

Sonderdrucke aus
DIAGNOSTICA

Heft 4/1980

Erfassung von moralischem Urteil mit standardisierten Fragebogen

Georg Lind & Roland Wakenhut

1. Einleitung

Es gibt kaum ein Meßinstrument in der Psychologie, das in so engem Dialog mit Theoriebildung und empirischer Forschung entwickelt wurde wie Kohlbergs Interview zur Erfassung der kognitiv-moralischen Entwicklung. Diese Methode wurde seit ihrem ersten Einsatz (K o h l b e r g 1958) bereits zweimal grundlegend revidiert. Sie hat deutliche Belege für eine neue, nicht nur die Moralpädagogik, sondern auch die Sozialisationsforschung herausfordernde Vorstellung von dem Erwerb moralischer Regeln erbracht, worüber es inzwischen zahlreiche Veröffentlichungen von Kohlberg selbst wie auch von anderen Autoren gibt (vgl. K o h l b e r g 1969, L i c k o n a 1976, D ö b e r t et al. 1977, R e s t 1979).

Die Methode ist jedoch nicht unumstritten. Da eine Meßmethode nicht besser sein kann als die immer noch sehr begrenzten und unsicheren Kenntnisse über den komplexen Gegenstand „Moral“, gibt es verständliche Zweifel an der Bedeutung der Meßwerte und an ihrer Fähigkeit, alle wichtigen Komponenten des moralischen Urteils adäquat wiederzugeben. Auch methodenimmanent werden Probleme gesehen, so bei der „Vercodung“ der Interviewantworten und bei der Verrechnung der Codezahlen zu einem Meßwert.

Mit der Konstruktion von Quasi-Interviews mit vorgegebenen Antworten, allgemein als „geschlossene oder standardisierte Fragebogen“ bezeichnet, sollte dieser Kritik begegnet und Instrumente geschaffen werden, die Kohlbergs Interview zu ergänzen und für bestimmte Zwecke zu ersetzen vermögen. Wir wollen in dieser Arbeit die Konstruktionsprinzipien erläutern und diskutieren, die für den Aufbau und Auswertung solcher Fragebogen von Bedeutung sind: Welche Informationen kann man mit diesen Verfahren gewinnen und welche nicht? Wie wird bei der Zusammenstellung der Fragen und Antworten vorgegangen und welche Auswertungsmethoden sind sinnvoll? Und schließlich, welche Kriterien sind vorhanden, an denen die Güte des Fragebogens bei einer Neuentwicklung oder bei Anwendung an neuen Personengruppen geprüft werden kann? Die zentrale Frage nach der Validität, die hier angesprochen ist, lautet: Erfasse ich überhaupt das, was ich erfassen möchte? Es ist offensichtlich, daß es auf diese Frage keine einfache, immer gültige Antwort geben kann. Wir werden zunächst versuchen, die für die Messung von moralischem Urteil bedeutsamen Vorgaben der Theorie Kohlbergs noch einmal zu rekapitulieren, um dann ausgehend von einer Gegenüberstellung der Interviewmethoden und der standardisierten Fragebogen den von uns eingeschlagenen empirischen Zugang eingehend darzustellen.

Die theoretischen Vorgaben lassen sich knapp zusammenfassen:

1. Die sechs von Kohlberg (1969, S. 376) identifizierten Stufen der moralischen Entwicklung implizieren qualitativ unterschiedene Denkformen.

2. Diese Denkformen entwickeln sich in einer invarianten Sequenz. Sie manifestieren sich in einer geordneten Rangreihe der Präferenz für die Stufen und ihrer kognitiven Verstehbarkeit.
3. Jede der unterschiedlichen Denkformen bildet ein strukturelles Ganzes; sie repräsentieren jeweils eine zugrundeliegende Organisation des Denkens.
4. Die Entwicklungsstufen bilden eine Ordnung von zunehmend differenzierteren Strukturen, wobei die höheren Stufen die auf den niedrigeren Stufen bestehenden Strukturen jeweils ersetzen bzw. integrieren.

Insgesamt sind diese Vorgaben recht offen und erlauben keine zwingende Entscheidung zugunsten einer spezifischen Form der Operationalisierung. Um die Entwicklung eines bestimmten Meßinstrumentes zu rechtfertigen, bedarf es anderer, wie durch das Erkenntnisinteresse des Forschers begründeter Kriterien. Geht es darum, die moralische Artikulationsfähigkeit mit zu erfassen, empfiehlt es sich, entsprechend dem Vorgehen von Kohlberg, in offenen Interviews die spontane Produktion von moralischen Argumenten zu registrieren. Akzeptiert man dagegen mit R e s t (1975b, S. 748), daß „in vielen Fällen im Alltag das wichtig ist, was die Leute an moralischen Argumenten wahrnehmen und präferieren, und weniger, welche moralische Argumentationen sie selbst produzieren,“ dann ist auch aus theoretischen Gründen geboten, Reaktionen auf vorgegebene moralische Argumentationen zu erfassen.

2. Interviewmethoden versus Fragebogen

Im Gegensatz zu einer grundsätzlichen methodisch-formalen Behandlung des Themas „offene“ versus „geschlossene“ Verfahren (vgl. K o h l i 1978), gehen wir davon aus, daß sich hinter dieser Gegenüberstellung Probleme verbergen, die sich eher auf der Grundlage von konkreten Fragestellungen beantworten lassen, zumal mit jeder Verfahrensvariante, wie wir oben gesehen haben, eine eigenständige Meßintention verbunden werden kann und das rein „methodische“ Probleme dann zum inhaltlichen wird. Entscheidend für die Validität der Meßwerte ist ihr Bezug auf das damit zu erreichende Verständnis von Natur und Bedeutung des moralischen Bewußtseins.

Eine rein äußerliche Annäherung geschlossener Fragebogen an Kohlbers Interviewmethode, die die zentralen Elemente der Theorie Kohlbergs unreflektiert läßt, kann jedenfalls nicht zu einer Aufhebung der grundsätzlichen Unterschiede zwischen Präferenz und Produktion moralischer Argumente führen, wie S c h u h l e r (1978) zu Recht gegen derartige Versuche einwendet. Es ist vielmehr zu berücksichtigen, daß die Präferenz in einer besonderen Beziehung zur Produktion von verbalen Urteilen steht. So stellt auch K o h l b e r g (1969, S. 387) fest, daß „das Argumentieren einer Person durch zwei sequentielle Ordnungen diktiert wird, durch eine Rangordnung der *Präferenz* und durch eine der *Schwierigkeit*, wobei die modale Stufe eines Individuums die am stärksten präferierte Stufe darstellt, die es leicht benutzen kann“¹).

Welche dieser beiden Komponenten mit welcher Methode besser gemessen wird, läßt sich am besten erkennen, wenn man sich die verschiedenen Operationen der Be-

¹ Übersetzung durch die Verfasser

fragung und Auswertung vor Augen führt. In Kohlbergs Interview, das hinsichtlich Durchführung und Auswertung auch weitgehend standardisiert ist, werden mehrere Dilemmata als kurze Geschichten dargeboten. Diesen folgen jeweils eine Reihe von Nachfragen, die der Befragte selbst mit eigenen Worten zu beantworten und zu begründen hat (Hat die Person X richtig gehandelt? Warum? etc.).

Die von den Befragten gegebenen Antworten werden im nachhinein mit den Beispielantworten („criteria“) des Auswertungshandbuchs Stufe für Stufe verglichen und so für jeden Aspekt bzw. jeden Issue ein Stufenwert bestimmt, aus denen dann durch arithmetische Operationen ein Gesamtstufenwert gebildet wird. Dies heißt, daß die Stufe einer Person nicht einfach aus ihren Antworten abgelesen, sondern über eine Reihe von Zwischenschritten konstruiert wird. Mit dem Meßverfahren von Kohlberg verbinden sich zwei wesentliche Probleme. Es ist offensichtlich, daß durch die eben beschriebenen Operationen die Stufenzuweisung ungenau oder gar arbiträr werden kann. In fast allen uns bekannten Untersuchungen streuen die Antworten einer Person über mehrere Stufen, weisen also nicht einheitlich auf eine Stufe hin. In solchen Fällen hängt die Zuweisung einer Stufe weitgehend von der gewählten Datenaggregation ab. Eine Mittelwertberechnung, wie sie meistens durchgeführt wird, führt in der Regel zu anderen Stufenzuweisungen als die Berechnung eines Median- oder Modalwertes. Der am weitesten verbreitete „Globalwert“, der sich aus der mit der Ordnungszahl der Stufe multiplizierten Summe aller Prozentanteile der Antworten jeder der sechs Stufen ergibt (z.B. $1 \times 15 \% + 2 \times 20 \% + 3 \times 15 \% + 4 \times 30 \% + 5 \times 20 \% + 6 \times 0 \% = 320 = \text{Globalwert}$), entspricht der Mittelwertbildung. Der Vergleich von Global- oder Modalwerten verschiedener Personen oder Untersuchungsgruppen ist auch deshalb problematisch, da durch diese technische Reduktion der Antwortmuster einer Person ganz wichtige Informationen verlorengehen. Vor allem aber entspricht dieses Vorgehen nicht Piagets und Kohlbergs Forderung, die Persönlichkeit des Individuums so umfassend wie möglich zu sehen, nämlich als „strukturelles Ganzes“.

Das zweite Problem, das speziell mit der Interviewmethode verbunden ist, betrifft die Vermengung verschiedener Komponenten des moralischen Urteils. Der üblicherweise berechnete Indikator reflektiert nicht nur die Fähigkeit der Person, auf bestimmten Stufen zu argumentieren, sondern auch seine Motivation, innerhalb dieses Kompetenzbereichs die ihm adäquat erscheinende Argumentation zu benutzen. Überdies hängt die Produktion eigener Argumente von der Artikulationsfähigkeit und der Vertrautheit einer Person mit einer solchen Aufgabenstellung ab. Dies alles kann dazu führen, daß die Stufe des moralischen Bewußtseins bei manchen Personengruppen (z.B. bei Kindern oder bei Erwachsenen mit geringer Bildung) oder bei bestimmten Dilemmata (z.B. „nichtmoralischen“) stark unterschätzt wird.

Mit welcher Erhebungs- oder Auswertungsmethode nun die „wahre Stufe“ erfaßt werden kann, ist angesichts solcher Probleme eher eine müßige Frage. Je nach Fragestellung, Forschungsinteresse und Forschungsstand wird man die eine oder die andere Methode der Erfassung und Stufenzuordnung für die bessere halten oder dazu übergehen, statt nur einem mehrere Parameter zur Kennung der individuellen Spannweite des moralischen Urteils heranzuziehen. Ein Beispiel für eine sehr tiefgehende Auslotung von Feinmechanismen des moralischen Argumentierens sind die Interviewstudien von Dörbert & Nunner-Winkler (1975) (vgl. Nunner-Winkler 1978).

Bei bestimmten Fragestellungen kann es aber methodisch und inhaltlich ein Vorzug sein, daß ein Fragebogen keine Anforderungen an die Artikulationsfähigkeit des Befragten stellt und auch Personen erreicht, die sich nicht gerne interviewen lassen, wenn gleich nun der Befragte vor die Aufgabe gestellt ist, seine Gedanken in standardisierte Beispielsätze zu übersetzen. Bei sorgfältiger Vortestung und durch Offenlassen von Kommentarmöglichkeiten für den Befragten kann aber auch dieses Problem minimiert werden.

Weitere Vorteile von „geschlossenen“ Fragebogen liegen in der durch die Automatisierung von Datenaufnahme und Verrechnung erreichten „Objektivität“.

Hinzu kommt, daß die Intensität der Präferenz für eine bestimmte moralische Orientierung nicht wie im Interview aus der Anzahl ihrer Nennung geschlossen zu werden braucht, sondern durch die Antwortkategorien direkt erfaßt werden kann.

Bei den standardisierten Fragebogen gibt es inzwischen zwei unterschiedliche Vorgehensweisen, was die verlangte Bewertung der vorgegebenen Argumente und die Auswertung betrifft. Der „Defining-Issues-Test“ von R e s t (1975a) beruht auf der Annahme, daß durch die Stufe des moralischen Urteils auch bestimmt ist, wie eine Person ein gegebenes moralisches Dilemma „definiert“. Eine hohe empirische Korrelation von $r = .68$ (R e s t 1975a, S. 80) weist auf einen engen Bezug zu Kohlbergs Interviewverfahren hin.

Eine weitere mögliche Verfahrensweise baut auf Untersuchungen von R e s t , T u r i e l und K o h l b e r g (1969) und R e s t (1973) über den Zusammenhang zwischen der eigenen Produktion von Argumenten und der Präferenz für vorgegebene Argumente auf. Es wird nicht Auskunft über elaborierte, eigene Überlegungen, sondern die Beurteilung vorliegender Argumentationen für und gegen eine vorgegebene Konfliktlösung verlangt. Dieses Design bietet sich an, wenn weiter-differenzierten Fragestellungen bezüglich der individuellen moralischen Urteilsstruktur nachgegangen werden soll, als dieses mit R e s t ' s D-I-T möglich wäre, etwa solchen wie sie von K e a s e y (1975) und D e P a l m a (1975) aufgeworfen wurden: In welchem Maße werden im individuellen Antwortverhalten überhaupt moralische Urteilsstrukturen wirksam? D.h., gibt es einen Zusammenhang zwischen dem Antwortverhalten des Individuums in der Befragungssituation und dem Niveau seines moralischen Bewußtseins?

Man kann davon ausgehen, daß diese Beurteilungsaufgabe alltäglichen Situationen, in denen die Befragten über moralische Probleme diskutieren, recht nahekommt, und in vielen Fällen nach sorgfältiger Analyse auch Rückschlüsse auf die Ebene des moralischen Bewußtseins zuläßt, auf der der Befragte denkt. Im deutschsprachigen Raum sind zwei „Meßinstrumente“ zugänglich, die nach diesen Überlegungen konstruiert werden.

Der „Moralisches-Urteil-Test“ (M-U-T), der ein Akzeptabilitätsurteil verlangt, wurde im Rahmen des Forschungsprojekts „Hochschulsozialisation“²⁾ entwickelt (L i n d e t

²⁾ Dieses Projekt der Deutschen Forschungsgemeinschaft ist Teil des Sonderforschungsbereichs 23 der Universität Konstanz. In ihm arbeiten mit: T i n o B a r g e l , B a r b a r a D i p p e l h o f e r - S t i e m , G e r h i l d F r a m h e i n , G e o r g L i n d , H a n s g e r t P e i s e r t (Leiter), J o h a n n - U l r i c h S a n d b e r g e r und H a n s - G e r h a r d W a l t e r. Die Forschungsgruppe führt gegenwärtig eine multi-national vergleichende Längsschnittuntersuchung durch, die sich mit dem Einfluß der Hochschule auf die Daseins-Vorstellungen und Wert-haltungen von Studenten befaßt.

al. 1976), woraus sich auch einige seiner Besonderheiten erklären (differenzierte Antwortskala, Themenauswahl). Er führt die Bezeichnung „Test“, weil ihm ein experimentelles Design zugrunde liegt, mit dessen Hilfe erst jene komplexen Analysen von individuellen Antwortstrukturen möglich sind, wie sie zur Überprüfung zentraler Hypothesen der Theorie der kognitiv-moralischen Entwicklung notwendig sind. Für den M-U-T wurde ein auf der Quadratsummenzerlegung basierendes Auswertungsprogramm geschaffen (L i n d 1980b), mit dessen Hilfe individuelle Antwortstrukturen analysiert werden können.

Der „Moralisches-Urteil-Fragebogen“ (M-U-F) von K r ä m e r - B a d o n i & W a k e n h u t (1978) stellt eine an die speziellen Untersuchungszwecke der Autoren angepaßte Variante des M-U-T dar, die ursprünglich für das Forschungsprojekt „Sozialisation in der Bundeswehr“ ³⁾ erstellt wurde, inzwischen aber auch in einer Reihe anderer Untersuchungen verwendet wurde. Für den M-U-F wurden neue Dilemmata, ein modifiziertes Antwortschema und besondere Kennwerte entwickelt, letzteres vor allem, um das Phänomen der Segmentierung im moralischen Bewußtsein (D ö b e r t & N u n n e r - W i n k l e r 1975, S e n g e r 1979) meßbar zu machen.

In der Reihenfolge der Arbeitsschritte stellen sich dem Konstrukteur drei Probleme: das der Auswahl und Konstruktion von moralischen Dilemmata, das der Auswahl von moralischen Argumenten und Bewertungsskalen und das der Verrechnung und Deutung individueller Bewertungsmuster, d.h. der Zuweisung von Stufen des moralischen Urteils im Sinne von K o h l b e r g s Schema.

Die hier vorgestellten Verfahren wurden für ganz spezifische Forschungsprobleme entwickelt. Wir wollen dennoch bei der Darstellung der Konstruktionsprinzipien dieser Verfahren auf einige allgemein gültige Erfahrungen hinweisen, die sich bei dem Einsatz der vorliegenden Fragebogen bzw. bei der Adaption ihrer Inhalte an neue Fragestellungen als hilfreich erweisen könnten.

3. Anlage und Konstruktion standardisierter Fragebogen zur Messung des Moralischen Urteils

Die Fragebögen sind folgendermaßen aufgebaut: Eine Kurzgeschichte beschreibt die Handlung einer oder mehrerer Personen in einer Situation, die durch ihre besondere Konstellation als ein moralisches Dilemma gelten kann. Zu der beschriebenen Handlung werden sechs Argumente *dafür* und sechs Argumente *dagegen* vorgegeben, wobei jedes Argument für genau eine Stufe des moralischen Urteils steht. Den Argumenten ist jeweils eine mehrstufige Antwortskala beigelegt, auf der die befragten Personen angeben, wie akzeptabel oder wie berechtigt ihnen das jeweilige Argument erscheint. Im Anhang finden sich zwei exemplarische Dilemmata, die Geschichte A r z t aus dem M-U-T und die Geschichte L ü d d e r s e n aus dem M-U-F. Der Aufbau der Fragebögen beinhaltet eine Reihe von Annahmen und Voraussetzungen, die im folgenden schrittweise diskutiert und erläutert werden.

³⁾ Das Projekt wurde am Sozialwissenschaftlichen Institut der Bundeswehr in München durchgeführt. Projektmitarbeiter waren: T h o m a s K r ä m e r - B a d o n i (jetzt Universität Bremen), Ekkehard Lippert, Paul Schneider (jetzt Staatsinstitut für Schulpädagogik, München) und R o l a n d W a k e n h u t (jetzt Universität Augsburg).

3.1. Das „moralische“ Dilemma

Was macht eine Kurzgeschichte bzw. eine Situationsbeschreibung zu einem Dilemma? Die Frage ist nicht ganz leicht zu beantworten, da ein Dilemma nicht schon „an-und-für-sich“ besteht, und auch nicht jedes Dilemma von jedem in gleicher Weise wahrgenommen wird.

In der Praxis der Fragebogenanwendung nehmen die Befragten zunächst Kenntnis von der Beschreibung des objektiv vorgegebenen Gehalts der Situation mitsamt dem raum-zeitlichen Hintergrund. Diese für alle Probanden gleiche Ausgangsinformation unterliegt der je subjektiven Interpretation und Definition, durch die das Individuum Sinnbezüge in die Situation hineinlegt. Nur wenn diese *Situationsdefinition* nach moralisch-ethischen Kriterien erfolgt, wird die Situation zu einer moralisch relevanten Situation bzw. zu einem moralischen Dilemma, das tatsächlich nach moralischen Kriterien bewertet wird. Um eine solche Situationsdefinition zu ermöglichen bzw. nahezulegen, sind einige Aspekte zu beachten.

Die Situation sollte dem Befragten soweit bekannt sein, daß er sich in die handelnde(n) Person(en) des Dilemmas hineinversetzen kann. Dieser Bedingung genügen in der Regel solche Situationen, die sich auf die Lebens- und Erfahrungswelt des Befragten beziehen und damit zumindest teilweise das Kriterium der *ökologischen Validität* (vgl. hierzu P a w l i k 1976) erfüllen. Bei diesen Situationen kann es sich um selbst erlebte Dilemmata handeln oder auch um solche, die man beobachtet oder durch Schilderung anderer erfahren hat.

Zur empirischen Absicherung der symbolischen Verfügbarkeit der Situation kann am Schluß eines Dilemmas die Frage gestellt werden, ob sich der Befragte tatsächlich in das Dilemma hineinversetzen konnte (vgl. Anhang).

Wenn dieser ersten Forderung Genüge getan ist, und der Befragte sich in die Situation hineinversetzen kann, muß noch erreicht werden, daß die Situation als ein *moralisches* Dilemma wahrgenommen und bewertet wird. Bei weitem nicht jede Umweltkonstellation ermöglicht bzw. erfordert auch eine moralische Thematisierung des Urteils. K o h l b e r g (1976) benennt insgesamt 11 moralische Themen („moral Issues“): Strafe, Eigentum, Gefühl, Autorität, Recht, Leben, Freiheit, Gleichheit, Wahrhaftigkeit, Gewissen, Geschlechtlichkeit.

Die Problemlage wird komplexer, wenn berücksichtigt wird, daß Befragte auf den höheren Stufen mehr Situationen moralisieren als Befragte der unteren Stufen. In der Praxis der Formulierung von Dilemmata können diese Probleme bislang nur approximativ gelöst werden. Neben einer Orientierung an den genannten moralischen Themen („issues“) empfiehlt es sich, zumindest einige der verwendeten Dilemmata so auszuwählen, daß grundsätzlich eine Diskussion auf den postkonventionellen Stufen ermöglicht wird und bei Personen mit postkonventioneller Moralität auch tatsächlich erfordert. Solche Dilemmata sind dann jedenfalls gegeben, wenn in ihnen eine Gefährdung der Menschenwürde und des menschlichen Lebens thematisiert sind.

Schließlich ist es Teil der Forschungsfrage, in welchem Maße ein Dilemma von dem einzelnen Befragten als „moralisches“ aufgefaßt wird. Dazu näheres unten.

Welche weiteren Themen und welche Situation für die Dilemmata ausgewählt werden, läßt sich nicht unabhängig vom Untersuchungsinteresse bestimmen. Steht eine generelle Analyse des moralischen Urteils im Vordergrund, so könnte man versuchen, möglichst alle oben aufgeführten „moral issues“ in den Dilemmata zu thematisieren. Richtet sich das Interesse mehr auf organisations- oder institutionsspezifische Bezüge des moralischen Urteils, so wären genau Situationen aus diesen Kontexten anzusprechen. Soll die Abhängigkeit des moralischen Urteils von spezifischen Lebensbereichen, die sog. *Segmentierung*, untersucht werden, müßten die interessierenden Bereiche auch in den Dilemmata repräsentiert sein. So ist ein Teil der Dilemmata des M-U-F auf die militärische Lebenswelt von wehrpflichtigen Soldaten der Bundeswehr zugeschnitten, während die anderen Dilemmata sich auf die zivile Lebenswelt beziehen.

Auswahl und Zusammensetzung der Dilemmata sind also immer in engem Zusammenhang mit dem jeweiligen Forschungsproblem zu sehen und von daher zu lösen. Die plausibel erscheinende Forderung nach *vollständiger* ökologischer Validität der Situationen, die auf eine *repräsentative* Abbildung der jeweiligen Umwelt zielt, ist kaum zu erfüllen. Solange die Grundgesamtheit der moralisch relevanten Situationen in dem interessierenden Lebensbereich nicht bestimmbar ist, kann auch keine repräsentative Auswahl gezogen werden. Andererseits kann man davon ausgehen, daß sich intuitiv einigermaßen sicher ermitteln läßt, welche Situationen ein „moralisches“ Dilemma darstellen. Dies wird unterstützt durch eine Untersuchung von L i n d (1979). Aufgefordert anzugeben, als welche Art von Problem sie das Arzt-Dilemma (siehe Anhang) ansehen, gaben Abiturienten (N = 538) folgende Antworten:

Definition des Arzt-Dilemmas als ein . . .

– humanitäres	51.3 %
– moralisches	25.8 %
– religiöses	10.6 %
– rechtliches	6.5 %
– gesellschaftliches	5.0 %
– wissenschaftliches Problem	0.7 %

Die Antworten zeigen, daß der alltägliche Sprachgebrauch „moralisch“ enger faßt als wir es hier in Anlehnung an Kohlberg tun.

Die Anzahl der Dilemmata für einen Fragebogen kann nicht allgemein verbindlich festgelegt werden. Es gibt Untersuchungen mit einem, dem Arzt-Dilemma, die belegen, daß bereits eines zu hinreichend aufschlußreichen Einsichten in die moralischen Orientierungen von Befragten führen kann. Der Einfluß situativer Faktoren läßt sich natürlich nur analysieren, wenn mindestens zwei Geschichten verwendet werden.

3.2. Das Argument

Jedes Argument muß sich anhand der bei K o h l b e r g (1969) beschriebenen Merkmale genau einer Stufe zuordnen lassen. In dem Maße, wie sich die neuen Stufen-

beschreibungen von Kohlberg (1976) bewähren, können in Zukunft diese zur Konstruktion von Argumenten herangezogen werden. Es erscheint auf jeden Fall notwendig, durch ein Experten-Rating feststellen zu lassen, ob die Argumente auch hinreichend gut die Stufen repräsentieren, die sie repräsentieren sollen.

Die Auswahl der Argumente sollte unabhängig von der Entscheidungsrichtung des Arguments und seiner Zugehörigkeit zu einer moralischen Urteilsstufe geschehen. D.h. sowohl für jede Entscheidungsrichtung (pro *und* contra) als auch für jede der sechs Stufen sollte (mindestens) ein Argument in den Fragebogen aufgenommen werden, damit die Stufe der moralischen Orientierung unabhängig von der Meinung zu dem moralischen Dilemma bestimmt werden kann. Diese Forderung impliziert die Trennbarkeit von Struktur und Inhalt moralischer Urteile. Sie muß sich darin zeigen, daß es dem Konstrukteur gelingt, sinnvolle Argumente auf allen sechs Stufen des moralischen Urteils zu formulieren. Mit „sinnvoll“ ist hier auch gemeint, daß die Argumente weder lächerlich noch künstlich wirken sollen. Als empirischer Beleg für den Erfolg solcher Konstruktionen kann u.a. gewertet werden, daß die Argumente auf der gleichen Stufe etwa gleich bewertet werden, genauer, daß die Pro-Argumente von Befürwortern ähnlich bewertet werden wie die entsprechenden Contra-Argumente von Gegnern der geschilderten Handlung. Bei den vorliegenden Dilemmata des M-U-T und des M-U-F ist dies annäherungsweise der Fall.

In neueren Arbeiten verweist Kohlberg (1976) darauf, daß bestimmte Urteilmuster, wie z.B. rassistische Vorurteile, auf einer postkonventionellen Ebene nicht vorstellbar sind. Dies verweist auf psychologische Barrieren, die sich auch bei der Fragebogenkonstruktion als große oder gar als unüberwindbare Hindernisse erweisen können.

3.3. Erfassung von Stärke und Richtung der Präferenz

Zu jedem der insgesamt 12 Argumente pro Dilemma wird eine Reaktionsskala vorgegeben, auf der die Befragten angeben, wie akzeptabel ihnen das Argument erscheint, so beim M-U-T, oder wie gut man mit diesem Argument die geschilderte Handlung rechtfertigen bzw. ablehnen kann, so beim M-U-F. Aus dem Vergleich bisheriger Untersuchungen sind keine großen Unterschiede im Antwortverhalten zu erkennen. Ein direkter Vergleich steht noch aus.

Die Anzahl der Reaktionsmöglichkeiten bestimmt sich zunächst aus der Notwendigkeit, *Grade* der Akzeptabilität zu erheben, die in dem Interviewverfahren von Kohlberg nur aus der Häufigkeit der Antworten abgeleitet werden können. Ihre Differenziertheit sollte der angezielten Befragtenpopulation angepaßt sein. Bei Personen mit einiger Fragebogenerfahrung und höherem kognitiv-intellektuellen Niveau können sehr fein abgestufte Reaktionsmöglichkeiten vorgesehen werden, deren Skalencharakter durch Zahlen symbolisiert und verdeutlicht werden kann. Bei Personen, die diese Voraussetzungen nicht mitbringen, sind drei bis fünf Reaktionsmöglichkeiten völlig ausreichend (vgl. Rohrmann 1978). Eine verbale Umschreibung aller oder nur der extremen Reaktionsmöglichkeiten erleichtert in jedem Fall die Beantwortung.

Um die *Meinung* zum in der Geschichte geschilderten moralischen Dilemma zu erfassen, wird die sogenannte Bilanzfrage gestellt. Bei der Auswertung kommt dieser Bilanzfrage eine besondere Bedeutung zu. Von ihrer Beantwortung hängt ab, ob beim einzelnen Befragten die Pro- und Contra-Argumente als „meinungskonform“ oder als „meinungskonträr“ zu klassifizieren sind.

4. Die empirische Zuweisung zur modalen Stufe

Die Auswertung der Fragebögen zum moralischen Urteil zielt auf eine Vielzahl von Kennwerten von Struktur und Inhalt der individuellen Urteilmuster. Diese vielfältigen Auswertungsmöglichkeiten sollen im folgenden aber nur insoweit dargestellt werden, als sie es gestatten, empirische Befunde, die anhand des M-U-T und des M-U-F gewonnen wurden, nachzuvollziehen und einzuordnen. Für eine detaillierte Darstellung von Auswertungsstrategien sei auf Lind (1978) und Wakenhut (1980) verwiesen.

Als Beispiel dient die Matrix der Antworten einer Befragungsperson auf den M-U-F, die aufgrund ihres prägnanten Aufbaus zur Demonstration besonders geeignet ist (Tabelle 1).

Tabelle 1:
Ausgangsdatenmatrix eines Befragten ⁴⁾

	Dilemma ARBEITER		Dilemma SCHNEIDER		Dilemma NEUMANN		Dilemma LÜDDERSEN	
	Pro	Con	Pro	Con	Pro	Con	Pro	Con
Stufe 1	3	3	3	2	2	3	3	2
Stufe 2	2	4	4	3	4	3	3	3
Stufe 3	2	5	5	2	5	2	4	4
Stufe 4	5	5	5	1	5	2	4	5
Stufe 5	3	3	4	2	3	2	3	4
Stufe 6	3	5	4	1	4	2	3	4
Bilanz (Meinung)	eher falsch		eher richtig		eher richtig		eher falsch	
Hineinver- setzen in das Dilemma	ja		ja		ja		etwas	

⁴⁾ Die Werte reichen von der Bewertung „1“ (= ein Argument, mit dem man die Handlung nicht rechtfertigen bzw. nicht ablehnen kann) bis „5“ (= ein sehr gutes Argument, um die Handlung zu rechtfertigen bzw. abzulehnen). Im M-U-T reichen die Präferenzurteile von „-4“ (= völlig unakzeptabel) bis „+4“ (= völlig akzeptabel). Die Dilemmata sind nach den jeweils handelnden Personen benannt.

Mit dem Befragten wollen wir annehmen, daß er sich in alle Dilemmata ausreichend hineinversetzen konnte. Es braucht also kein Dilemma von der Auswertung ausgeschlossen zu werden. Als ein erster Indikator für die modale Stufe läßt sich der höchste Mittelwert der Präferenzscores pro Stufe über die vier Dilemmata ermitteln:

Stufe 1: $M_1 = 2.63$

Stufe 2: $M_2 = 3.25$

Stufe 3: $M_3 = 3.63$

Stufe 4: $M_4 = 4.00$

Stufe 5: $M_5 = 3.00$

Stufe 6: $M_6 = 3.25$

Der höchste Mittelwert findet sich auf Stufe 4 mit $M_4 = 4.00$.

Ein weiterer, entscheidender Indikator ist aus der *Abfolge* der Mittelwerte zu ersehen. Bis hin zur modalen Stufe 4 steigt die mittlere Präferenz der Stufen *monoton* an, um dann auf den darüberliegenden Stufen unregelmäßig bzw. inkonsistent zu verlaufen. Diese *partielle*, bis zur modalen Stufe reichende *Konsistenz* spielt eine besondere Rolle bei der Auswertung. Sie steht im Widerspruch zu der für das Traitkonzept der Persönlichkeitsforschung charakteristischen Annahme der relativen, d.h. interindividuellen Konsistenz, die – übertragen auf die vorliegende Problemstellung – implizieren würde, daß die Befragten über *alle* sechs Stufen hinweg konsistent reagieren müßten. Das würde bedeuten, daß z.B. eine Person mit modaler Stufe 3 auch die Argumente zu den darüberliegenden Stufen konsistent bearbeitet. K o h l b e r g (1974, S. 79) diskutiert freilich ein sog. „nicht-kumulatives Sequenzmodell“, nach dem sich die *Verwendung* der Stufen richten soll. Das Modell sieht vor, daß bis hin zur modalen Stufe die relativen Häufigkeiten, mit denen die Stufen verwendet werden, *monoton* ansteigen, auf der modalen Stufe ihr Maximum erreichen, um dann wieder *monoton* abzunehmen. Diese Annahme ist für die hier interessierenden *Präferenzurteile* reichlich restriktiv, wenn man in Betracht zieht, daß Argumente über der modalen Stufe, auch wenn sie möglicherweise kognitiv gar nicht verarbeitet werden können, mit zunehmendem Abstand zur modalen Stufe immer weniger präferiert werden sollen. Die theoretisch sparsamere Annahme erscheint ausreichend, daß moralische Argumente *unter* der modalen Stufe mit zunehmender Nähe zur modalen Stufe stärker präferiert werden, während für die *über* der modalen Stufe liegenden Argumente keine Annahmen getroffen werden. Präferenzurteile zu den letztgenannten Argumenten enthalten demnach keine oder nur eine sehr bedingt auf die Struktur des moralischen Urteils rückführbare Variabilität. Präferenzen über der modalen Stufe sind nicht vergleichbar mit Präferenzen unter der modalen Stufe. Eine Aggregation von Präferenzurteilen über alle Stufen hinweg ist damit kaum möglich bzw. ist abhängig von der erreichten Moralstufe. Nur bei modaler Stufe 6 geht die *partielle* Konsistenz in eine vollständige Konsistenz der Präferenzurteile über. Die Zuweisung zu einer Stufe der kognitiv-moralischen Entwicklung kann entsprechend auf einer empirischen Konsistenzprüfung der Präferenzurteile aufbauen, wie schon vorher am Verlauf der Präferenzmittelwerte gezeigt wurde.

Entscheidend ist bei allen Möglichkeiten der Konsistenzprüfung, daß immer der *vollständige*, alle Stufen umfassende Präferenzvektor bzw. die Relationen zwischen den Scores *aller* Stufen berücksichtigt werden. Dadurch, daß das individuelle Antwortmuster als „strukturelles Ganzes“ gesehen und diese Sichtweise auch in der Auswertung realisiert wird, hebt sich dieses Vorgehen deutlich von der in der psychologischen Ein-

stellungs- und Persönlichkeitsmessung gängigen Praxis ab, globale *Summenwerte* über die Items – hier vergleichbar mit den Argumenten zu den Stufen – zu bilden und als Meßwerte zu interpretieren. Auch ist dieses idiographische Skalierverfahren stichproben-*unabhängig*. Die Konsistenzprüfung wird *intraindividuell* und nicht *interindividuell* (wie in der Einstellungs- und Persönlichkeitsmessung) vorgenommen.

Bevor weitere Auswertungsschritte an der Beispieldatenmatrix durchgeführt werden, soll an die impliziten Voraussetzungen erinnert werden, auf denen die vorher skizzierte Möglichkeit der Konsistenzprüfung basiert. Es wird nämlich strikte Vergleichbarkeit *aller* Präferenzvektoren unabhängig von der Richtung der Argumente und unabhängig von der Spezifität der Dilemmata unterstellt. Ob und inwieweit diese Voraussetzung von einer individuellen Datenmatrix erfüllt wird, läßt sich mit Hilfe eines dreifaktoriellen, varianzanalytischen Designs abschätzen, das es gestattet, die Variabilität der Präferenzurteile nach den drei wesentlichen varianzstiftenden Faktoren Stufe, Dilemma und die „Meinungsrichtung“⁵⁾ aufzulösen (vgl. Lind 1977). Bei der Spezifikation des dritten Faktors ist zu beachten, das sich zwar die Richtung der Argumente objektiv festmachen läßt, daß sie sich aber in der subjektiven Wahrnehmung des Befragten danach bestimmt, ob die Pro- oder ob die Contra-Argumente dem eigenen Standpunkt entsprechen (Meinungskonformität).

Entscheidend ist der Grad, zu dem sich die Beurteilung der vorgegebenen Argumente an deren Stufenzugehörigkeit orientiert. Der Varianzanteil zu Lasten dieses Faktors gibt Auskunft darüber, inwieweit der Befragte in der Lage ist, moralische Argumente unabhängig von ihrer Meinungskonformität, zur Kenntnis zu nehmen und qualitativ zu unterscheiden. Ergibt sich dagegen ein hoher Varianzanteil für den Faktor Dilemma und für die Wechselwirkung von Faktor Dilemma und Faktor Stufe, so liegt der Schluß nahe, daß die Präferenzurteile stark von der Spezifität der Dilemmata beeinflusst werden. Übersteigen diese Varianzanteile den Faktor Stufe, so ist u.E. die oben genannte Voraussetzung der Vergleichbarkeit nicht mehr erfüllt. Das Vorgehen bei der Konsistenzprüfung ist dann entsprechend zu modifizieren und an Teilmatrizen der Ausgangsmatrix mit vergleichbaren Präferenzvektoren durchzuführen. Dies heißt, daß in dem Maße, wie sich die individuelle Antwortstruktur als Funktion anderer Faktoren als der sechs Typen der moralischen Orientierung darstellt, eine Einstufung nach Kohlbergs Schema gegenstandslos wird.

Die Varianzanteile machen deutlich, daß dem Faktor Meinungskonformität bei dieser Person das relativ größte Gewicht zukommt. Für den Befragten ist also weniger die moralische Qualität der Argumente wichtig als vielmehr die Relation von Inhaltsrichtung der Argumente zum persönlichen Standpunkt. Daraus ergibt sich für die weitere Auswertung die Konsequenz, aus der Ausgangsmatrix zwei Teilmatrizen nach der Meinungskonformität der Argumente zu bilden und die modale Stufe erneut zu ermitteln.

⁵⁾ Wir haben diesen Begriff in früheren Publikationen manchmal synonym mit „Inhaltsrichtung“ gebraucht. Inzwischen hat es sich als notwendig erwiesen, neu zwischen Inhalt und Meinung zu differenzieren und letzteren ausschließlich zur Kennzeichnung eines Aspekts der Urteilmuster zu verwenden.

Zerlegt man nach diesen Ergebnissen die Ausgangsmatrix (Tabelle 1) anhand der Bilanzfrage in eine *meinungskonforme* und in eine *meinungskonträre* Antwortmatrix, so resultieren folgende Matrizen (Tabelle 2):

Zur Illustration seien die Ergebnisse der dreifaktoriellen Varianzanalyse an der Beispieldatenmatrix kurz mitgeteilt:

			Faktor ⁶⁾	Varianzanteil
			Stufe	14.80 %
			Dilemma	4.98 %
			Meinungs- konformität	34.45 %
Wechselwirkung Stufe	X		Dilemma	10.77 %
Wechselwirkung Stufe	X		Meinungs- konformität	14.40 %
Wechselwirkung Dilemma	X		Meinungs- konformität	9.96 %
			Restvarianz	10.63 %

Tabelle 2:
Meinungskonforme Antwortmatrix

	Dilemma ARBEITER Con	Dilemma SCHNEIDER Pro	Dilemma NEUMANN Pro	Dilemma LÜDDERSEN Con	M (Stufe)
Stufe 1	3	3	2	2	2.50
Stufe 2	4	4	4	3	3.75
Stufe 3	5	5	5	4	4.75
Stufe 4	5	5	5	5	5.00
Stufe 5	3	4	3	4	3.50
Stufe 6	5	4	4	4	4.25

⁶⁾ Auf eine nähere Diskussion der verschiedenen Wechselwirkungen zwischen den Hauptfaktoren wird aus Gründen der Übersichtlichkeit verzichtet.

Meinungskonträre Antwortmatrix

	Dilemma ARBEITER Pro	Dilemma SCHNEIDER Con	Dilemma NEUMANN Con	Dilemma LÜDDERSEN Pro	M (Stufe)
Stufe 1	3	2	3	3	2.75
Stufe 2	2	3	3	3	2.75
Stufe 3	2	2	2	4	2.50
Stufe 4	5	1	2	4	3.00
Stufe 5	3	2	2	3	2.50
Stufe 6	3	1	2	3	2.25

Die Konsistenzprüfung anhand des Verlaufs der mittleren Präferenzen weist in beiden Matrizen auf Stufe 4 als modale Stufe hin.

Die meinungskonträre Antwortmatrix signalisiert noch ein anderes Phänomen. Die Präferenzen zu den beiden Dilemmata aus der militärischen Lebenswelt – Dilemma S c h n e i d e r und Dilemma N e u m a n n – sind ab Stufe 3 durchgehend niedriger als bei den beiden anderen nicht-militärischen Dilemmata. Diese Diskrepanz zeigt, daß Gegenargumente auf den konventionellen und postkonventionellen Stufen kaum mehr akzeptiert werden, d.h. moralische Argumente werden für die militärische Lebenswelt weniger präferiert und zugelassen, militärische und nicht-militärische Lebenswelt werden möglicherweise unterschiedlich moralisiert. Der Aufbau der meinungskonträren Antwortmatrix könnte damit als ein Indikator für eine Segmentierung des moralischen Urteils nach militärischem und nicht-militärischem Bereich interpretiert werden. Dieser erste Hinweis auf eine Segmentierung kann in einer zweifaktoriellen Varianzanalyse der meinungskonträren Antwortmatrix weiter verfolgt werden. Der erste Faktor wird durch die moralischen Stufen, der zweite Faktor durch die zu zwei Dilemmagruppen (militärisch vs. nicht-militärisch) zusammengefaßten Dilemmata definiert. Im vorliegenden Fall ergeben sich folgende Varianzanteile:

Faktor	Varianzanteil
Stufe	7.01 %
Dilemma-Gruppe	35.88 %
Wechselwirkung	34.18 %
Restvarianz	22.93 %

Deutlich zeigt sich, daß die moralischen Präferenzen des Befragten sich stärker an der im Dilemma thematisierten Lebenswelt orientieren als an der Stufenzugehörigkeit der Argumente und damit auf Segmentierungsprozesse verweisen. Entscheidend hierfür ist der Varianzanteil, der sich aus der Wechselwirkung von Faktor Stufe und Faktor Dilemma ergibt.

Weiterhin fällt an der meinungskonträren Antwortmatrix dieser Person auf, daß die mittleren Präferenzen, ausgenommen Stufe 1, durchgehend niedriger liegen, wenn auch die gleiche modale Stufe 4 angezeigt wird. Daraus läßt sich ablesen, daß es dem Befragten schwerer fällt, gegen seine eigene Meinung gerichtete Argumente zu präferieren. Im, hier allerdings nicht gegebenen Extremfall, kann dies soweit gehen, daß kein einziges gegenteiliges Argument mehr akzeptiert wird und die Matrix nur mehr den Score 1 aufweist. Ein solches Antwortmuster läßt darauf schließen, daß der Befragte noch nicht in der Lage ist, sich in die Position eines Gegners seiner Meinung hineinzusetzen. Die Fähigkeit aber, sich in den anderen hineinzusetzen, kann als notwendige Voraussetzung für das Erreichen von Stufe 3 und somit als ein zwischen präkonventioneller und konventioneller Moral differenzierendes Merkmal gelten. Das Antwortmuster verweist demnach auf präkonventionelle Tendenzen.

Mit diesen ausgewählten Beispielen für eine Auswertung der Fragebogen sollte nur grundsätzlich gezeigt werden, daß die beschriebenen Datenmatrizen eine Reihe von Auswertungsmöglichkeiten beinhalten, die sich direkt auf die Theorie des moralischen Urteils beziehen lassen. Gerade darin unterscheidet sich die Messung des moralischen Urteils von der Messung anderer psychologischer Konstrukte, wie etwa der Einstellungen, bei denen eine schwer überbrückbare Kluft zwischen theoretischer Konzeptualisierung und empirisch-operationalem Zugang besteht. Die hier vorgetragenen Überlegungen lassen sich bislang nur auf die Ansätze innerhalb der kognitiv-entwicklungstheoretischen Theorietradition übertragen. Als Anwendungsmöglichkeiten wären der Bereich der kognitiv-sozialen Entwicklung (S e l m a n 1976), der der Entwicklung des Ichs (L o e v i n g e r 1976) und andere Bereiche zu nennen, in denen eine vergleichbare Entwicklungslogik formuliert wurde.

5. Offene Fragen

Beide Fragebogen sind bereits in einer Reihe von Untersuchungen eingesetzt worden, um Bedingungen, Verlauf und Ergebnisse von Sozialisationsprozessen zu untersuchen (vgl. L i n d 1978, 1980a; L i n d et al. 1980, K r ä m e r - B a d o n i & W a k e n h u t 1979; W a k e n h u t 1979). Für die Theorie der kognitiv-moralischen Entwicklung bedeutsam ist, daß sich einige der von Kohlberg und seinen Mitarbeitern in Interviewstudien ermittelten Befunde auch durch die vorliegenden Fragebogen haben bestätigen lassen. So konnte die von K o h l b e r g und R e s t (1973) gefundene Präferenzhierarchie sowie die dadurch implizierte Quasi-Simplex-Struktur der Präferenzinterkorrelationen deutlich bestätigt werden (L i n d 1978a; W a k e n h u t 1980). Als zweites wichtiges Ergebnis ließ sich mit dem M-U-T nachweisen, daß der Inhalt der moralischen Orientierung (Richtung und Intensität) empirisch eng mit ihrer Struktur (Konsistenz) verbunden ist – die sechs Typen der moralischen Orientierung somit auch empirisch eine geordnete Skala darstellen.

Diesen die Fruchtbarkeit des Ansatzes bestätigenden Befunden stehen eine Reihe von offenen Fragen und Problemen methodologischer Art gegenüber, die jedoch nicht nur die hier vorgestellten Fragebogen betreffen. Aus der kognitiven Entwicklungstheorie geht hervor, daß es inadäquat wäre, diese Verfahren an den herkömmlichen Güte-

kriterien (etwa der klassischen Testtheorie) zu messen.⁷⁾ Dieses nicht gesehen zu haben, ist die grundlegende Schwäche der Diskussion bei K u r t i n e s & G r e i f (1974) (vgl. B r o u g h t o n 1978) und bei R e i n s h a g e n (1978). Es gibt keine Belege für einen zwischen „wahrer“ moralischer Orientierung und Meßwert vermittelnden Zufallsprozeß und somit auch keine sinnvolle Möglichkeit, die *Reliabilität* zu bestimmen. Vielmehr wird, wie K o h l b e r g (1979) in seiner Entgegnung auf psychometrische Ansätze betont, zunächst von einem „deterministischen“ Modell ausgegangen, in dem „Meßfehler“ als systematisch aufzufassen und theoretisch zu identifizieren sind. Eine Erklärung für „Meßfehler“ wurde z.B. in dem Phänomen der Segmentierung lokalisiert und empirisch nachgewiesen. Daran knüpfen sich eine Reihe offener Probleme, die der weiteren Forschung einen neuen Weg weisen. Um der Gefahr von Zirkelschlüsseln und Exhaurierungen zu entgehen, wird vor allem zu klären sein, in welchem Verhältnis die Segmentierung zu dem Postulat der strukturellen Ganzheit steht. Gewiß ist aber schon jetzt, daß im kognitiv-entwicklungstheoretischen Paradigma Antwortkonsistenz und -inkonsistenz als substantielle Bestandteile der Persönlichkeitsstruktur anzusehen sind und deshalb die Frage nach der Zuverlässigkeit eines Meßverfahrens hier anders gestellt werden muß.

Ähnlich wird mit dem Kriterium der *Validität* zu verfahren sein. Validität des Tests („wie gut mißt der Test, was er messen soll?“) setzt Validität der ihn begründenden Theorie voraus („wie gut kennt der Forscher das, was er messen will?“). Die hier vorgestellten Fragebogen sind in der Lage, hinsichtlich einiger zentraler Hypothesen die empirische Validität der kognitiven Entwicklungstheorie zu bestätigen (vgl. oben). Daraus läßt sich der Schluß ziehen, daß die Tests selbst einigermaßen theoretisch valide sind. Womöglich wird mit B r o u g h t o n (1978) aber noch anders zu argumentieren sein. Um theoretische Positionen prüfen zu können, muß zunächst angenommen werden, daß der theoretisch begründete Test valide ist und Abweichungen von der Prognose zu Lasten der Theorie gehen. Oder die Richtigkeit der Theorie gilt als hinreichend plausibel – dann ist bei der Übersetzung der Theorie in Meßoperationen nach Fehlern zu suchen. Beides läßt sich aber nicht gleichzeitig, sondern nur in Form eines iterativen Vorgehens in der Forschung bewerkstelligen. Es wird deutlich, daß die Validität von Meßverfahren eingebettet ist in die Validität des Forschungs- oder Diagnoseprozesses, innerhalb derer sie eingesetzt werden.

S u m m a r y

This article comparatively discusses methods of measuring moral evaluations in terms of K o h l b e r g 's theory, using nonstructured interviews and standardized questionnaires. The cognitive interests specifically associated with the above-mentioned methods are expounded. The authors then describe the planning and construction of the questionnaires which they developed to measure moral evaluations. The evaluation strategies presented by the authors, based on a theory-oriented type of individual measurement, are discussed in detail. Finally, unsolved problems regarding the reliability and validity of these questionnaires are pointed out.

⁷⁾ Vgl. hierzu die kritische Auseinandersetzung mit der klassischen Testtheorie im Rahmen der Fähigkeitsdiagnostik bei L u m s d e n (1976), K e m p f (1978) und H i l k e (1980).

Literatur

- Broughton, J. 1978. Criticism of the developmental approach to morality. A.P.A. Journal Supplement Abstract Service, November 1978.
- Döbert, R., Habermas, J. & Nunner-Winkler, G. 1977. Entwicklung des Ich. Köln: Kiepenheuer & Witsch.
- Döbert, R. & Nunner-Winkler, G. 1975. Adoleszenzkrise und Identitätsbildung. Frankfurt: Suhrkamp.
- Hilke, R. 1979. Zu den impliziten „anthropologischen“ Voraussetzungen der klassischen Testtheorie. *Diagnostica* 26, 99–111.
- Keasey, C. B. 1975. Implicators of cognitive development for moral reasoning. In: de Palma, D. & Foley, J. (Hg.): *Moral development*. New York: Wiley, S. 39–56.
- Kempf, W. 1978. Kritische Bemerkungen zu impliziten Voraussetzungen psychologischer Testtheorie und -praxis. Universität Konstanz: Hektographiertes Manuskript.
- Kohlberg, L. 1958. The development of modes of moral thinking and choice in the years ten to sixteen. University of Chicago: Unpublished Doctoral Dissertation.
- Kohlberg, L. 1969. Stage and sequence. The cognitive developmental approach to socialization. In: Goslin, D. (Hg.): *Handbook of socialization theory and research*. Chicago: Rand McNally, S. 347–480 (deutsch: Kohlberg, L. 1974).
- Kohlberg, L. 1974. Stufe und Sequenz: Sozialisation unter dem Aspekt der kognitiven Entwicklung. In: Kohlberg, L. (Hg.): *Zur kognitiven Entwicklung des Kindes*. Frankfurt: Suhrkamp, S. 7–255.
- Kohlberg, L. 1976. Moral stages and moralization. In: Lickona, T. (Hg.): *Moral development and behavior*. New York: Holt, Rinehart & Winston, S. 31–53.
- Kohlberg, L. 1979. The meaning and measurement of moral development. Heinz Werner Memorial Lecture, April 1979: manuscript.
- Kohli, M. 1978. „Offenes“ und „geschlossenes“ Interview. *Soziale Welt* 19, 2–25.
- Krämer-Badoni, T. & Wakenhut, R. 1978. Möglichkeiten der Skalierung von moralischem Urteilsverhalten. In: Eckensberger, L. H. (Hg.): *Entwicklung des moralischen Urteilens – Theorie, Methoden, Praxis*. Saarbrücken: Universitätsdruck, S. 379–391.
- Krämer-Badoni, T. & Wakenhut, R. 1979. Moralisches Urteil und politische Einstellungen bei unterschiedlichen sozialen Gruppen. In: Eckensberger, L. H. (Hg.): *Bericht über den 31. Kongreß der Deutschen Gesellschaft für Psychologie in Mannheim 1978*. Göttingen: Hogrefe.
- Kurtines, W. & Greif, E. B. 1974. The development of moral thought: Review and evaluation of Kohlbergs approach. *Psychological Bulletin* 81, 453–470.

- L i c k o n a , T. (Ed.) 1976. *Moral development and behavior*. New York: Holt, Rinehart & Winston.
- L i n d , G., N i e l s e n , A. & S c h m i d t , U. (Hg.) 1976. *Moralisches Urteil und Hochschulsozialisation – Materialien, Beiträge*. Universität Konstanz: Arbeitsunterlage 40 des Projekts Hochschulsozialisation.
- L i n d , G. 1977. Über die Grundlagen einer integrationistischen Meßtheorie zur Erforschung des moralischen Urteils. Konstanz: Arbeitsunterlage 41 des Projekts Hochschulsozialisation.
- L i n d , G. 1978. Wie mißt man moralisches Urteil? Probleme und alternative Möglichkeiten der Messung eines komplexen Konstrukts. In: P o r t e l e , G. (Hg.): *Sozialisation und Moral*. Weinheim: Beltz, S. 171–201.
- L i n d , G. 1979. *Einstellung zur Euthanasie*. Universität Konstanz: Unveröffentlichtes Manuskript.
- L i n d , G. 1980a. Toward an evaluation of the impact of university on students' moral development. In: T r o t s e n b u r g , E. (Hg.): *Higher education – A field of study*, Vol. 3. Frankfurt: Lang (in press).
- L i n d , G. 1980b. Intraindividuelle Varianzkomponentenzerlegung der Antwortmuster des „Moralisches-Urteil-Test“ (Algorithmus für das Programmsystem K o s t a s). Universität Konstanz: Arbeitsunterlage 65 des Projekts Hochschulsozialisation.
- L i n d , G., S a n d b e r g e r , J.-U. & B a r g e l , T. 1980. Moral judgement, ego-strength and democratic orientations. Some theoretical contiguities and empirical findings. Universität Konstanz: Arbeitsunterlage 64 des Projekts Hochschulsozialisation.
- L o e v i n g e r , J. 1976. *Ego development*. San Francisco: Jossey-Bass.
- L u m s d e n , J. 1976. Test theory. *Annual Review of Psychology* 27, 251–280.
- N u n n e r - W i n k l e r , G. 1978. Probleme bei der Messung des moralischen Urteils mit standardisierten Verfahren. In: E c k e n s b e r g e r , L. H. (Hg.): *Entwicklung des moralischen Urteilens – Theorie, Praxis, Methoden*. Saarbrücken: Universitätsdruck, 337–358.
- D e P a l m a , D. J. 1975. Research and theory in moral development: A commentary. In: D e P a l m a , D. J. & F o l e y , J. M. (Eds.) *Moral development. Current theory and research*. New York: Wiley
- P a w l i k , K. 1976. Ökologische Validität: Ein Beispiel aus der Kulturvergleichsforschung. In: K a m i n s k i , G. (Hg.) *Umweltpsychologie*. Stuttgart: Klett.
- R e i n s h a g e n , H. 1978. Forderungen für die Messung des moralischen Urteils mit standardisierten Verfahren. *Diagnostica* 24, 137–145.
- R e s t , J. R. 1973. The hierarchical pattern of moral judgment: A study of patterns of comprehension and preference of moral stages. *Journal of Personality* 41, 86–109.

- R e s t , J. R. 1975a. Recent research on an objective test of moral judgment: How the important issues of a moral dilemma are defined. In: de P a l m a , D & F o l e y , J. (Hg.): Moral development. New York: Wiley, 75–93.
- R e s t , J. R. 1975b. The validity of tests of moral judgment. In: M e y e r , J. R. et al. (Hg.): Values education. Theory, practice, problems, prospects. Waterloo, Ont., Canada: Laurier.
- R e s t , J. R. 1979. Development in judging moral issues. Minneapolis: University of Minnesota Press.
- R e s t , J. R., T u r i e l , E. & K o h l b e r g , L. 1969. Level of moral development as a determinant of preference and comprehension of moral judgment made by others. *Journal of Personality* 37, 225–252.
- R o h r m a n n , B. 1978. Empirische Studien zur Entwicklung von Antwortskalen für die sozialwissenschaftliche Forschung. *Zeitschrift für Sozialpsychologie* 9, 222–245.
- S c h u h l e r , P. 1978. Probleme bei der Erfassung des moralischen Urteils mit offenen und geschlossenen Verfahren. In: S i l b e r e i s e n , R. (Hg.): Newsletter Soziale Kognition 2, 162–168.
- S e l m a n , R. 1976. Social-cognitive understanding: A guide to educational and clinical behavior. In: L i c k o n a , T. (Hg.): Moral development and behavior. New York: Holt, Rinehart & Winston, 299–316.
- S e n g e r , R. 1979. Formen der Segmentierung des moralischen Bewußtseins bei Bundeswehrsoldaten. In: L i n d , G. (Hg.): Bericht über den Workshop „Moralische Entwicklung und soziale Umwelt“ in Konstanz 1979. Universität Konstanz: Zentrum I Bildungsforschung.
- W a k e n h u t , R. 1979. Effects of military service on political socialization of draftees. *Armed Forces and Society* 5, 626–641.
- W a k e n h u t , R. 1980. Messung von moralischem Urteil mit dem Moralisches-Urteil-Fragebogen (M-U-F). München: Unveröffentlichtes Manuskript.

Anschrift der Verfasser:

Prof. Dr. Dr. Roland Wakenhut
WiSo-Fakultät der Universität Augsburg
Memminger Straße 14
8900 Augsburg

Dipl.-Psych. Georg Lind
Zentrum I Bildungsforschung
Universität Konstanz
Postfach 55 60
7750 Konstanz

Anhang

Geschichte „LÜDDERSEN“ (M-U-F)

Der Bauer Lüddersen, der seine Felder in der Nähe des geplanten Atomkraftwerkes bei X. hat, ist Mitglied einer Bürgerinitiative gegen den Bau von Kernkraftwerken. Während einer Demonstration, zu der die Bürgerinitiative aufgerufen hat, kommt es zu einer Besetzung des Baugeländes. Obwohl es dabei zu Sachschäden auf dem Baugelände und zu handfesten Schlägereien mit den Sicherungskräften kommt, nimmt auch Lüddersen daran teil.

Mit welchem der folgenden Argumente, die man für das Verhalten von Lüddersen vorbringen kann, läßt sich Ihrer persönlichen Meinung nach das Verhalten von Lüddersen am besten rechtfertigen?

Dieses Argument rechtfertigt das Verhalten von Lüddersen

Man kann vorbringen

	sehr gut	ziemlich gut	gut	weniger gut	gar nicht
1. daß er in gutem Glauben gehandelt hat, weil auch alle seine Nachbarn mitgegangen sind					
	5	4	3	2	1
2. daß er mitgegangen ist, weil er befürchtete, daß seine Felder durch den Bau des Kernkraftwerkes an Wert verlieren					
3. daß er sich nicht klar darüber war, daß er mit seiner Teilnahme an der Besetzung Recht und Ordnung verletzte					
4. daß er die Besetzung trotz ihrer Rechtswidrigkeit für angemessen hielt, weil durch ein Atomkraftwerk die Gesundheit der anässigen Bevölkerung z.B. im Falle eines Unfalles einen unabsehbaren Schaden davontragen könnte					
5. daß er keine rechtliche Möglichkeit sah, gegen den Bau von Kernkraftwerken zu protestieren und den Staat zu einer Diskussion dieser Problematik zu zwingen					
6. daß er sich bei der Vielzahl der Besetzer in Sicherheit fühlte					

Mit welchem der folgenden Argumente, die man *gegen* das Verhalten von Lüddersen bringen kann, läßt sich Ihrer persönlichen Meinung nach eine Ablehnung des Verhaltens von Lüddersen am besten rechtfertigen?

Dieses Argument rechtfertigt eine *Ablehnung* des Verhaltens von Lüddersen

Man kann vorbringen

1. daß er mit seiner Teilnahme an der Besetzung eine ganze Reihe von Gesetzen übertreten hat
2. daß er sich damit anders verhalten hat als die meisten Bauern in der Umgebung
3. daß er auch dann, wenn er den Bau von Kernkraftwerken für falsch hält, kein Recht hat, an gesetzeswidrigen Aktionen teilzunehmen, solange die rechtlichen Mittel gegen die Baumaßnahmen nicht völlig ausgeschöpft sind
4. daß er wegen seiner Teilnahme an der Besetzung bestraft werden wird
5. daß er nicht hätte teilnehmen dürfen, da es sich nicht um eine gewaltfreie Aktion handelte, Gewalt aber kein Mittel der Auseinandersetzung zwischen verschiedenen Meinungen sein kann
6. daß er vermutlich weniger Schere-reien gehabt hätte, wenn er an der Besetzung nicht teilgenommen hätte

sehr gut	ziemlich gut	gut	weniger gut	gar nicht
5	4	3	2	1

--	--	--	--	--

--	--	--	--	--

--	--	--	--	--

--	--	--	--	--

--	--	--	--	--

Nun sagen Sie uns bitte noch, was Sie von der geschilderten Situation halten:

Können Sie sich in die Situation von Lüddersen hineinversetzen?

ja	etwas	nein
3	2	1

Glauben Sie, daß eine solche Situation in der Wirklichkeit vorkommen kann?

ja	vielleicht	nein
3	2	1

Halten Sie das Verhalten von Lüddersen eher für richtig oder eher für falsch? Versuchen Sie bitte, sich für *eine* der beiden Möglichkeiten zu entscheiden!

eher richtig	eher falsch
2	1

wir Euthanasie von Mord unterscheiden können, darf sich keiner am Leben anderer vergreifen.

10. weil der Arzt sich damit eine Menge Unannehmlichkeiten zuziehen kann. Andere sind dafür schon empfindlich bestraft worden.
11. weil er es hätte wesentlich leichter haben können, wenn er gewartet und nicht in das Sterben der Frau eingegriffen hätte.
12. weil der Arzt gegen das Gesetz verstoßen hat. Wenn man Zweifel bezüglich der Rechtmäßigkeit der Euthanasie hat, dann darf man solchen Bitten nicht nachgeben.

-4	-3	-2	-1	0	+1	+2	+3	+4

--	--	--	--	--	--	--	--	--

--	--	--	--	--	--	--	--	--

--	--	--	--	--	--	--	--	--

Wenn Sie keines der angeführten Argumente für völlig akzeptabel halten, wie würden Sie selbst Ihre Beurteilung des Arztverhaltens begründen?