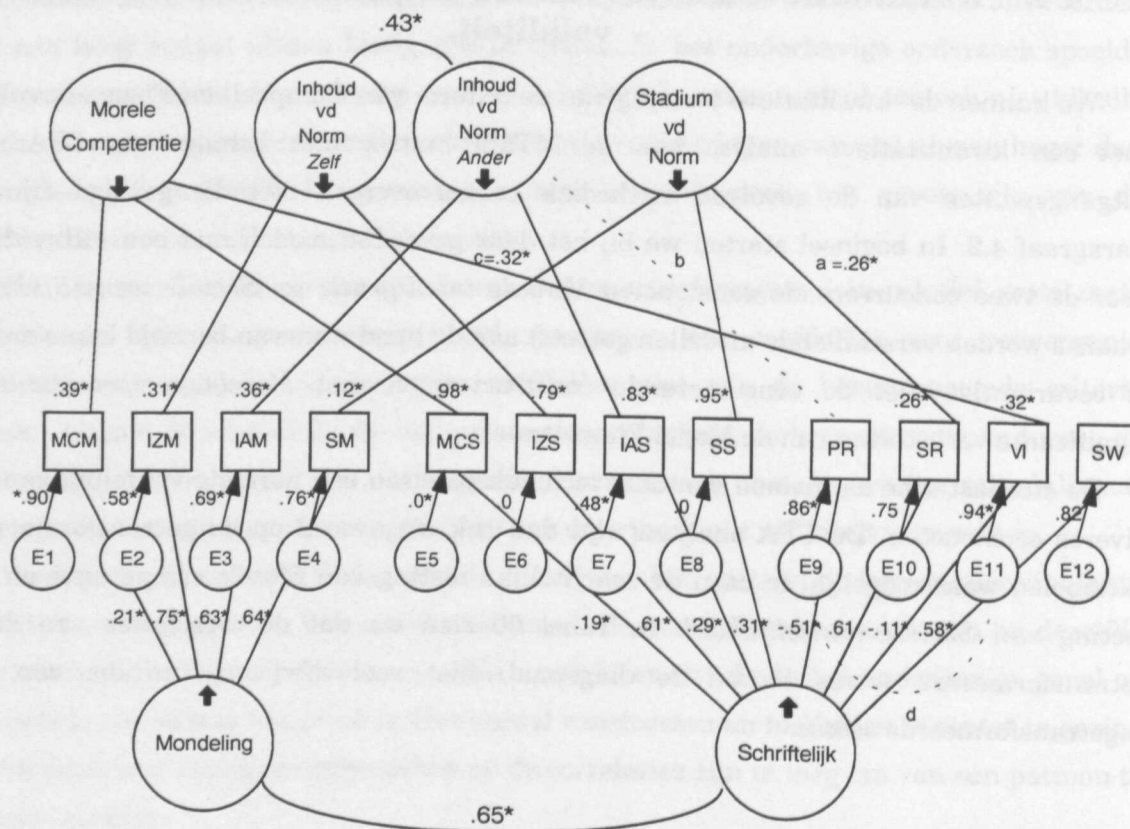
Volgens de Campbell - Fiske criteria dient voor het vaststellen van de discriminante validiteit ook het patroon van correlaties tussen de verschillende constructen bij dezelfde en andere methoden hetzelfde te zijn. We constateren dat dit in het onderhavige geval op een enkele plaats niet het geval is. Het aantal constructen en methoden is echter te gering om dit criterium zinvol te onderzoeken en de correlaties zijn te laag om van een patroon te kunnen spreken.

Samenvattend leert een kwalitatieve analyse van de MTMM-matrix ons dat de constructen *Morele competentie* en *Inhoud van de norm* een matige tot zwakke convergente en discriminante validiteit lijken te bezitten. Het construct *Stadium van de norm* kent een geringe convergente en discriminante validiteit. Het construct *Stadium van gemeenschapsgevoel* beschikt eveneens over een vorm van discriminante validiteit t.o.v. de overige klimaat constructen, maar vertoont een zorgwekkende samenhang met *Sociale wenselijkheid*. Tot slot kunnen we vermelden dat het concurrerende construct *Verbale intelligentie* slechts een relatie vertoont met *Morele competentie* en daarmee de discriminante validiteit van de nieuwe constructen ondersteunt, maar dat het concurrerende construct *Sociale wenselijkheid* zwak tot matig samenhangt met de klimaat constructen, hetgeen onwenselijk is uit het oogpunt van discriminante validiteit, maar in omvang niet onoverkomelijk.

4.4. Kwantitatieve analyse van convergente en discriminante validiteit.

We kunnen de kwalitatieve toetsing van de criteria van Campbell en Fiske aanvullen met een kwantitatieve analyse van de MTMM-matrix met behulp van CFA. De uitgangspunten van de gevolgd methodiek komen overeen met die gevolgd zijn in Paragraaf 4.2. In beginsel starten we bij het daar gevonden model, met een uitbreiding voor de twee concurrerende constructen *Verbale intelligentie* en *Sociale wenselijkheid*. Daarna worden verschillende modellen getoetst aan de hand waarvan bepaald kan worden of covarianties met de concurrerende constructen en met *Morele competentie* een significante verbetering van de Model Fit geven.

De stochastische aannamen van deze techniek vereisen een normale verdeling van de diverse constructen. De CFA analyses zijn dan ook uitgevoerd op de getransformeerde variabelen waar mogelijk, in casu de schriftelijke meting van *Morele competentie* en de meting van *Sociale wenselijkheid*. In Tabel 50 zien we dat de correlaties van deze getransformeerde scores (boven de diagonaal) niet veel afwijken van die van de ongetransformeerde scores.



Figuur 25. De discriminante en convergente validiteit voor alle constructen. Structureel model met gestandaardiseerde vergelijkingen (lijnen) en correlaties (bogen). Parameters met een * zijn vrij te schatten. (Model 4: 1+a+c)

In Figuur 25 vinden we het uiteindelijke model. In het model zijn een viertal parameters aangegeven (a t/m d) die de belangrijkste geneste varianten van het model

Tabel 51. Fit van structureel model en varianten

Model	χ^2	df	p	Model Fit			Verschil met			
				Fit	N-Fit	C-Fit	Model	χ^2	df	p
1 zonder a,b,c,d	63.9	44	.03	.80	.89	.93	-	-	-	-
2 1 + a	55.7	43	.09	.83	.93	.95	1	8.1	1	<.01
3 1+a+b	52.8	42	.12	.84	.94	.96	2	2.9	1	.10
4 1+a+c	44.3	42	.38	.86	.99	.99	2	11.5	1	<.01
5 1+a+c+d	40.7	41	.48	.88	1.0	1.0	4	3.5	1	.07

weergegeven.

In Tabel 51 vinden we voor deze varianten de fit-gegevens. Het eerste model is een uitgekledde variant van het model uit Paragraaf 4.2, zonder de covarianties tussen *Stadium van de norm* en *Sociale relaties* en zonder de covariantie tussen *Inhoud van de norm - Ander* en *Persoonlijk rendement*. In eerste instantie is getoetst of deze parameters in het model moesten worden opgenomen, met model 2 en 3 uit Tabel 51. De

laatstgenoemde covariantie bleek geen significante verbetering van de model-fit te geven. Op grond van een residu-analyse is daarna gekeken welke parameters voor opname in het model in aanmerking kwamen. De twee belangrijkste residuen zijn in model 4 en 5 getoetst. Uiteindelijk is voor model 4 gekozen. Dit model kent een goede fit. De χ^2 waarde is voldoende klein ($p=.38$) en de Non-Normed Fit en de Comparative Fit zijn zelfs .99. De lagere waarde van de Normed-Fit (.86) is waarschijnlijk het gevolg van een schaalfactor. De verdeling van de residuen kent wederom een evenwichtige opbouw met een lichte nadruk op residuen groter dan nul. Ook dit model onderschat dus de oorspronkelijke correlaties, maar onvoldoende om het model niet te kunnen accepteren. Een analyse van de ongetransformeerde metingen met hetzelfde model geeft sterk vergelijkbare resultaten, met een iets minder gunstige χ^2 waarde, hetgeen te verwachten is. De schatting van alle (gestandaardiseerde) parameters staat in Figuur 25 weergegeven.

Om het model te kunnen identificeren zijn voor de vier constructen waar alleen een enkelvoudige schriftelijke meting van bestaat geen latente variabelen opgenomen. De relatie tussen deze constructen en de overige latente variabelen laat zich door middel van de regressie parameters aflezen.

In het algemeen ondersteunt het model een redelijke mate van convergente validiteit. De construct componenten van de metingen van de klimaat constructen op het schriftelijk instrument zijn relatief hoog t.o.v. de methode componenten en de unieke error-variantie is steeds laag. De *Stadium van gemeenschapsgevoel* constructen *Persoonlijk rendement* en *Sociale relaties* kennen in het model geen afzonderlijke schatting van een latente variabele omdat er slechts één meting voor ieder van de constructen beschikbaar is. In dat geval zijn de error-componenten een schatting voor de unieke construct variantie, en deze zijn voldoende hoog ten opzichte van de methode ladingen. Voor de metingen op het mondelinge instrument zijn de klimaat construct-componenten minder sterk en de methode componenten relatief sterker. Samen met de eveneens substantiële error-componenten (de mondelinge meting van *Morele competentie* kent een zeer hoge error-component) wijst dit erop dat de mondelinge metingen van relatief mindere kwaliteit zijn, een resultaat dat in de literatuur over interview technieken al menigmaal gerapporteerd is.

Er lijkt daarnaast eveneens sprake van een behoorlijke mate van discriminante validiteit. Alle constructen kennen een eigen latente variabele in het model, zonder dat er covariantie parameters moesten worden geschat, behalve de samenhang tussen de *Zelf* en de *Ander* - meting van *Inhoud van de norm*. Deze relatie is in Paragraaf 2 van dit hoofdstuk al bestudeerd. De metingen van het *Stadium van gemeenschapsgevoel* kennen een hoge unieke variantie (error-componenten) zonder dat ze laden op de overige

constructen, uitgezonderd de relatie tussen *Stadium van de norm* en *Sociale relaties*. Deze relatie is echter niet al te sterk. Evenzo wordt er in het model slechts één directe relatie tussen de morele constructen en de concurrerende constructen *Verbale intelligentie* en *Sociale wenselijkheid* gespecificeerd. Het betreft hier een relatie tussen *Morele competentie* en *Verbale intelligentie*, die te maken zou kunnen hebben met de staat van cognitieve ontwikkeling van de respondenten, hetgeen zowel op de intelligentie als de *Morele competentie* van invloed is.

We kunnen voor het bepalen van de discriminante validiteit echter niet alleen uitgaan van het ontbreken of voorkomen van directe relaties tussen constructen. Ook indirect kunnen substantiële relaties voorkomen tussen constructen. Het model kent sterke methode-componenten voor de verschillende metingen. Bovendien hangen deze methode componenten onderling samen. Met andere woorden, veel van de covariantie tussen de morele constructen en de concurrerende constructen worden in het model beschreven d.m.v. methode variantie. In het algemeen zijn de methode componenten te sterk ten opzichte van de construct componenten om van een goede discriminante validiteit te kunnen spreken.

Daarnaast laadt de component voor het schriftelijke instrument hoog op *Sociale wenselijkheid* (.58) en de component voor het schriftelijk instrument heeft een - niet rechtstreeks gemodelleerde - correlatie van .38 met *Sociale wenselijkheid*. Met name de discriminante validiteit van *Sociale relaties* staat daardoor onder druk: de geschatte correlatie tussen deze twee constructen bedraagt .36. We constateren daarnaast dat een aspect van *Sociale wenselijkheid* een belangrijk onderdeel uitmaakt van de methode component van het schriftelijk instrument. We zien hier een voorbeeld van een interpretatie probleem in het onderscheid tussen methode - en construct variantie. Overigens kan hierbij aangetekend worden dat de correlatie tussen de twee methoden-componenten significant blijft wanneer ook relaties tussen het mondeling instrument en *Sociale wenselijkheid* worden gemodelleerd. We concluderen hieruit dat de relatie van *Sociale wenselijkheid* met het schriftelijk instrument sterker is dan die met het mondelinge instrument. Daarnaast gelden de twee methoden componenten als twee afzonderlijke componenten waartussen de samenhang niet alleen beschreven kan worden in termen van *Sociale wenselijkheid* of *Verbale intelligentie*.

Deze aanwijzingen voor een relatief gebrek aan discriminante validiteit nemen niet weg dat de correlaties tussen de *latente variabelen* voor de *Inhoud* en het *Stadium van de norm* en het construct *Morele competentie* onderling en met *Sociale wenselijkheid* in het model op nul geschat worden. Al met al concluderen we dat de discriminante validiteit tussen *Morele competentie* en de nieuwe klimaat constructen goed is, en de discriminante

validiteit van de klimaat constructen op zich zelf voldoende is, maar beperkingen kent, gezien de matige convergente validiteit en relatieve grote methoden-variantie.

Samenvattend concluderen we dat er sprake is van een beperkte mate van convergente en discriminante validiteit. De grote methode component voor de mondelinge metingen staat vermoedelijk (in de vorm van ruis) een grotere convergente validiteit in de weg. Anderzijds vertonen de schriftelijke metingen enige samenhang met het construct *Sociale wenselijkheid*. Met name de discriminante validiteit van het construct *Stadium van gemeenschapsgevoel* staat bij ontstentenis van een convergente validiteitscoëfficiënt ter discussie. Anderzijds zijn de univariate eigenschappen van dit construct sterk, met name de betrouwbaarheid, zodat een substantiële convergente validiteits coëfficiënt verwacht zou mogen worden. Tot slot kennen de nieuwe klimaat constructen alle voldoende discriminante validiteit t.o.v. het construct *Morele competentie*.

4.5. De relatie tussen Morele competentie, Inhoud en Stadium van de norm.

In de voorgaande paragrafen zijn, uitgaande van de vraagstelling, steeds modellen geschat zonder alle covariantie termen tussen de constructen. Dit is voornamelijk gebeurd omdat het bij de, in aantal methoden beperkte, opzet van deze studie niet mogelijk en evenmin noodzakelijk was alle covariantie parameters tussen de inhoudelijke constructen te schatten. Het feit dat het mogelijk is een model te schatten waarin al deze covarianties op nul gerespecteerd zijn sluit niet uit dat deze covariantie wel aanwezig kunnen zijn.

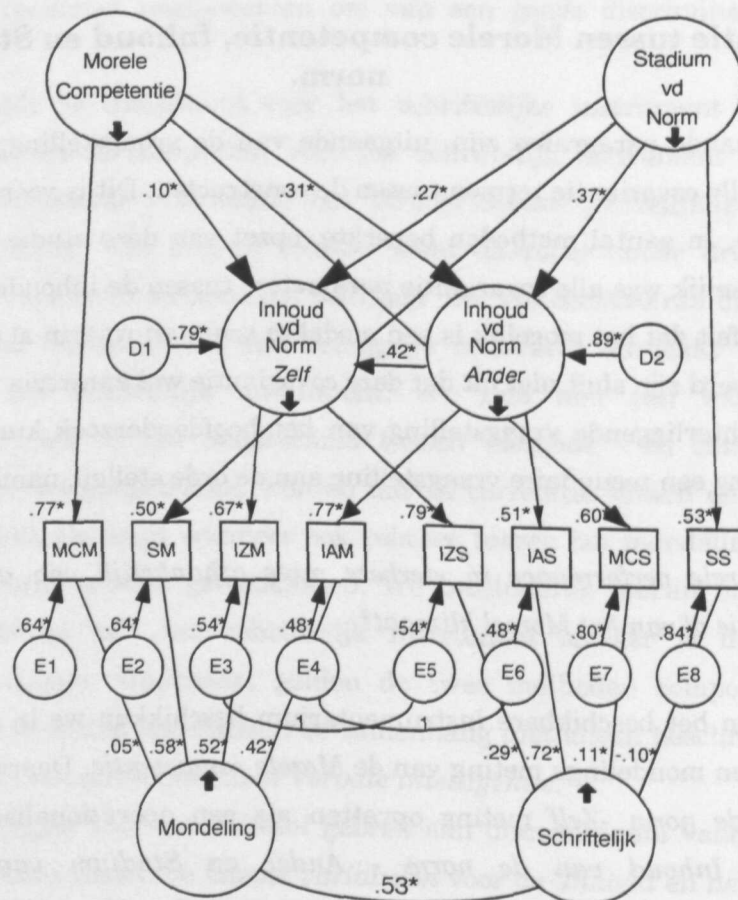
Vanuit de achterliggende vraagstelling van het hoofdonderzoek kunnen we bij deze validatiestudie nog een secundaire vraagstelling aan de orde stellen, namelijk:

Is de Morele performance in sterkere mate afhankelijk van de Morele competentie of van het Moreel klimaat?

Uitgaande van het beschikbare instrumentarium beschikken we in deze studie over een schriftelijke en mondelinge meting van de *Morele competentie*. Daarnaast kunnen we de *Inhoud van de norm -Zelf* meting opvatten als een operationalisatie van *Morele performance* en *Inhoud van de norm - Ander* en *Stadium van de norm* als operationalisaties van twee verschillende componenten van het *Moreel klimaat*.

Een eerste aanzet tot het beantwoorden van deze vraagstelling vormt het schatten van een CFA-model waarin bovengenoemde covarianties tussen constructen wel worden

geschat, teneinde inzicht te verkrijgen in de omvang van de invloed van *Morele competentie* en *Moreel klimaat* op *Morele performance*. We schatten in dit geval deze covariantie parameters ongeacht de vraag of ze significant bijdragen aan de fit van het model. In Figuur 26 vinden we zo'n model gespecificeerd. Uitgaande van de theorie zijn de relaties tussen *Morele competentie* en *Moreel klimaat* enerzijds en *Morele performance* anderzijds gespecificeerd als een afhankelijkheidsrelatie, waarbij *Morele performance* als de afhankelijke variabele optreedt. Om het aantal vrije parameters terug te brengen (het model kan anders niet geïdentificeerd worden) zijn de error-componenten van de schriftelijke en mondelinge metingen van de *Zelf*- en de *Ander*- componenten van *Inhoud van de norm* in het model aan elkaar gelijk gesteld. Met andere woorden, verondersteld wordt dat de meetfouten van dit construct op het schriftelijke instrument en het mondelinge instrument sterk met elkaar gecorreleerd zijn. Deze aanname wordt op louter praktische gronden gemaakt ten einde het model te kunnen identificeren.



Figuur 26. De relaties tussen *Morele performance*, *Morele competentie* en *Moreel klimaat*.

Dit model heeft een acceptabele fit en voldoende lage χ^2 -waarde (respectievelijk, Fit=.96, NFit=.92, CFit=.98, $\chi^2 = 8.9$, df = 6 en p=.18), hoewel de fit (relatief) minder goed is dan de fit van het model in Figuur 25, allereerst het gevolg van de relatieve overparametrisatie. De residuen zijn licht scheef verdeeld met een overwicht voor zwak positieve residuen: het model onderschat de meeste correlaties. Al met al is de fit van het model, objectief gesproken, minder goed dan die van eerdere modellen, maar de interpretatieve mogelijkheden van het model staan hier voorop.

We zien in het model dat de *Morele performance (Inhoud van de norm - Zelf)* allereerst afhangt van de verwachte handeling van de anderen (*Inhoud van de norm-Ander*), daarna van het *Stadium van de norm*, beide componenten van *Moreel klimaat*, en slechts in geringe mate van *Morele competentie*.

Een manier om het aantal vrije parameters van bovengenoemd model terug te brengen, anders dan het correleren van verschillende error-termen, is het gelijkstellen van de error-termen aan 1 minus de betrouwbaarheid van de geobserveerde variabelen. Op basis van de resultaten zoals beschreven in Hoofdstuk 3 kunnen deze betrouwbaarheden geschat worden op grond van de interne consistentie van de verschillende schalen. Een dergelijk model is eveneens geschat. In het algemeen zijn de aldus gefixeerde error-termen geringer in omvang dan de error-termen uit het model in Figuur 26. De Fit van dit alternatieve model is dan ook iets lager. De relaties tussen *Morele performance*, *Inhoud van de norm* en *Stadium van de norm* blijven echter gehandhaafd: *Morele performance (Inhoud van de norm - Zelf)* hangt wederom hoofdzakelijk van de verwachte handeling van de anderen af (*Inhoud van de norm-Ander*). Dit effect bedraagt .54 de andere effecten zijn niet groter dan .30.

De hier gevonden relaties vormen een bevestiging van de theoretische veronderstellingen. In het hoofdonderzoek zullen deze relaties vergaand onderzocht worden.

5. SAMENVATTING EN DISCUSSIE

In deze studie zijn de convergente en discriminante validiteit van de instrumenten voor moreel schoolklimaat en morele competentie nader onderzocht aan de hand van een multi-trek multi-methode (MTMM) matrix. Moreel klimaat en morele competentie zijn gemeten met behulp van (half) gestructureerde interviews en schriftelijke vragenlijsten. Verder zijn de woordmatrijs van de GIT en een schaal voor sociale wenselijkheid afgenomen.

De studie vond plaats in het kader van de constructie van een instrument voor het meten van de wijze waarop leerlingen van het voortgezet onderwijs het moreel klimaat op hun school waarnemen. Moreel schoolklimaat wordt omschreven als de waarden en normen die in een school gelden en de mate waarin deze waarden en normen worden gedeeld. Hoe leerlingen het moreel klimaat op hun school waarnemen wordt verder gemakshalve aangeduid met moreel schoolklimaat.

Hoofdstuk 1 beschrijft vanuit het gezichtspunt van de cognitieve ontwikkelingsbenadering van Kohlberg de kenmerken van de ontwikkeling van moreel klimaat en morele competentie, en de wijze waarop klimaat en competentie tot uitdrukking komen in het moreel gedrag. Het moreel klimaat wordt uiteengelegd in zes dimensies die elk op hun beurt een aantal categorieën bevatten. Deze categorieën bedoelen het democratische proces van constructie, bescherming en handhaving van normen weer te geven dat openbare scholen voor voortgezet onderwijs in de VS doorlopen die zich ontwikkelen tot 'just community' scholen. Deze scholen variëren zeer sterk qua leerlingpopulatie. Verondersteld wordt dat met behulp van deze dimensies en categorieën ook verschillen in moreel klimaat van normale scholen voor voortgezet onderwijs in Nederland vastgesteld kunnen worden. Deze scholen lopen nogal uiteen in de mate waarin en de wijze waarop democratisering op school gestalte krijgt.

Drie van de dimensies van moreel klimaat worden hier gebruikt: *De Inhoud van de norm* heeft betrekking op de gedragskeuze, bijvoorbeeld wel of niet een impopulaire medeleerling helpen, het *Stadium van de collectieve norm* verwijst naar de aard van de argumenten die gelden binnen de groep / school ter rechtvaardiging van deze gedragskeuze, en het *Stadium van gemeenschapsgevoel* verwijst naar de sociale relaties in een gemeenschap die aan deze argumenten ten grondslag liggen, bijvoorbeeld een machtsrelatie of een concreet wederkerige relatie. De waarneming door de leerlingen van de eigen keuze voor het gedrag in schoolgebonden morele dilemma's en de aard van de argumenten daarbij (*Zelf*) wordt onderscheiden van de waarneming van de gedragskeuze

en aard van de argumenten van de medeleerlingen (*Ander*). Kohlberg et al. (1974) en Higgins et al. (1984) zien *Zelf* als een operationalisatie van morele performance en *Ander* van moreel klimaat. Later wordt afstand van dit gezichtspunt genomen (Power et al., 1989) en lijkt het zelf-perspectief te vervallen. De houdbaarheid van een onderscheid tussen *Zelf* en *Ander* oordelen wordt afzonderlijk onderzocht.

Hoofdstuk 2 beschrijft de instrumenten en de steekproef. De instrumenten ter bepaling van het moreel klimaat zijn: (a) het Moreel Schoolklimaat Instrument-Interview (MSI-i) en (b) Moreel Schoolklimaat Instrument-Schriftelijk (MSI-s). De instrumenten waarmee de morele competentie wordt vastgesteld zijn: (c) de interview versie van de Sociomoral Reflection Measure-Short Form (SRM-SF), en (d) de Sociomoral Reflection Objective Measure-Short Form (SROM-SF). De variatie in methode bij deze instrumenten (schriftelijke vragenlijst versus mondeling interview) verwijst tevens naar een onderscheid dat wordt gemaakt tussen de daarmee gemeten begrippen: 'morele competentie-produktie' in geval van interview en 'morele competentie-herkenning' in geval van vragenlijst. Voor moreel klimaat kan wellicht een overeenkomstig onderscheid gelden. Voor het Nederlandse taalgebied ontbreken betrouwbare versies van schriftelijke produktietaken of mondelinge herkenningstaken voor zowel morele competentie als moreel klimaat. Internationaal gezien geldt dit alleen voor moreel klimaat. De uitspraken van leerlingen tijdens de interviews zijn gecodeerd. In geval van de morele competentie (SRM-SF) vond dit plaats aan de hand van een vertaald uitgebreid codeboek afkomstig van Gibbs et al. (1992), in geval van moreel klimaat (MSI-i) moest volstaan worden met een summier codeboek gebaseerd op de omschrijving van de dimensies en categorieën door Power et al. (1989).

De steekproef werd samengesteld uit een grotere steekproef van bijna 1600 leerlingen van de tweede, derde en vierde klas van het voortgezet onderwijs die deelnamen aan een omvangrijker project naar moreel klimaat en morele competentie ('hoofdonderzoek'). De onderhavige studie is uitgevoerd bij 120 derde klassers van alle niveaus van het voortgezet onderwijs (VBO / MAVO / HAVO / VWO). Om de studie niet te compliceren hebben we ons beperkt tot 1 leerjaar.

Hoofdstuk 3 geeft een descriptie van de uitval, de representativiteit van de items en de schaalgegevens van achtereenvolgens de MSI-i, MSI-s, SRM-SF en SROM-SF.

Voor de MSI-i is de mate van uitval op *Inhoud van de norm* acceptabel. De representativiteit van de items en de interne consistentie zijn eveneens bevredigend. De uitval op item-niveau op *Stadium van de norm* is echter te groot als gevolg van de

gebruikte interview-techniek¹. Bovendien is de representativiteit van de verschillende items c.q. de interne consistentie zeer gering. Daarnaast discrimineert het construct in de onderzochte populatie slecht. De constructie van de MSI-i *Stadium van de norm* constructen behoeft dan ook enige aanpassing, enerzijds om de mate van uitval terug te dringen, anderzijds om de variantie van de meting te vergroten. Gedacht kan worden aan een vergroting van de variatie in situaties ten koste van de variatie en diepte van het type vragen. Voor de *Zelf - Inhoud* meting en de *Ander - Stadium* meting zijn getransformeerde scores berekend.

Voor de MSI-s zijn de eigenschappen van de schalen voor het meten van het construct *Inhoud van de norm* bevredigend. De eigenschappen van de schalen voor het meten van het construct *Stadium van de norm* zijn minder bevredigend. De uitval op deze schalen is iets te hoog. Om deze reden is er voor gekozen ook respondenten met slechts één scorebaar item op deze schalen toch een totaal-score te geven. De interne consistentie is wel voldoende, maar de schalen discrimineren in de onderhavige populatie in beperkte mate. Tot slot kunnen we de eigenschappen van de schalen die het construct *Stadium van gemeenschapsgevoel* meten eveneens bevredigend noemen. Ook dit construct kent echter een geringe discriminatoire waarde voor de onderhavige populatie.

Voor de SRM-SF zijn de (geringe) uitval en de verdelingskarakteristieken bevredigend. De intercodeur-overeenstemming is goed. De interne consistentie is matig en blijft achter bij de interne consistentie die andere validatiestudies rapporteren. Voor gegevens ten aanzien van de test-hertest betrouwbaarheid kunnen we verwijzen naar andere studies (Gibbs et al., 1992). Het instrument discrimineert slechts in geringe mate in deze onderzoekspopulatie. De verschillen tussen de eigenschappen van het instrument in deze studie en in andere studies zijn vooreerst te interpreteren in het licht van een relatief homogene onderzoekspopulatie.

Voor de SROM-SF zijn de uitval en de verdelingskarakteristieken bevredigend. De SROM-SF is wel scheef verdeeld, maar een transformatie in de vorm van een eenvoudige 2e graads functie van de oorspronkelijke SROM-SF kan dit verhelpen. Ten aanzien van de

¹De hier gebruikte interview techniek is direct afgeleid van technieken zoals die na intensieve research zijn ontwikkeld voor het meten van morele competentie. Het voordeel van deze methoden-parallelliteit is dat de discriminante validiteit van het stadium van de norm t.o.v. morele competentie niet beschreven kan worden in termen van methode-variantie. Zowel empirisch als theoretisch zijn echter argumenten voorhanden om deze methoden parallelliteit te verlaten. Wanneer alleen de methode aangepast wordt voor het mondelinge instrument, d.w.z. daar waar het ook het meest noodzakelijk en plausibel is, kan dat wellicht deels het hierbovengenoemde probleem rond het vaststellen van discriminante validiteit verhelpen.

betrouwbaarheid kunnen we voornamelijk alleen rapporteren over de interne consistentie. Deze is voldoende. Tien van de twaalf items zijn daarbij voldoende representatief.

De Woordmatrijs heeft een bevredigend geringe uitval. De verdelingskarakteristieken zijn eveneens bevredigend. De interne consistentie van de test is in de onderhavige steekproef matig, maar het patroon van representativiteit van de items komt in hoge mate overeen met het normeringsonderzoek van Kooreman en Luteijn. Voor overige testgegevens wordt naar dit onderzoek verwezen.

De Sociale Wenselijkheidsschaal heeft een zeer geringe uitval. De verdelingskarakteristieken zijn echter niet zoals we ons zouden wensen en een schaaltransformatie kan dat slechts ten dele verbeteren. De interne consistentie van de schaal is bovendien zwak. Eerder gebruik van de schaal gaf in dit opzicht betere resultaten (Brugman, 1988). Er is daarom geen reden het construct niet te handhaven voor een validatie-onderzoek van de SRM en MSI-constructen.

In hoofdstuk 4 wordt de onderzoeksvraag naar de convergente en discriminante validiteit beantwoord met behulp van een kwalitatieve analyse van de MTMM correlatiematrix en de kwantitatieve CFA analyses. De voordelen van CFA boven een variantie-analytische aanpak kunnen in dit onderzoek slechts ten dele worden benut door de aard van de stochastische eigenschappen van de verschillende constructen. Bij een CFA aanpak is de kans op misinterpretatie reëel. Bovendien is in dit onderzoek slechts gebruik gemaakt van twee methoden en worden slechts enkele constructen met beide methoden gemeten. Dit beperkte design geeft problemen met de identificatie van onderdelen van de modellen-taxonomie zoals Widaman die voor MTMM-data ontwikkeld heeft. Deze problemen zijn opgelost door slechts enkele covariantie-termen tussen construct componenten te schatten, en de overige covariantie-termen te fixeren op nul.

Gegeven het grote aantal subvormen van constructen en slechts twee methoden wordt eerst de empirische basis voor het *Zelf-Ander* onderscheid bij de klimaat constructen onderzocht (*Inhoud van de norm*, *Stadium van de norm*, *Stadium van gemeenschapsgevoel*). De kwalitatieve analyse wijst uit dat de samenhang tussen de *Zelf*- en de *Ander*-meting matig tot zeer hoog is. Geconstateerd wordt dat de *Zelf*- en *Ander*-metingen een gebrek aan discriminante validiteit bezitten, met als mogelijke uitzondering het construct *Inhoud van de norm*. In een aanvullende CFA analyse zijn twee soorten modellen met elkaar vergeleken: equivalentie modellen met voor elk construct één latente variabele en geneste modellen met een latente variabele voor zowel de *Zelf* als de *Ander* meting per construct. Een model van het eerste type heeft een acceptabele fit, maar een niet erg overtuigende χ^2 waarde. Toetsing van het tweede type modellen wijst uit dat voor

de *Zelf- Ander* meting van *Stadium van de norm* onvoldoende discriminante validiteit bestaat. Modellen met separate latente variabelen voor *Zelf* en *Ander* voor *Inhoud van de norm* daarentegen zijn wel significant beter dan het equivalentie model.

De beantwoording van de vraag naar de convergente en discriminante validiteit van morele competentie en moreel schoolklimaat vindt plaats met alleen het *Zelf - Ander* onderscheid voor het construct *Inhoud van de norm*. Een kwalitatieve analyse van de MTMM-matrix wijst uit dat de constructen *Morele competentie* en *Inhoud van de norm* een matige tot zwakke convergente en discriminante validiteit bezitten. Het construct *Stadium van de norm* kent een geringe convergente en discriminante validiteit. Het construct *Stadium van gemeenschapsgevoel* bezit eveneens discriminante validiteit t.o.v. de overige klimaatconstructen, maar vertoont een zorgwekkende samenhang met *Sociale wenselijkheid*. Het concurrerende construct *Verbale intelligentie* heeft slechts een relatie met *Morele competentie* en ondersteunt daarmee de discriminante validiteit van de klimaat constructen. Het concurrerende construct *Sociale wenselijkheid* hangt zwak tot matig samen met de klimaat constructen, wat onwenselijk is vanuit het gezichtspunt van discriminante validiteit; de hoogte is echter niet onoverkomelijk.

Bij de kwantitatieve analyse van de MTMM matrix met behulp van CFA is gekozen voor een model waarin zowel alle afzonderlijke methoden als alle afzonderlijke constructen zijn opgenomen. Daarnaast zijn er relaties gespecificeerd tussen de twee methode-componenten, tussen *Morele competentie* en *Verbale intelligentie* en tussen *Stadium van de norm* met *Sociale relaties*. Het model kent relatief sterke methode componenten voor de verschillende metingen, die bovendien onderling samenhangen. De grote methode component van met name de mondelinge constructen staat een grote convergente validiteit in de weg. De schriftelijke component laadt daarnaast hoog op *Sociale wenselijkheid* (en *Inhoud van de norm* en *Sociale relaties*) en de mondelinge component laadt in mindere mate eveneens op *Sociale wenselijkheid*. We concluderen dat in de geschatte methoden componenten een bepaalde vorm van sociale wenselijkheid een rol speelt. Met name de discriminante validiteit van *Sociale relaties* staat door deze indirecte samenhang met *Sociale wenselijkheid* onder druk. Ook de discriminante validiteit van *Inhoud van de norm* en in mindere mate *Stadium van de norm* staat hierdoor op het niveau van de geobserveerde variabelen onder druk. Weliswaar zijn de geschatte latente componenten voor deze constructen niet gerelateerd aan *Sociale wenselijkheid*, maar de convergente validiteit van deze (latente) constructen is onvoldoende om dit een doorslaggevende rol te geven. Voorts merken we op dat uit de analyse van discriminante validiteit van de *Zelf-* en de *Ander-* metingen bleek dat deze relaties hoofdzakelijk optreden op het niveau van de *Zelf-*metingen. We duiden deze

gemeenschappelijke component, gerelateerd aan de methode en *Sociale wenselijkheid*, dan ook als *prosociale aanpassing*.

De correlaties tussen de latente variabelen voor de *Inhoud* en het *Stadium van de norm* en *Morele competentie* onderling en met *Sociale wenselijkheid* worden op nul geschat. Al met al concluderen we dat de discriminante validiteit tussen morele competentie en de constructen van moreel klimaat goed is, en de discriminante validiteit van de klimaat constructen voldoende is maar beperkingen kent.

De convergente validiteit van de diverse constructen in dit onderzoek laat te wensen over. De convergente validiteit van de SRM-SF en de SROM-SF is in dit onderzoek veel lager dan op grond van de resultaten van de try-out mocht worden verwacht (Stams, 1994). Ook werd een lagere negatieve en positieve samenhang gevonden met respectievelijk sociale wenselijkheid en verbale intelligentie. Daarbij moet worden aangetekend dat we de bijgestelde versie gebruikten waarin de eerste try-out heeft geresulteerd, die ook in de tweede try-out van Stams tot betere resultaten leidde (bij uitsluitend VBO leerlingen). Ook de uitval in dit onderzoek op grond van het feit dat de antwoorden van de leerlingen niet voldeden aan de gestelde normen, was beduidend lager dan in de try-out.

Een verklaring voor het verschillend resultaat uit een verschil in afname zoals Stams (pers. comm.) ons heeft voorgesteld lijkt ons vanwege die geringe uitval niet waarschijnlijk. Die verschillen in afname hebben betrekking op: (a) een klassikale schriftelijke afname in aanwezigheid van de docent in de try-out versus een groepswijze afname onder leiding van een stagiaire bij leerlingen geselecteerd uit verschillende klassen in dit onderzoek; (b) een verschil in afnameduur: deze was in dit onderzoek langer dan in de try-out, er werden nu in totaal meer instrumenten afgenomen. Verder werd in dit onderzoek dezelfde normering gebruikt als in het onderzoek van Stams. De resultaten die nu zijn gevonden voor de SR(O)M-SF stroken overigens wel met die Gibbs zelf heeft gevonden met deze instrumenten bij een qua leeftijd betrekkelijk homogene groep, echter met uitsluitend schriftelijke procedures. Overeenkomstige resultaten worden door Gibbs (Gibbs et al., 1984, Table 3, p. 533) ook gerapporteerd bij gebruikmaking van de uitgebreide versies van de hier gebruikte instrumenten, de SRM en SROM. Correlaties van 0.21 (!) tot 0.61 bij homogene groepen en 0.58 voor de hele groep. Door verwijdering van de groep met de uitzonderlijke lage correlatie van 0.21 (dit is een jonge leeftijdsgroep van gemiddeld 12,5 jaar) wordt de correlatie 0.70 voor de groep die in leeftijd varieert van 14 tot 18,5 jaar. Vooralsnog schrijven we de wat teleurstellende resultaten dan ook toe aan de betrekkelijk homogene steekproef in dit onderzoek.

De resultaten van het schoolklimaat instrument, MSI-i/-s, blijven duidelijk achter bij die van de morele competentie, SR(O)M-SF, maar zullen wellicht enigszins verbeteren als de gegevens van het tweede dilemma over Diefstal verwerkt zijn. In dit geval bestaan er onvoldoende empirische gronden het *Zelf-Ander* onderscheid te handhaven bij het *Stadium van de norm* en *Stadium van gemeenschapsgevoel*. Als beide van elkaar worden onderscheiden, moet niet alleen onafhankelijke codering plaatsvinden, zoals ook in dit onderzoek is gebeurd, maar ook onafhankelijke dataverzameling. Dergelijke studies zijn ons niet bekend. In ieder geval voldoen de studies van Kohlberg et al. (1974), Higgins et al. (1984) en Power (1988), waarin met name interviews zijn gebruikt en verder 'spontane' interacties over actuele morele vraagstukken, niet aan dit criterium. Voor de *Inhoud van de norm* is het *Zelf-Ander* onderscheid wel zinvol te handhaven. De meest waarschijnlijke verklaring voor dit verschijnsel, die in een ander verband ook door Power et al. (1989) naar voren is gebracht, lijkt ons dat kennis van het gedrag van anderen normaliter wel aanwezig is, maar niet van de redenen van dit gedrag.

Hoewel de resultaten mate met betrekking tot de convergente en discriminante validiteit strikt genomen niet meevallen kunnen we hier nog enkele relativerende kanttekeningen bij plaatsen. In het algemeen blijkt het vinden van convergente validiteitscoëfficiënten hoger dan .50 zeldzaam. Met name in onderzoek waarin zowel schriftelijke als mondelinge instrumenten gebruikt worden, worden veelvuldig convergente validiteitscoëfficiënten gerapporteerd van .30 of zelfs minder. Dit doet de vraag naar voren komen in hoeverre in sociaal-wetenschappelijk onderzoek in het algemeen mondelinge en schriftelijke metingen in een MTMM-studie met elkaar vergeleken kunnen en dienen te worden.

Voorts is in de analyse-fase van het onderzoek naar voren gekomen dat de afbakening van de onderzoeksgroep voor deze studie ongewenste gevolgen had: de klimaat constructen differentiëren onvoldoende tussen de respondenten. De keuze van deze onderzoeksgroep is tot stand gekomen op pragmatische gronden. Het beperken van het aantal scholen in het onderzoek vereenvoudigt het vinden van respondenten aanzienlijk, terwijl het beperken van het aantal leerjaren multi-modale verdelingen van constructen kan voorkomen. Op theoretische gronden is echter eveneens te verwachten dat bij een bredere opzet van de studie de variantie van de verschillende metingen toeneemt. Dit zou tot een aanzienlijke stijging van de convergente validiteit kunnen leiden.

Tenslotte, welke empirische relaties tussen moreel klimaat, morele competentie en morele performance kunnen we afleiden uit de matrices en modellen? De waarneming van individueel moreel gedrag (*Inhoud van de norm-Zelf*) blijkt conform de theorie sterker gerelateerd aan de waarneming van het gedrag van anderen (*Inhoud van de norm-Ander*)

REFERENTIES

- dan aan de individuele morele competentie. Of deze relatie zich wijzigt bij een hogere morele competentie in vergelijking met een lagere, zal in het hoofdonderzoek nader worden onderzocht.
- Bentler, P.M. (1980). Validation of the sociomoral reflection objective measure - Short Form. *Psychological Reports*, 51, 135-146.
- Bentler, P.M. (1983). *BGJ structural equations program manual*. Los Angeles: BMDP Statistical Software.
- Berkowitz, M. W. (1985). The role of discussion in moral education. In M. W. Berkowitz & F. Oser (Eds.), *Moral Education: Theory and Application* (pp. 179-219). Hillsdale, New Jersey: Lawrence Erlbaum.
- Berkowitz, M.W. & Keller, M. (1994). Transitional processes in social cognitive development: A longitudinal study. *International Journal of Behavioral Development*, 17, 447-467.
- Boon, J. (1930). *The Concept Developmental Stage*. Academisch proefschrift KUN.
- Boon, J. & Meulenaar, P.G.M. (1989). A developmental model of hierarchical stage structure in objective moral judgements. *Developmental Review*, 9, 112-134.
- Brugman, D. (1985). *Personologie van de leerstof*. Academisch proefschrift, RUG.
- Brugman, D. (1994). De school als een morele gemeenschap. In T. Othof & D. Brugman (red.), *De ontdekking van moraal leer* (pp. 177-202). Lisse: Swets & Zeitlinger.
- Brugman, D., Hout, K., Van Roozendaal, M., & Tsvetkova, L. W. C. (1994). Morele klimaat variabelen voor verschillende onderlagen de perceptie van de leerling. *Comenius*, 14, 54-61.
- Sykes, J.P. (1956). Domain specificity and the logic of using general ability as an independent variable or covariate. *Merrill-Palmer*, 41, 1-24.
- Campbell, B.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 17-23.
- Charlier, M. (1981). Alternative readings of the competence-performance relation. In M. Charlier & M. Chapman (Eds.), *Criteria for Competence* (pp. 24-44). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. (1988). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 26, 37-45.
- Colby & Kohlberg (1987). *The Measurement of Moral Judgments*. Vol. 1: Theoretical Foundations and Research Validation. Cambridge: Cambridge University Press.
- Damon, W. & Hillen (1982). Peer interaction and development of change in children's moral reasoning. *Merrill-Palmer Quarterly*, 29, 241-262.
- Davison, M. L., Robbins, S. & Bruggen, D. B. (1982). Stage structure in objective moral judgments. *Developmental Psychology*, 24, 137-149.

REFERENTIES

- Andersen, C.S. (1982). The search for schoolclimate: A review of the search. *Review of Educational Research, 52*, 368-420.
- Basinger, K.S. & Gibbs, J.C. (1987). Validation of the sociomoral reflection objective measure - Short Form, *Psychological Reports, 61*, 139-146.
- Bentler, P.M. (1989). *EQS structural equations program manual*. Los Angeles: BMDP Statistical Software.
- Berkowitz, M. W. (1985). The role of discussion in moral education. In M. W. Berkowitz & F. Oser (Eds.), *Moral Education: Theory and Application* (pp. 179-219). Hillsdale, New Jersey: Lawrence Erlbaum.
- Berkowitz, M.W. & Keller, M. (1994). Transitional processes in social cognitive development: A longitudinal study. *International Journal of Behavioral Development, 17*, 447-467.
- Boom, J. (1992). *The Concept 'Developmental Stage'*. Academisch proefschrift KUN.
- Boom, J. & Moolenaar, P.C.M. (1989). A developmental model of hierarchical stage structure in objective moral judgements. *Developmental Review, 9*, 133-145.
- Brugman, D. (1988). *Personaliseren van de leerstof*. Academisch proefschrift. RUL.
- Brugman, D. (1994). De school als een morele gemeenschap. In T. Olthof & D. Brugman (red.), *Het ontstaan van moreel besef* (pp. 177-202). Lisse: Swets & Zeitlinger.
- Brugman, D., Høst, K., Van Roosmalen, M., & Tavecchio, L. W. C. (1994). Moreel klimaat van scholen voor voortgezet onderwijs: de perceptie van de leerlingen. *Comenius, 14*, 64-82.
- Byrnes, J.P. (1995). Domain specificity and the logic of using general ability as an independent variable or covariate. *Merrill-Palmer, 41*, 1-24.
- Campbell, D.T. & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 9*, 13-27.
- Chandler, M. (1991). Alternative readings of the competence-performance relation. In M. Chandler & M. Chapman (Eds.), *Criteria for Competence* (pp. 5-18). Hillsdale, NJ: Lawrence Erlbaum.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.
- Colby & Kohlberg (1987). *The Measurement of Moral Judgment. Vol. 1: Theoretical Foundations and Research Validation*. Cambridge: Cambridge University Press.
- Damon, W. & Killen (1982). Peer interaction and the process of change in children's moral reasoning. *Merrill-Palmer Quarterly, 28*, 347-368.
- Davison, M. L., Robbins, S. & Swanson, D. B. (1978). Stage structure in objective moral judgments. *Developmental Psychology, 14*, 137-146.

- Eisenberg, N. (1989). Development of prosocial values. In N. Eisenberg, J. Reykowski & E. Staub (Eds.), *Social and Moral Values. Individual and Societal Perspectives* (87-103). Lawrence Erlbaum, New Jersey.
- Elbers, E. (1991a). The development of competence and its social context. *Educational Psychology Review*, 3, 73-94.
- Elbers, E. (1991b). Context, culture, and competence. *Educational Psychology Review*, 3, 137-148.
- Evans, Ch. (1986). Kohlberg's Moral Judgment Interview: Is there a need for additional research. In S. Modgil & C. Modgil (Eds.), *Lawrence Kohlberg, Consensus and Controversy* (pp. 471-480). Londen: Falmer Press.
- Gibbs, J.C. (1979). Kohlberg's moral stage theory. A Piagetian revision. *Human Development*, 22, 89-112.
- Gibbs, J.C. (1991). Toward an integration of Kohlberg's and Hoffman's theories of morality. In W.L. Kurtines & J.L. Gewirtz (Eds.), *Handbook of Moral Behavior and Development. Vol. 1: Theory* (pp. 183-222). Hillsdale, N.J.: Lawrence Erlbaum.
- Gibbs, J.C., Arnold, K.D., Morgan, R.L., Schwartz, E.S., Gavaghan, M.P. & Tappan, M.B. (1984). Construction and Validation of a Multiple-Choice Measure of Moral Reasoning. *Child Development*, 55, 527-536.
- Gibbs, J.C., Basinger, K.S., & Fuller, D. (1992). *Moral Maturity. Measuring the Development of Sociomoral Reflection*. Hillsdale, N.J.: Lawrence Erlbaum.
- Gibbs, J.C., Widaman, K.F. & Colby, A. (1982). Construction and validation of a simplified, group administrable equivalent to the Moral Judgment Interview. *Child Development*, 53, 895-910.
- Gilligan, G. & Belenky, M. F. (1980). A naturalistic study of abortion decisions. In R. L. Selman & R. Yando (Eds.), *Clinical-Developmental Psychology. New Directions for Child Psychology* 7 (pp. 69-90). San Francisco: Jossey-Bass Inc.
- Haan, N. (1975). Hypothetical and actual moral reasoning in a situation of civil disobedience. *Journal of Personality and Social Psychology*, 32, 255-270.
- Higgins, A. (1991). The just community approach to moral education: Evolution of the idea and recent findings. In W.L. Kurtines & J.L. Gewirtz (Eds.), *Handbook of Moral Behavior and Development. Vol. 3: Application* (pp. 111-141). Hillsdale, N.J.: Lawrence Erlbaum.
- Higgins, A., Power, C., & Kohlberg, L. (1984). The relationship of moral atmosphere to judgments of responsibility. In W.L. Kurtines & J.L. Gewirtz (Eds.), *Morality, Moral Behavior and Moral Development* (pp. 74-106). New York: Wiley.
- Hoeks, C., Dudink, A. & Wouters, L. (1984). *Overwegingen bij sociale beslissingen. Vragenlijst*. Universiteit van Amsterdam.
- IJzendoorn, M.H. van (1986). Morele argumentatie en bezorgdheid over kernbewapening. *Pedagogische Studiën*, 63, 1-13.

- Jöreskog, K.G. & Sörbom, D. (1988). *LISREL 7: A guide to the program and applications*. Chicago: SPSS, Inc.
- Kelman, H.C. & Hamilton, V.L. (1989). *Crimes of Obedience. Toward a social psychology of authority and responsibility*. New Haven en Londen: Yale Univ. Press.
- Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of Socialization Theory and Research* (pp. 347-480). Chicago: Rand McNally.
- Kohlberg, L. (1970). The moral atmosphere in the school. Reprinted in D. Purpel & K. Ryan (Eds., 1976), *Moral Education...It Comes with the Territory* (pp. 196-220). Berkeley, Californië: McCutchan.
- Kohlberg, L. (1981). Exploring the moral atmosphere of institutions: A bridge between moral judgment and moral action. In L. Kohlberg, *The Meaning and Measurement of Moral Development* (pp. 35-53). Worcester, Mass.: Clark University Press.
- Kohlberg (1981). *Essays in Moral Development. Vol 1. The Philosophy of Moral Development. Moral Stages and the Idea of Justice*. San Francisco: Harper & Row.
- Kohlberg, L. (1984). *Essays in Moral Development. Vol 2. The Psychology of Moral Development*. San Francisco: Harper & Row.
- Kohlberg, L. (1985). The just community approach to moral education in theory and practice. In M.W. Berkowitz & F. Oser (Eds.), *Moral education: Theory and application* (pp. 27-87). Hillsdale, New Jersey: Lawrence Erlbaum.
- Kohlberg, L. & Candee, D. (1984). The relationship of moral judgment to moral action. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Morality, Moral Behavior and Moral Development* (pp. 52-73). New York: Wiley.
- Kohlberg, L. & Higgins, A. (1987). School Democracy and Social Interaction. In W.M. Kurtines & J.L. Gewirtz (Eds.), *Moral Development through Social Interaction* (pp. 102-128). New York: Wiley.
- Kohlberg, L., Levine, C., & Hower, A. (1983). *Moral Stages: A Current Formulation and a Response to Critics*. Basel: Karger.
- Kohlberg, L., Scharf, P., & Hickey, J. (1974). The justice structure of the prison - a theory and an intervention. *Prison Journal*, 51, 3-14.
- Kooreman, A. & Luteijn, F. (1987). *Groninger Intelligentie Test - schriftelijke verkorte vorm*. Lisse: Swets & Zeitlinger.
- Krebs, D. L., Vermeulen, S. C. A., Carpendale, J. I., & Denton, K. (1991). Structural and situational influences on moral judgment: The interaction between stage and dilemma. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of Moral Behavior and Development. Vol. 2: Research* (pp. 139-169). Hillsdale, N.J.: Lawrence Erlbaum.
- Krebs, D.L. & Van Hesteren, F. (1994). The development of altruism: Toward an integrative model. *Developmental Review*, 14, 103-158.

- Kutnick, P. (1986). The relationship of moral judgment to moral action: Kohlberg's theory, criticism and revision. In S. Modgil & C. Modgil (Eds.), *Lawrence Kohlberg, Consensus and Controversy* (pp. 125-148). Londen: Falmer Press.
- Laak, J.J.F. & Aleva, A.L. (1994). Rechtvaardig verdelen. In T. Olthof & D. Brugman (red.), *Het ontstaan van moreel besef* (pp. 117-146). Lisse: Swets & Zeitlinger.
- Leming, J.S. (1976). An exploratory inquiry into the multi-factor theory of moral behavior. *Journal of Moral Education*, 5, 179-188.
- Levine, C. G. (1979). Stage acquisition and stage use. *Human Development*, 22, 145-164.
- Mey, L. de (1994). *Morele oordelen in ontwikkelings- en opvoedingsperspectief*. Academisch proefschrift VU.
- Nucci, L. & Weber, E.K. The domain approach to values education: From theory to practice. In: W.M. Kurtines & J.L. Gewirtz (Eds.), *Handbook of Moral Behavior and Development. Vol. 3: Application* (pp. 251-266). Hillsdale, N.J.: Lawrence Erlbaum.
- Nunnally, J.C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Oppenheimer, L.J.T. (1978). Moral Judgment: The investigation of an unidimensional or multidimensional construct? In L.J.T. Oppenheimer, *Social Cognitive Development* (pp. 121-132). Academisch proefschrift KUN.
- Piaget, J. (1932/1977). *The Moral Judgement of the Child*. Harmondsworth: Penguin books.
- Pollard, S.R., Kurtines, W.M., Carlo, G., Dance, M., & Maycock (1991). Moral education from the perspective of psychosocial theory. In: W.M. Kurtines & J.L. Gewirtz (Eds.), *Handbook of Moral Behavior and Development. Vol. 3: Application* (pp. 289-315). Hillsdale, N.J.: Lawrence Erlbaum.
- Ponemon, L.A. (1992). Ethical reasoning and selection socialization in accounting. *Accounting Organizations and Society*, 17, 239-258.
- Power, C. (1988). The just community approach to moral education. *Journal of Moral Education*, 17, 195-208.
- Power, C., Higgins, A., & Kohlberg, L. (1989). *Lawrence Kohlberg's Approach to Moral Education*. New York: Columbia University Press.
- Pritchard, M.S. (1991). *On becoming responsible*. University Press of Kansas.
- Puka, B. (1991). Towards the redevelopment of Kohlberg's theory: Preserving essential structure, removing controversial content. In: W.M. Kurtines & J.L. Gewirtz (Eds.), *Handbook of Moral Behavior and Development. Vol. 1: Theory* (pp. 373-393). Hillsdale, N.J.: Lawrence Erlbaum.
- Rest, J. R. (1979). *Development in Judging Moral Issues*. Minneapolis: University of Minnesota press.
- Rest, J. R. (1983). Morality. In J. H. Flavell & E. Markman (Eds.), *Manual of Child Psychology: Vol. 3. Cognitive Development* (pp. 555 - 629). New York: Wiley.
- Rest, J.R. (1986). *Moral development. Advances in Research and Theory*. New York: Praeger.
- Rest, J.R. (1986a). Moral research methodology. In S. Modgil & C. Modgil (Eds.), *Lawrence Kohlberg, Consensus and Controversy* (pp. 455-469). Londen: Falmer Press.

- Rest, J.R. (1986b). Interchange. Rest replies Evans. In S. Modgil & C. Modgil (Eds.), *Lawrence Kohlberg, Consensus and Controversy* (pp. 455-469). Londen: Falmer Press.
- Rest, J.R. & Narvæz, D. (1994) (Eds.). *Moral Development in the Professions: Psychology and Applied Ethics*. Hillsdale, N.J.: Lawrence Erlbaum.
- Schmitt, N. & Stults, D.M. (1986). Methodology review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement, 10*, 1-22.
- Siegel, M. (1980). Kohlberg versus Piaget: To what extent has one theory eclipsed the other? *Merrill-Palmer Quarterly, 26*, 285-297.
- Sprinthall, N.A., Stewart, D., Siemienska, R., Jasinska-Kania, A. (1993). *Public Administration and the Public Trust: Assessing Ethical Reasoning in the United States and Poland* (A progress report). Paper presented at the AME conference 'From Moral Judgment to Moral Action', Tallahassee, November.
- Stams, G.J. (1994). *De sociomorele competentie van scholieren*. Ongepubliceerde doctoraalscriptie. Vakgroep Algemene Pedagogiek, RUL.
- Sternberg, R.J. & Kolligian, J. (1990) (Eds.). *Competence Considered*. New Haven: Yale.
- Turiel, E. (1983). *The Development of Social Knowledge: Morality and Convention*. Cambridge: Cambridge University Press.
- Vandenberg, R.J. & Scarpello, V. (1991). Multitrait-multimethod validation of the satisfaction with my supervisor scale. *Educational and Psychological Measurement, 52*, 203-212.
- Walker, L.J., De Vries, B.J., & Trevethan, S.D. (1987). Moral stages and moral orientations in real-life and hypothetical dilemmas. *Child Development, 58*, 842-858.
- Waters, E. & Sroufe, L.A. (1983). Social competence as a developmental construct. *Developmental Review, 3*, 79-97.
- Weber, J. (1990). Managers' moral reasoning: Assessing their responses to three moral dilemmas. *Human Relations, 43*, 687-702.
- Widaman, K.F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9*, 1-26.
- Yeates, K.O. & Selman, R.L. (1989). Social competence in the schools: toward an integrative developmental model for intervention. *Developmental Review, 9*, 64-100.
- Zwart-Woudstra, H.A., Meijer, T., Fintelman, M., & IJzendoorn, M.H. van (1993). *Vragenlijst Sociale Relaties*. Leiden: Centrum voor Gezinsstudies/Vakgroep Algemene Pedagogiek, RUL.

BIJLAGE A. Overzicht van de scoringsregels voor het stadium van de collectieve norm

Stadium 1

De redeneringen bevatten verwijzingen naar *angst voor fysieke sancties of concrete straf*.

Bv.:

- Als je helpt dan slaan ze je in elkaar
- Als je niet helpt dan krijg je straf van de leraar.

Gedragskeuze: kan zowel helpen als niet helpen zijn.

Stadium 2

In de redeneringen wordt verwezen naar *eigenbelang* (het eigen voordeel/nadeel wordt benadrukt). Gebrek aan interesse voor anderen (klasgenoten). Concrete wederkerigheid, afschuiven van verantwoordelijkheid. Bv.:

- Als jij hulp nodig hebt en niemand zou jou willen helpen dan zou je dat ook niet leuk vinden.
- Ik help niet want er zijn genoeg anderen die het wel kunnen doen

Gedragskeuze: kan zowel helpen als niet helpen zijn.

Stadium 2/3

In de redeneringen staat centraal dat *leerlingen erbij willen horen*, zijn bang voor afkeuring door klasgenoten. Angst om buiten de groep te vallen. Bv.:

- Als je helpt dan hoor je niet meer bij de groep.
- Ze helpen niet omdat ze bang zijn om uitgelachen te worden

Gedragskeuze: niet helpen

Stadium 3/2

De redeneringen zijn *pro-sociaal*. Empathie, medelijden. Een bezorgdheid voor anderen die het eigenbelang overstijgt. Leerlingen worden als individuen gezien, er wordt *niet* naar ze verwezen als klasgenoten. Bv.:

- Je helpt omdat het lullig is om niet te helpen
- Je helpt omdat hij die hulp hard nodig heeft

Gedragskeuze: helpen

Stadium 3

In de redeneringen wordt verwezen naar *sociale relaties* tussen de leerlingen, naar het belang van helpen voor goede sociale relaties. Saamhorigheid, onderlinge verbondenheid. Als klasgenoten heb je wat voor elkaar over. Bv.:

- Je helpt want het is een leerling in de klas, hij hoort bij de klas ook al is hij niet populair.
- Je helpt want het is toch een klasgenoot.

Gedragskeuze: helpen

Stadium 4

In de redeneringen wordt verwezen naar *de klas als geheel* en de consequenties van het wel of niet helpen voor de hele groep. Gedeelde verantwoordelijkheid. Bv.:

- Je hebt geen klas meer over als je niet gaat helpen.
- Je helpt omdat je je verantwoordelijk voelt.

Gedragskeuze: helpen

Voorbeeld-vragen MSI-i Helpen-dilemma.

A. Soms zitten er leerlingen in de klas die niet bij de groep horen, er een beetje buiten vallen of niet populair zijn. Bert is zo'n onpopulaire leerling. Hij presteert slecht in een vak waar de meeste leerlingen goed in zijn, en blijft misschien wel zitten. Na het gezamenlijk gedeelte van de les kunnen jullie de rest van de tijd zelf werken. Bert vraagt of iemand hem wil helpen.

- Herken je deze situatie?

Is een dergelijke situatie bij jou in de klas wel eens voorgekomen?

- Wat verwachten de leerlingen van deze school van elkaar dat ze in een dergelijke situatie zullen doen?

- Wat zou jij doen in een dergelijke situatie? Waarom?

- Wat denk je (verwacht je) dat je klasgenoten in een dergelijke situatie zullen doen?

Hoeveel van je klasgenoten zullen Bert in deze situatie helpen, denk je? Waarom zouden ze (wel/niet) helpen?

- Wat vind je dat je klasgenoten in een dergelijke situatie **zouden moeten** doen? Waarom?
- Zou jij het afkeuren als niemand Bert zou helpen? Waarom wel/niet?
- Zouden je klasgenoten het afkeuren als niemand Bert zou helpen? Waarom wel/niet?
- Zou jij je verantwoordelijk voelen als Bert zou blijven zitten? Waarom wel/niet?
- Zouden je klasgenoten zich verantwoordelijk voelen als Bert zou blijven zitten? Waarom wel/niet?
- Zou iemand er iets van zeggen als niemand Bert helpt?

BIJLAGE B. Voorbeeld van het helpen dilemma uit de MSI-s

Helpen

Soms zitten er leerlingen in de klas die niet bij de groep horen, er een beetje buiten staan of niet populair zijn. Bert is zo'n onpopulaire leerling, ook jij mag hem niet echt. Bovendien presteert hij slecht in een vak waar de meeste leerlingen goed in zijn. Misschien blijft hij dit jaar wel zitten. Na een gezamenlijk deel van de les kunnen de leerlingen de rest van de tijd voor zichzelf werken. Bert vraagt of iemand hem wil helpen met zijn huiswerk.

Stel dat Bert een leerling in jouw klas is.

1. WAT GEBEURT ER?

- A. Herken je deze situatie? Ja
 Nee
- B. Wat zou jij in een dergelijke situatie doen? Helpen
 Misschien helpen
 Niet helpen
- C. Hoeveel van je klasgenoten zouden een leerling als Bert helpen, denk je? Niemand
 Sommigen
 Ongeveer de helft
 Allen of de meesten
- D. Wat vind je dat je klasgenoten in een dergelijke situatie **zouden moeten** doen? Ze moeten helpen
 Dat moeten ze zelf weten

Je kunt verschillende redenen hebben om Bert wel of niet te helpen. Hieronder vind je er een aantal. Geef voor elke reden aan of deze lijkt op een reden die je zelf zou geven. Geef daarnaast aan of je denkt dat de meeste van je klasgenoten die reden zouden geven.

2. REDENEN OM BERT WEL TE HELPEN

A. Als klasgenoten elkaar niet helpen dan houd je geen klas meer over.

Lijkt dit op een reden die jij zou geven?

- Ja, heel veel
 Ja, een beetje
 Nee
 Ik begrijp deze reden niet

Lijkt dit op een reden die de meeste van je klasgenoten zouden geven?

- Ja, heel veel
 Ja, een beetje
 Nee, ik denk van niet
 Ik begrijp deze reden niet

B. Als jij hulp nodig hebt en niemand zou jou willen helpen, dan zou je dat ook niet leuk vinden.

Lijkt dit op een reden die *jij* zou geven?

- Ja, heel veel
- Ja, een beetje
- Nee
- Ik begrijp deze reden niet

Lijkt dit op een reden die *de meeste van je klasgenoten* zouden geven?

- Ja, heel veel
- Ja, een beetje
- Nee, ik denk van niet
- Ik begrijp deze reden niet

C. Het is een leerling in je klas, hij hoort bij de klas ook al is hij niet populair.

Lijkt dit op een reden die *jij* zou geven?

- Ja, heel veel
- Ja, een beetje
- Nee
- Ik begrijp deze reden niet

Lijkt dit op een reden die *de meeste van je klasgenoten* zouden geven?

- Ja, heel veel
- Ja, een beetje
- Nee, ik denk van niet
- Ik begrijp deze reden niet

D. Als je Bert niet helpt, wordt de leraar misschien wel boos.

Lijkt dit op een reden die *jij* zou geven?

- Ja, heel veel
- Ja, een beetje
- Nee
- Ik begrijp deze reden niet

Lijkt dit op een reden die *de meeste van je klasgenoten* zouden geven?

- Ja, heel veel
- Ja, een beetje
- Nee, ik denk van niet
- Ik begrijp deze reden niet

Welke van de vier bovenstaande redenen (A t/m D) komt het meest overeen met de reden die *jij* zou geven? Kruis één hokje aan.

- A. Als klasgenoten elkaar niet helpen dan houd je geen klas meer over.
- B. Als jij hulp nodig hebt en niemand zou jou willen helpen, dan zou je dat ook niet leuk vinden.
- C. Het is een leerling in je klas, hij hoort bij de klas ook al is hij niet populair.
- D. Als je Bert niet helpt, wordt de leraar misschien wel boos.

Welke reden komt het meest overeen met de reden die *de meeste van je klasgenoten* volgens jou zouden geven? Kruis één hokje aan.

- A. Als klasgenoten elkaar niet helpen dan houd je geen klas meer over.
- B. Als jij hulp nodig hebt en niemand zou jou willen helpen, dan zou je dat ook niet leuk vinden.
- C. Het is een leerling in je klas, hij hoort bij de klas ook al is hij niet populair.
- D. Als je Bert niet helpt, wordt de leraar misschien wel boos.

BIJLAGE C. Items van het Stadium van gemeenschapsgevoel

1) Machtsgebruik (stadium 1)

Item

- 6 Je moet oppassen wat je doet anders word je gepakt door de andere leerlingen.
- 15 Je moet oppassen wat je doet anders pakken de leraren je weer.
- 24 Op school houd ik me koest want een aantal leerlingen maakt de dienst uit.
- 31 Op school houd ik me koest want de leraren bepalen wat er gebeurt.
- 38 De leerlingen bedriegen de leraren als het hun goed uitkomt.

2) Concrete wederkerigheid (stadium 2)

- 7 Ik help alleen een leraar als ik er iets voor terug krijg.
- 16 Ik help alleen een medeleerling als hij mij ook ergens mee kan helpen.
- 25 De leerlingen van deze school komen alleen voor hun vrienden op.
- 26 Het hoort erbij dat de "bruggers" op deze school gepest worden.
- 32 Leerlingen van deze school laten iedereen links liggen die niet tot hun vrienden behoort.
- 39 Je doet mee aan het pesten van de "bruggers" want als "brugger" werd je zelf ook getreiterd.

3) Relaties en delen/ vriendschap (stadium 3)

- 8 Ik vind dat deze school een gemeenschap is, waar leerlingen en leraren om elkaar geven
- 17 Als een leerling buiten de groep valt probeer ik hem of haar erbij te betrekken.
- 33 Soms zijn we net een grote familie hier op school, vind ik.
- 40 Je kunt op deze school op de andere leerlingen vertrouwen.
- 44 Als een leerling tegenslag heeft, veel ziek is of problemen heeft, dan krijgt hij of zij steun van medeleerlingen en leraren.

4) Sociaal contract (stadium 4)

- 9 Je hebt als leerling als het ware een contract met de school, daarom vind ik dat je niet mag spijbelen.

- 18 Leerlingen en leraren houden zich aan de regels, omdat dat het beste is voor de school als gemeenschap.
- 27 Leerlingen en leraren zijn samen verantwoordelijk voor wat er op school gebeurt.
- 34 Ik vind dat je als leerling verplicht bent om te zorgen dat alle leerlingen zich op school thuis voelen.
- 41 Je moet je niet afvragen wat de school voor jou kan doen, maar wat jij voor de school kunt doen.

BIJLAGE D. Voorbeeld-vragen SRM-sf.

1. Denk eens terug aan een keer dat je iets hebt beloofd aan een vriend(in) van je. Hoe belangrijk is het dat mensen beloftes tegenover vrienden nakomen, als ze kunnen?

Noem één van de volgende antwoorden: *Heel belangrijk* / *belangrijk* / *niet belangrijk*.

Waarom is het heel belangrijk/ belangrijk/ niet belangrijk?

2. Hoe denk je over het nakomen van een belofte aan een willekeurig persoon? Hoe belangrijk is het dat mensen hun beloftes nakomen, zelfs tegenover iemand die ze nauwelijks kennen?

Noem één van de volgende antwoorden: *Heel belangrijk* / *belangrijk* / *niet belangrijk*.

Waarom is het heel belangrijk/ belangrijk/ niet belangrijk?

3. Hoe denk je over het nakomen van een belofte aan een kind? Hoe belangrijk is het voor ouders om zich aan beloftes te houden tegenover hun kinderen, als ze dat kunnen?

Noem één van de volgende antwoorden: *Heel belangrijk* / *belangrijk* / *niet belangrijk*.

Waarom is het heel belangrijk/ belangrijk/ niet belangrijk?

PROBLEEM 1

In Amerika zou een vrouw spoedig sterven aan een bijzonder soort kanker. Er was één medicijn dat haar volgens de dokters zou kunnen redden. Het was een vorm van radium, die de apotheker uit dezelfde stad pas ontdekt had. Het medicijn was duur om te maken, maar de apotheker wilde dat mensen tien keer de prijs zouden betalen die het hem gekost had om het medicijn te maken.

De man van de zieke vrouw, Heinz, probeerde bij iedereen die hij kende het geld te lenen, maar hij kreeg slechts de helft van het bedrag bij elkaar. Heinz vertelde de apotheker dat zijn vrouw op sterven lag en vroeg of hij het medicijn voor minder geld mocht kopen of het geld later mocht betalen. De apotheker zei echter: "Nee, ik heb het medicijn uitgevonden en ik wil er geld aan verdienen." De enige manier waarop Heinz het medicijn nu nog kon krijgen, was door in te breken bij de apotheker en het medicijn te stelen.

Heinz heeft een probleem. Hij zou zijn vrouw moeten helpen en haar leven moeten redden. Dit betekent echter dat hij moet inbreken en daardoor de wet overtreedt.

- Wat moet Heinz doen? Hij moet het medicijn stelen
 Hij moet het medicijn niet stelen
 Ik weet het niet

Waarom:

Er worden nu wat dingen in het verhaaltje veranderd. We willen graag weten of je dan nog dezelfde mening hebt (wel stelen, niet stelen of ik weet het niet). Verder willen we ook graag weten welke zaken *jij* belangrijk vindt bij het oplossen van het probleem en *waarom* je die zaken belangrijk vindt. Wil je daarom alle volgende vragen beantwoorden.

1. Stel dat de stervende persoon niet de vrouw van Heinz is maar een vriend van hem en deze vriend heeft niemand anders om hulp aan te vragen. Wat moet Heinz dan doen? Stelen
 Niet stelen
 Ik weet het niet
- 1a. Hoe belangrijk is het om alles te doen wat je kunt om het leven van een vriend te redden, zelfs als je daarvoor de wet moet overtreden? Heel belangrijk
 Belangrijk
 Niet belangrijk

REDENEN OM HET LEVEN VAN EEN VRIEND TE REDDEN

- 1b. Stel dat je een reden zou moeten geven *waarom* het voor jou *belangrijk* is om alles te doen wat je kunt om het leven van een vriend te redden. Welke reden zou *jij* geven? Lijkt één van de volgende redenen op een reden die *jij* zou geven? Als je een reden niet begrijpt, kruis je 'ik weet het niet' aan.

Lijkt dit op een
reden die *jij* zou geven?

- | | |
|---|---|
| A. Omdat jouw vriend misschien ook wel eens dingen voor jou heeft gedaan, dus je zou wat voor je vriend moeten doen als je wilt dat hij jou in de toekomst ook helpt. | <input type="checkbox"/> Ja
<input type="checkbox"/> Nee
<input type="checkbox"/> Weet niet |
| B. Omdat het bij vriendschap gaat om wederzijds respect en samenwerking. | <input type="checkbox"/> Ja
<input type="checkbox"/> Nee
<input type="checkbox"/> Weet niet |
| C. Omdat het je vriend is, die misschien een machtig persoon is. | <input type="checkbox"/> Ja
<input type="checkbox"/> Nee
<input type="checkbox"/> Weet niet |
| D. Omdat je veel om je vriend geeft en zou verwachten dat je vriend jou ook zou helpen. | <input type="checkbox"/> Ja
<input type="checkbox"/> Nee
<input type="checkbox"/> Weet niet |
| 1c. Welke van de vier genoemde redenen lijkt het meest op de reden die <i>jij</i> zou geven? | <input type="checkbox"/> A <input type="checkbox"/> B <input type="checkbox"/> C <input type="checkbox"/> D |

REDENEN OM HET LEVEN VAN EEN ONBEKENDE TE REDDEN

- 2a. Stel dat het om een onbekende gaat? Hoe belangrijk is het om alles te doen wat je kunt om het leven van een onbekende te redden, zelfs als dit betekent dat je de wet overtreedt?
- Heel belangrijk
 Belangrijk
 Niet belangrijk
- 2b. Stel dat je een reden zou moeten geven *waarom* het voor jou *belangrijk* is om alles te doen wat je kunt om het leven van een onbekende te redden, zelfs als dit betekent dat je de wet overtreedt? Welke reden zou *jij* geven?

Lijkt dit op een
reden die *jij* zou geven?

- | | |
|--|---|
| A. Omdat je altijd aardig moet zijn. | <input type="checkbox"/> Ja
<input type="checkbox"/> Nee
<input type="checkbox"/> Weet niet |
| B. Omdat het leven kostbaar is en het onmenselijk is om iemand te laten lijden, terwijl zijn leven gered kan worden. | <input type="checkbox"/> Ja
<input type="checkbox"/> Nee
<input type="checkbox"/> Weet niet |
| C. Omdat het leven heel erg veel waard is en de wetgeving hierop gebaseerd zou moeten zijn. | <input type="checkbox"/> Ja
<input type="checkbox"/> Nee
<input type="checkbox"/> Weet niet |
| D. Omdat de onbekende het medicijn nodig heeft en iedereen wil leven. | <input type="checkbox"/> Ja
<input type="checkbox"/> Nee
<input type="checkbox"/> Weet niet |
| 2c. Welke van de vier genoemde redenen lijkt het meest op de reden die <i>jij</i> zou geven? | <input type="checkbox"/> A <input type="checkbox"/> B <input type="checkbox"/> C <input type="checkbox"/> D |