

Methoden der Entwicklungspsychologie

Edgar Erdfelder, Christian Rietz und Georg Rudinger

Mit Baltes, Reese und Nesselroade (1977, S. 4) kann man den Gegenstand der Entwicklungspsychologie wie folgt charakterisieren: „Developmental psychology deals with the description, explanation, and modification (optimization) of intraindividual change in behavior across the life span, and with interindividual differences (and similarities) in intraindividual change“. Diese Aufgabenbeschreibung beinhaltet einerseits die Forderung nach der Identifizierung von Formen, Sequenzen und Mustern des Entwicklungsgeschehens und andererseits die Forderung nach Aufdeckung zugrundeliegender Mechanismen, Ursachen und Gesetze, so daß interindividuelle Unterschiede und Gemeinsamkeiten in der Entwicklung adäquat beschrieben und erklärt werden können. Will man diese Ziele erreichen, so erscheint es naheliegend, die interessierenden Untersuchungseinheiten wiederholt, d.h. *längsschnittlich* zu beobachten. Eine solche ausschließliche Fokussierung auf den Längsschnitt verkennt jedoch, daß Beschreibung und Erklärung Forderungen sind, die man sinnvoll nur an entwicklungspsychologische Theorien, nicht aber an Methoden im allgemeinen und Designs und Daten im speziellen richten kann. Daher werden in diesem Kapitel nicht nur längsschnittliche Methoden, sondern auch andere Verfahren behandelt, die potentiell einen Beitrag zur Theorieüberprüfung in der Entwicklungspsychologie leisten können. Dieses Kapitel bietet keine Einführung in Datenerhebungs- und Auswertungsmethoden (vgl. hierzu z.B. Baltes et al., 1977; Trautner, 1992), sondern es sollen vor allem neuere Methodenentwicklungen skizziert werden, die für die Entwicklungspsychologie von Interesse sein können.

Quer- und längsschnittliche Designs können im Bezugsrahmen *Personen* \times *Zeit* \times *Variablen* systematisch dargestellt werden. Bei einer einfachen Querschnittstudie werden mehrere Stichproben (Angehörige verschiedener Altersgruppen) zu einem Zeitpunkt ($T = 1$) bezüglich einer oder mehrerer Variablen untersucht ($V = 1$ oder $V > 1$). Bei einer einfachen Längsschnittstudie wird dagegen eine Stichprobe zu mehreren Zeitpunkten (z.B. im Abstand von fünf Jahren) wiederholt untersucht ($T > 1$). Während Querschnitte zwangsläufig auf mehreren beobachteten Personen basieren ($N > 1$), können längsschnittliche Daten für mehrere ($N > 1$) oder für nur eine Person ($N = 1$) erhoben werden. Eine weitere Differenzierung kann schon für den Fall $N = 1$ über die Anzahl der Variablen vorgenommen werden (vgl. Nesselroade, 1977): *measuring change (or differences)* bei einer Variablen ($V = 1$), *structuring change (or differences)* bei mehreren Variablen ($V > 1$).

Quer- wie längsschnittliche Daten bieten keinen direkten Blick auf genuine Entwicklungsprozesse, da querschnittliche Daten von *Kohorteneffekten* und längsschnittliche von *Meßzeitpunkt-* sowie *Retest-Effekten* überlagert sein können (z.B. Baltes et al., 1977). Welchen Wert die verschiedenen Designs zur Prüfung von Entwicklungs-

theorien haben, richtet sich danach, inwieweit diese Störeinflüsse theoriespezifisch geschätzt werden können, so daß eine Überprüfung von Hypothesen anhand der jeweiligen Datenbasis möglich wird (Kendler, 1979). Unter Umständen kann sogar ein Design mit einmaliger Erhebung ($T = 1$) einer altershomogenen Stichprobe seine Berechtigung haben, wenn nämlich eine Entwicklungstheorie ein bestimmtes Ergebnismuster für Personen eines bestimmten Alters zu einem bestimmten Zeitpunkt impliziert.

Die o.g. Taxonomie beinhaltet nicht nur die quantitative Dichotomie „1 versus > 1 “ für jede Facette, sondern auch qualitative Aspekte, wie abhängige versus unabhängige Stichproben, diskretes versus kontinuierliches Zeitkonzept, diskrete versus kontinuierliche Variablen, fehlerfreie versus fehlerbehaftete Variablen etc. Die methodologischen Implikationen dieser Aspekte werden im folgenden behandelt.

1 Personen, Variablen und Zeit

Der Bezugsrahmen *Personen* \times *Zeit* \times *Variablen* definiert ein hypothetisches Universum aller möglichen Personen, Meßzeitpunkte und Variablen (vgl. Cattell, 1988), aus welchem Beobachtungen gewonnen werden. Bei diesem Vorgehen kommt es, und zwar unabhängig vom spezifischen Design, zu *multimodalen Selektionseffekten*. Zunächst ist hier der Selektionseffekt der freiwilligen Teilnahme zu nennen. Bei Längsschnittstudien kommt unter anderem selektive Stichprobenminderung hinzu, die dann als systematischer Fehler (Schäffer, in diesem Band) zu werten ist, wenn sie kein Abbild des zugrundeliegenden Populationsprozesses ist.

Auch bei der Auswahl und Definition der Variablen sind Selektionsaspekte zu beachten: Ist Variable Y_1 oder Y_2 der validere Indikator für ein Konstrukt? Welche Konsequenzen hat die Entscheidung für die eine oder andere manifeste Variable? Damit sind Probleme der Konstruktvalidität angesprochen, wie sie schon von Cook und Campbell (1979) unter den beiden Aspekten *cause construct validity* (Validität der unabhängigen Variablen, UV) und *effect construct validity* (Validität der abhängigen Variablen, AV) erwähnt wurden. Mit diesen Überlegungen ist die Unterscheidung in manifeste (beobachtete) und latente (theoretische) Variablen verbunden. Validität bezieht sich hier auf die Angemessenheit der postulierten Verknüpfung zwischen manifesten und latenten Variablen, Reliabilität auf die Enge dieser Beziehung; je weniger eng diese Beziehung, desto größer der Meßfehler.

Zeit ist die wesentliche Variable im Rahmen entwicklungspsychologischer Forschung bzw. jeglicher Forschung, die sich mit Veränderung beschäftigt (vgl. Rudinger & Wood, 1990). Die Anzahl von und die Abstände zwischen Meßzeitpunkten, die aus dem Universum möglicher Meßzeitpunkte ausgewählt werden, sollten über eine *causal lag theory* theoretisch begründbar sein (vgl. Heise, 1975). Natürlich ist der Meßzeitpunkt ebenso wie Alter und Kohorte eine Stellvertretervariable für eine Vielzahl von konkreten Einflüssen. Die Wahl gleicher Abstände zwischen den Meßzeitpunkten erscheint auf diesem Hintergrund oft willkürlich. Mit dem Status der Zeit als „unabhängige“ Designvariable (Wie beeinflusst „Zeit“ die Entwicklungsprozesse?) ist die Spezifikation des Zeitbegriffes als diskret, ordinal oder metrisch eng verknüpft. Das gleiche gilt für die Betrachtung von Zeit als AV (Wann tritt ein Entwicklungsphänomen auf, wie lange ist die Zeitdauer eines Entwicklungsprozesses?).

Im folgenden werden ausgewählte Kombinationen der Facetten *Personen*, *Zeit* und *Variablen* behandelt, wobei wir zunächst querschnittliche Designs und danach längsschnittliche Designs und Analysemethoden in den Mittelpunkt rücken.

2 Querschnittliche Designs

Finden sich in einer Querschnittstudie Altersgruppenunterschiede, so kann zunächst nicht entschieden werden, ob hierfür das Lebensalter oder die Kohortenzugehörigkeit (z.B. via kohortenspezifischer Schulbildung) verantwortlich ist. Analog kann man bei Längsschnittstudien nicht ohne weiteres entscheiden, ob beobachtete Veränderungen mit dem Lebensalter oder dem Meßzeitpunkt zusammenhängen. Schaie (1965) und Baltes (1968) haben Lösungsvorschläge für beide Probleme unterbreitet. Diese laufen darauf hinaus, *Querschnittsequenzen* (mehrere Querschnittuntersuchungen zu verschiedenen Meßzeitpunkten) bzw. *Längsschnittsequenzen* (mehrere Längsschnittuntersuchungen mit Probanden verschiedener Kohorten) zu betrachten. Die neueste Darstellung dieser Designs und damit verbundener Probleme gibt Schaie (1994).

Wendet man geeignete statistische Modelle auf Quer- oder Längsschnittsequenzen an, so können die primär interessierenden genuinen Entwicklungsprozesse von Kohorten- oder Meßzeitpunkteffekten potentiell getrennt werden. Dies sei exemplarisch anhand der simultanen Schätzung von Alters-, Kohorten- und Meßzeitpunkteffekten im Rahmen von Querschnittsequenzen demonstriert. Das scheinbar naheliegende Modell der dreifaktoriellen Varianzanalyse mit den Haupteffekten Alter, Kohorte und Meßzeitpunkt sowie deren Interaktionen kann dazu nicht verwendet werden, da die drei UVn linear abhängig sind: Das Modell ist in dieser Form nicht identifizierbar. Die von Schaie (1965) in Erwägung gezogene Auswertung über mehrere zweifaktorielle Varianzanalysen (Alter \times Kohorte, Alter \times Meßzeitpunkte, Kohorte \times Meßzeitpunkte) ermöglicht keine unverfälschte Schätzung der Effekte (vgl. Adam, 1978; Schaie & Hertzog, 1982). Man muß daher entweder – was theoretisch unbefriedigend ist – jeweils eine der drei Effektquellen „ignorieren“ oder aber dreifaktorielle Varianzanalysen mit bestimmten Restriktionen berechnen, was erstmals von Mason, Mason, Winsborough und Poole (1973) vorgeschlagen wurde.

Mason et al. (1973) gehen von einer additiven Beeinflussung der AV Y durch die drei Faktoren A (Alter), Z (Meßzeitpunkt) und K (Kohorte) aus, so daß für den bedingten Erwartungswert der a -jährigen Personen zum Zeitpunkt z

$$E(Y|A = a, Z = z, K = k) = f(a) + g(z) + h(k) \quad (1)$$

gilt. Hierbei ist $k = z - a$ das Geburtsjahr (die Kohorte) der a -jährigen Personen; f , g und h bezeichnen unbekannte reellwertige Funktionen. Rechnerisch ist es günstig, das additive Modell (1) als Regressionsmodell mit Dummy-Variablen als UVn darzustellen. Die Dummy-Variablen für eine exemplarische Querschnittsequenz mit drei Altersstufen (20-, 40-, 60-Jährige), drei Meßzeitpunkten (1960, 1980, 2000) und folglich fünf Kohorten (1900, 1920, 1940, 1960, 1980) veranschaulicht Tabelle 1. Wendet man ein multiples Regressionsmodell mit diesen acht Dummy-Variablen als Prädiktoren auf eine in der genannten Querschnittsequenz erhobene AV Y an, so entspricht das unstandardisierte Regressionsgewicht b_1 des Prädiktors X_1 gerade dem von additiven Meßzeitpunkt- und Kohorteneffekten bereinigten Y -Mittelwertsunterschied

TABELLE 1. Dummy-Kodierung zum additiven Alter \times Meßzeitpunkt \times Kohorten-Modell von Mason et al. (1973) (leicht verändert übernommen aus Erdfelder, 1987, S. 177).

UV _n			Dummy-Variablen								
			für A			für Z			für K		
A	Z	K	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	
20	1960	1940	0	0	0	0	0	1	0	0	
40	1960	1920	1	0	0	0	1	0	0	0	
60	1960	1900	0	1	0	0	0	0	0	0	
20	1980	1960	0	0	1	0	0	0	1	0	
40	1980	1940	1	0	1	0	0	1	0	0	
60	1980	1920	0	1	1	0	1	0	0	0	
20	2000	1980	0	0	0	1	0	0	0	1	
40	2000	1960	1	0	0	1	0	0	1	0	
60	2000	1940	0	1	0	1	0	1	0	0	

zwischen 40- und 20-Jährigen und b_2 dem zwischen 60- und 20-Jährigen. Analog repräsentieren die Regressionsgewichte b_3 und b_4 Meßzeitpunkteffekte der Jahre 1980 und 2000 relativ zum Bezugsmeßzeitpunkt 1960 und b_5 bis b_8 Kohorteneffekte der Kohorten 1920 bis 1980 relativ zur Bezugskohorte 1900. Die additive Regressionskonstante b_0 ist der (bereinigte) Y -Mittelwert der 20-Jährigen und legt zusammen mit b_1 und b_2 die „reine“, unverfälschte Entwicklungsfunktion $f(A) = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2$ in Modell (1) fest. Zu beachten ist allerdings, daß b_0 von der willkürlichen Wahl der Bezugskohorte und des Bezugsmeßzeitpunktes abhängt, während b_1 und b_2 hierdurch unbeeinflusst bleiben.

Das geschilderte additive Modell ist letztlich ein restringiertes dreifaktorielles varianzanalytisches Modell *ohne* Interaktionseffekte. Selbst dieses Modell weist noch linear abhängige Prädiktoren auf und ist deshalb nicht identifizierbar. Um ein identifizierbares Modell zu bekommen, muß mindestens eine (inhaltlich fundierte) Zusatzannahme gemacht werden, die z.B. die Fixierung eines Regressionsgewichts auf einen bestimmten Wert erlaubt (z.B. Nullsetzen von b_5 , gleichbedeutend mit der Annahme identischer Effekte der Kohorten 1900 und 1920) oder die Gleichsetzung von Regressionsgewichten (z.B. $b_7 = b_8$, d.h. gleiche Effekte der Kohorten 1960 und 1980). Denkbar ist auch, die Menge der „erlaubten“ Entwicklungsfunktionen auf bestimmte Funktionstypen einzuschränken, z.B. auf

$$f(A) = b_0 + b_1 \cdot \ln(A) - b_2 \cdot (\ln(A))^2, \quad b_1 > 0, \quad b_2 > 0. \quad (2)$$

Dieser Funktionstyp läßt sich auf dem Hintergrund der G_f - G_c -Theorie der Intelligenzentwicklung (Cattell, 1971; Horn, 1982) begründen, wenn als AV_n Intelligenztestrohwerte untersucht werden (Erdfelder, 1987). Setzt man (2) in (1) ein, erhält man ohne zusätzliche Annahmen über die Gleichheit bestimmter Kohorten und/oder

Meßzeitpunkteffekte ein identifizierbares Modell, sofern mehr als drei Altersgruppen pro Meßzeitpunkt untersucht werden. Wenn die Anzahl der untersuchten Altersgruppen groß ist, kann man dieses Modell sogar um Wechselwirkungseffekte zwischen Alter und Meßzeitpunkt sowie Alter und Kohorte erweitern (Erdfelder, 1987). Kohorten- und Meßzeitpunkteffekte müssen also bei hinreichend restringierter Entwicklungsfunktion nicht notwendig als additiv angenommen werden, sie können sich z.B. im Alter von 60 anders als im Alter von 20 Jahren auswirken.

Das in diesem Abschnitt für quantitative AVn dargestellte Vorgehen zur Separierung von Alters-, Meßzeitpunkt- und Kohorteneffekten läßt sich auch auf kategoriale AVn im Rahmen log-linearer Modelle übertragen (Fienberg & Mason, 1978). Auch auf Längsschnittsequenzen läßt sich ein analoges Verfahren anwenden, wobei allerdings die durch wiederholte Messungen erzeugte Abhängigkeit der Daten zu berücksichtigen ist. Praktisch ist dies durch Einbeziehung der Personenvariable in das Modell möglich (vgl. Pedhazur, 1977).

3 Längsschnittliche Designs

Das einfachste Längsschnittdesign umfaßt eine Variable ($V = 1$), die an vielen Personen ($N > 1$) zu zwei Meßzeitpunkten ($T = 2$) erhoben wird (Y_{t1} und Y_{t2}). An diesem Design werden elementare methodische Probleme, die bei allen längsschnittlichen Designs auftreten, illustriert. Dieses *prä-experimentelle Design* erlaubt eine konzeptuelle Trennung der Facetten von Konstanz und Wandel, jedoch noch keine eindeutige *Interpretation* von Unterschieden und Veränderungen. Wir werden zunächst den Fall dieses „kleinsten“ Längsschnitts behandeln, um dann auf Längsschnitte mit mehr als zwei Meßzeitpunkten einzugehen.

3.1 Facetten von Konstanz und Wandel

Prinzipiell können die Mittelwerte für die wiederholt untersuchte Stichprobe zu beiden Meßzeitpunkten gleich oder unterschiedlich, die Varianzen ($Var(Y_{t1})$ und $Var(Y_{t2})$) homogen oder heterogen und die Korrelationen zwischen den beiden Meßzeitpunkten (r_{t1t2} , häufig als *Test-Retest-Korrelationen* bezeichnet) hoch oder niedrig sein. Behalten Personen ihre relative Position über die Zeit bei, so ist r_{t1t2} positiv und hoch. Diese Korrelation spiegelt also das Ausmaß der Persistenz interindividueller Differenzen in der beobachteten Variablen wider. Die drei Parameter (Mittelwert, Varianz und Korrelation) können *unabhängig* voneinander in ihren Werten variieren. Jede erdenkliche Kombination von Parameterwerten ist möglich.

Potentielle Störeinflüsse wie Meßzeitpunkt- oder *Retest*-Effekte wirken sich auf die Parameter unterschiedlich aus. Unterstellt man z.B. fixierte additive Störeffekte (gleicher meßzeitpunktabhängiger Störeinfluß für alle Beobachtungseinheiten), so sind lediglich die Mittelwerte, nicht aber die anderen o.g. Parameter verfälscht.

Geht man weiterhin davon aus, daß Ergebnismuster wie die gerade beschriebenen auf der Ebene beobachtbarer Variablen angesiedelt sind und es eine theoretisch „wahre Welt“ (latente Variablen) hinter diesen beobachtbaren Mustern (manifeste Variablen) gibt, so wird schon in dem einfachen Fall zweier Meßzeitpunkte die Situation dadurch komplex, daß vier Facetten zur Beschreibung der Veränderung dieses

„Mini-Systems“ über die Zeit berücksichtigt werden müssen: (1) Die Unterscheidung in manifeste Variablen und ihnen zugrundeliegende latente Variablen, (2) die Korrelation auf manifester Ebene einerseits und latenter Ebene andererseits, (3) die Varianzen auf beiden Ebenen und (4) die Mittelwerte auf beiden Ebenen. Zur Verdeutlichung dieser Facetten: Daß die beobachtete Varianz vom ersten zum zweiten Meßzeitpunkt größer wird, kann durchaus auf eine fächerförmige Entwicklung auf der latenten Ebene zurückführbar sein und muß nicht einem Anwachsen des Meßfehlers angelastet werden.

3.2 Stabilität und Reliabilität

Die Korrelation zwischen zwei Meßzeitpunkten, d.h. die Korrelation der Variablen Y_{t1} und Y_{t2} (r_{t1t2}), wird häufig als Indikator für *Stabilität* betrachtet.¹ Da mit „Stabilität“ aber im Regelfall bestimmte Eigenschaften eines Entwicklungsprozesses angesprochen werden sollen, sollte man diesen Begriff besser auf Ebene der latenten bzw. „wahren“ und *nicht* der beobachtbaren Werte ansiedeln. Stabilität wäre somit als Korrelation (ρ_{t1t2}) zwischen den latenten Variablen η_{t1} und η_{t2} zu definieren, die den beobachtbaren Variablen Y_{t1} und Y_{t2} zugrunde liegen. Diese Korrelation spiegelt das Ausmaß der Persistenz interindividueller Differenzen in den latenten Variablen wider. Reliabilität (hier $rel(Y_{t1})$ bzw. $rel(Y_{t2})$) bezieht sich auf die Qualität der Messung der untersuchten Konstrukte zum jeweiligen Meßzeitpunkt (vgl. Rietz, Rudinger & Andres sowie Stumpf, in diesem Band). Bezüglich der Korrelation der beobachteten Variablen muß also gefragt werden, ob sie eine angemessene Abbildung des Entwicklungsprozesses ist oder ob die Schlußfolgerung auf diesen Prozeß durch mangelnde Reliabilitäten verfälscht ist. Als eine Konsequenz dieser Überlegungen führt eine „Dekomposition“ der Korrelation von Y_{t1} und Y_{t2} zu

$$r_{t1t2} = \rho_{t1t2} \cdot \sqrt{rel(Y_{t1}) \cdot rel(Y_{t2})} . \quad (3)$$

Die *Test-Retest*-Korrelation ist also eine Kombination aus Stabilität und Reliabilität. Eine niedrige Korrelation schließt weder perfekte Reliabilität noch perfekte Stabilität aus. $r_{t1t2} = 0.20$ kann z.B. über $1.0 \cdot \sqrt{0.2 \cdot 0.2}$ (perfekte Stabilität, niedrige Reliabilität) oder $0.2 \cdot \sqrt{1.0 \cdot 1.0}$ (niedrige Stabilität, perfekte Reliabilität) zustande kommen. Es sind alle Kombinationen von Reliabilität und Stabilität möglich, die zum gleichen Produkt führen. Diese Konfundierung ist für den Fall $T = 2$ und $V = 1$ nur dann auflösbar, wenn man *a priori* Informationen über Stabilität oder Reliabilität besitzt. Für $T > 2$ und $V = 1$ können Stabilität und Reliabilität allerdings aus den Daten geschätzt werden.

Die konzeptuelle Trennung von theoretischen Annahmen über latente Variablen einerseits und meßfehlerbelasteten beobachtbaren Variablen andererseits ist nicht an quantitative Variablen gebunden: Bei kategorialen Variablen beispielsweise kann auch zwischen theoretischen Annahmen über den Entwicklungsprozeß (z.B. Stabilität, Synchronizität) und *Klassifikationsfehlern* unterschieden werden (Erdfelder, 1990; Rudinger, Chaselon, Zimmermann & Henning, 1985).

¹Man beachte, daß das Symbol r hier und im folgenden nicht für eine Korrelation in der Stichprobe steht, sondern ein Parameter ist, der die Korrelation beobachtbarer Variablen in der zugrundeliegenden Population bezeichnet.

3.3 Differenzwerte

Zur Beschreibung des Ausmaßes von Veränderungen zwischen zwei Meßzeitpunkten wird häufig das „natürliche Maß für Veränderungen“ (Lord, 1963), nämlich der Differenzwert ($D = Y_{t2} - Y_{t1}$) verwendet. Die Vorbehalte gegen Differenzwerte als reliable und valide Indikatoren für Veränderung in statistischen Analysen sind so alt wie die Differenzwerte selbst (Cronbach & Furby, 1970; Harris, 1973) und spiegeln unterschiedliche Aspekte des gleichen Sachverhaltes wider: Die Ungenauigkeiten der einzelnen Beobachtungen werden in deren Differenzwerte „transportiert“. Die Reliabilität von Differenzwerten entspricht dem folgenden Quotienten (Lord, 1963, S. 32):

$$\frac{\text{Var}(Y_{t1}) \cdot \text{rel}(Y_{t1}) + \text{Var}(Y_{t2}) \cdot \text{rel}(Y_{t2}) - 2 \cdot \sqrt{\text{Var}(Y_{t1}) \cdot \text{Var}(Y_{t2})} \cdot r_{t1t2}}{\text{Var}(Y_{t1}) + \text{Var}(Y_{t2}) - 2 \cdot \sqrt{\text{Var}(Y_{t1}) \cdot \text{Var}(Y_{t2})} \cdot r_{t1t2}} \quad (4)$$

Sie ist also eine nichtlineare Funktion von fünf Veränderlichen. Praktisch alle der mit der Differenzbildung verbundenen und in der Vergangenheit so intensiv diskutierten Schwierigkeiten rühren daher, daß eine wesentliche Informationsquelle außer acht gelassen wird, nämlich die *Varianzen* der eingehenden Meßwerte. Berücksichtigt man diese und trägt der Unterscheidung zwischen manifesten und latenten Variablen Rechnung, so stellt sich heutzutage die Bewertung des Differenzenansatzes wie auch der Meßfehlerproblematik generell (*Ausgangswertproblem*, *Regressionseffekt*) differenzierter dar (vgl. Rogosa, 1988).

3.4 Simplex-Modelle

Im Rahmen eines Längsschnittes mit mehr als zwei Meßzeitpunkten ist man in der Lage, die im vorherigen Kapitel beschriebenen Konzepte in testbare Modelle zu überführen und zu numerischen Schätzungen zu gelangen. Dies leisten beispielsweise für quantitative Variablen lineare Strukturgleichungsmodelle (vgl. Rietz, Rudinger & Andres, in diesem Band); für diskrete Variablen existieren analoge Modellansätze (vgl. Collins & Cliff, 1990; Hamerle & Ronning, 1995; Langeheine & van de Pol, 1990). Ein lineares Strukturgleichungsmodell, mit dem man die in Kapitel 3.2 dargestellte Trennung von Stabilität und Reliabilität leisten kann, ist das *Quasi-Markoff-Simplex-Modell*, in dem ein autoregressiver Prozeß erster Ordnung auf Ebene der latenten Variablen formuliert wird (vgl. Abbildung 1). Dieses Modell ist dadurch charakterisiert, daß es auf theoretischer Ebene die Gerichtetheit der Zeit berücksichtigt, auf empirischer die Meßfehler. Die individuellen latenten Anteile von Meßwerten zum Meßzeitpunkt t werden durch latente Anteile zum Meßzeitpunkt $t - 1$ vorhergesagt. Damit erhält man eine „Kette“ von Stabilitätskoeffizienten, die eine Theorie über die Kovarianzmatrix der wiederholten zeitgebundenen Beobachtungsvariablen impliziert. Dies kann auch unter der zusätzlichen Annahme invarianter Stabilitäten (Stationarität) bezüglich des Entwicklungsprozesses für gleiche Zeiteinheiten bei gleichen Reliabilitäten geschehen (Einzelheiten bei Rudinger & Rietz, 1993).

In derartigen Modellen können weiterhin auch Hypothesen über längsschnittliche Veränderungen von Varianzen der latenten Variablen (z.B. Entwicklung als Ausdifferenzierung) und sogar über (latente) Mittelwerte formuliert und getestet werden (McArdle & Epstein, 1987; Rudinger, 1995; Rudinger, Andres & Rietz, 1994).

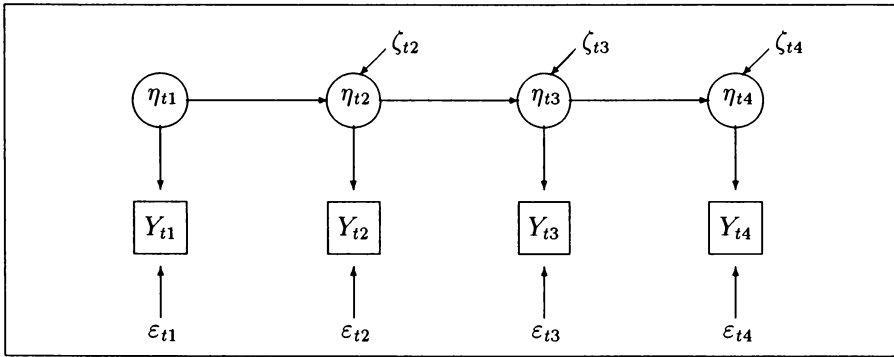


ABBILDUNG 1. Quasi-Markoff-Simplex über vier Meßzeitpunkte.

3.5 „Cross-lagged panel correlations“ und Pfadanalyse

Ziel der Analyse von *cross-lagged panel correlations* (CLPC) ist es, auf der Basis von Korrelationen Aussagen über die „kausale Ordnung“ zweier Variablen ($V = 2$), die zu (mindestens) zwei Meßzeitpunkten ($T = 2$) an derselben Stichprobe wiederholt erhoben wurden, zu machen. Wenn zwei Variablen zu zwei Meßzeitpunkten erhoben werden (X_{t1} , X_{t2} , Y_{t1} und Y_{t2}), ergeben sich sechs Korrelationen: zwei synchrone Korrelationen ($r(X_{t1}, Y_{t1})$, $r(X_{t2}, Y_{t2})$), zwei Autokorrelationen ($r(X_{t1}, X_{t2})$, $r(Y_{t1}, Y_{t2})$) sowie zwei „verzögerte Kreuzkorrelationen“, die *cross-lagged correlations* ($r(X_{t1}, Y_{t2})$, $r(Y_{t1}, X_{t2})$). Während synchrone Korrelationen und Autokorrelationen in der CLPC-Technik den „interpretativen Kontext“ bestimmen, sind die *cross-lagged correlations* von zentraler Bedeutung für die Beantwortung der Frage nach *kausaler Priorität* (vgl. Rudinger & Bierhoff, 1980). X wird genau dann als kausal dominant gegenüber Y bezeichnet, wenn $r(X_{t1}, Y_{t2}) > r(Y_{t1}, X_{t2})$.

Auch Vertreter der CLPC-Technik betonen, daß diese Interpretation die Gültigkeit einer Reihe z.T. restriktiver Randbedingungen voraussetzt (Synchronizität, Stationarität, Stabilität, hohe und gleiche Reliabilität; vgl. dazu Kenny, 1975), die in der Praxis selten geprüft werden. Ferner schließt auch die Gültigkeit der Annahmen *nicht* aus, daß es starke Alternativhypothesen gibt, z.B. die eines allen Variablen zu allen Meßzeitpunkten zugrundeliegenden gemeinsamen Faktors (Kenny, 1975). Vor allem aber ist einzuwenden, daß das Konzept der „kausalen Priorität“ unter wissenschaftslogischen Gesichtspunkten ungeklärt bleibt. Rogosa (1980) hat beispielsweise gezeigt, daß kein Bezug zwischen „kausaler Priorität“ und „kausalen Effekten“ im Sinne von Pfadanalysen besteht, wenn man von einem Pfadmodell ausgeht, bei dem X_{t2} und Y_{t2} jeweils linear durch X_{t1} und Y_{t1} beeinflusst werden. Auch bei Gültigkeit der CLPC-Voraussetzungen ist ein Schluß von den *cross-lagged correlations* auf die entsprechenden Pfadkoeffizienten nicht möglich.

Da es sich bei klassischen Pfadanalysen um Systeme linearer multipler Regressionsgleichungen handelt, braucht auf diese Techniken an dieser Stelle nicht weiter eingegangen werden (vgl. hierzu Opp & Schmidt, 1976). Das der Pfadanalyse zugrundeliegende Modell ist klar definiert, die Interpretation des an die Größe von Pfadkoeffizienten gebundenen Kausalitätskonzepts ist eindeutig, und die Anwendung ist an weniger Voraussetzungen gebunden. Deshalb sind Pfadanalysen der CLPC-Technik

vorzuziehen. Die Pfadanalyse geht allerdings wie die CLPC-Technik von meßfehlerfreien (perfekt reliablen) Variablen aus. Diese Annahme ist häufig unrealistisch, und deshalb empfiehlt sich eine Behandlung der Problematik im allgemeineren Rahmen linearer Strukturgleichungsmodelle, die Pfadmodelle als Spezialfälle umfassen (vgl. Rudinger & Rietz, 1995; Bentler, Wu & Houck, in diesem Band).

3.6 Strukturen

Häufig interessieren im Rahmen multivariater Längsschnittstudien *Veränderungen von Strukturen* und die *Struktur von Veränderungen*. Eine mögliche Frage betrifft die Invarianz der untersuchten Konstrukte bzw. Entwicklungsdimensionen über die Zeit; eine andere Frage bezieht sich auf die Validität der Meßinstrumente für verschiedene Kohorten und/oder Altersstufen bzw. Zeitpunkte. Diese Fragen können unter dem Begriff der *faktoriellen Invarianz* zusammengefaßt werden, was gleiche Faktoren pro Meßzeitpunkt, „ähnliche“ Ladungen pro Faktor und Meßzeitpunkt und „ähnliche“ Beziehungen zwischen den Konstrukten pro Meßzeitpunkt (vgl. Horn, 1991; Rietz, 1995) beinhaltet.

In den genannten Aspekten können sich natürlich auch Veränderungen über die Zeit ergeben: Die Anzahl der Faktoren kann sich verändern (Differenzierung vs. Integration), die Ladungen können bei bedeutungsgleichen Konstrukten unterschiedlich hoch sein, ein Meßfehlermodell kann sich verändern und auch die Beziehung zwischen den Konstrukten. Das eigentliche Strukturproblem wird über die Beziehung zwischen den Konstrukten bzw. die Veränderung dieser Beziehung über die Zeit abgebildet. Invarianz der Konstrukte über die Meßzeitpunkte ist nicht gleichbedeutend mit Stabilität: Invarianz und Stabilität sind genauso unabhängige Konzepte wie Stabilität und Reliabilität. Das Problem von Invarianz und Strukturveränderung stellt sich noch komplexer dar, wenn z.B. im Rahmen von zahlreichen Entwicklungstests die *gleiche* Entwicklungsdimension durch altersspezifische Meßprozeduren erfaßt werden muß. Die meisten mit struktureller Veränderung bzw. Invarianz zusammenhängenden Fragestellungen lassen sich mit linearen Strukturgleichungsmodellen behandeln (Hertzog & Schaie, 1986, 1988; Rudinger, Andres & Rietz, 1994). Diese setzen natürlich eine explizite *Theorie über den Entwicklungsprozeß* voraus, die festlegt, was als Invarianz, als Veränderung und als Fehler zu betrachten ist.

Fragen nach Struktur und Invarianz stellen sich auch im Rahmen „qualitativer“ Entwicklungstheorien wie Stadien- und Stufentheorien, deren empirische Überprüfung in der Regel über (geordnete) kategoriale Variablen erfolgt. Analog zu quantitativen Veränderungen gibt es eine Reihe qualitativer Entwicklungsmodelle wie z.B. synchrone, divergente oder konvergente Progression, horizontale oder vertikale Verschiebung (*decalage*) oder reziproke Interaktion (van den Daele, 1969). Quantitativ-strukturelle Veränderungen haben im qualitativen Paradigma ihre Entsprechung vor allem in der Art der Verknüpfung von Entwicklungssequenzen: Diese Beziehungen können z.B. disjunktiv, konjunktiv oder kompensatorisch sein. Ansätze zur empirischen Überprüfung dieser Entwicklungsmodelle finden sich bei Andersen (1994), Erdfelder (1990), Formann (1994), Henning und Rudinger (1985), Meiser (1993), Meiser, Hein-Eggers, Rompe und Rudinger (in press) und Schröder, Edelstein und Hoppe-Graff (1994).

4 Weiterführende Literatur

Es konnten nur einige von uns für wichtig erachtete Methodenentwicklungen kurz skizziert werden. Detailliertere und vertiefende Darstellungen lassen sich folgenden Büchern entnehmen: Magnusson und Bergman (1990), Magnusson, Bergman, Rudinger und Törestad (1994), Nesselroade und Baltes (1979) und von Eye und Clogg (1994). Für die Entwicklungspsychologie wichtige Methoden behandeln auch die Kapitel von Krauth, von Langeheine und Rost sowie von Rost und Erdfelder (in diesem Band). Das Buch von Wohlwill (1973) ist immer noch empfehlenswert, da entwicklungspsychologische Theorien, Konzepte, Modelle und Methoden integrativ dargestellt werden.

Literaturverzeichnis

- Adam, J. (1978). Sequential strategies and the separation of age, cohort, and time-of-measurement contributions to developmental data. *Psychological Bulletin*, 85, 1309–1316.
- Andersen, E. B. (1994). Longitudinal studies for discrete data based on latent structure models. In D. Magnusson, L. R. Bergman, G. Rudinger & B. Törestad (Eds.), *Problems and methods in longitudinal research: Stability and change* (pp. 308–322). Cambridge: Cambridge University Press.
- Baltes, P. B. (1968). Longitudinal and cross-sectional sequences in the study of age and generation effects. *Human Development*, 11, 145–171.
- Baltes, P. B., Reese, H. W. & Nesselroade, J. R. (1977). *Life-span developmental psychology: Introduction to research methods*. Monterey: Brooks/Cole.
- Cattell, R. B. (1988). The data box: Its ordering of total resources in terms of possible relational systems. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 69–130). Chicago: Rand-McNally.
- Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.
- Collins, L. M. & Cliff, N. (1990). Using the longitudinal Guttman simplex as a basis for measuring growth. *Psychological Bulletin*, 108, 128–134.
- Cook, T. D. & Campbell, D. T. (1979). *Quasi-experimentation: Design & analysis issues for field settings*. Boston: Houghton Mifflin.
- Cronbach, L. J. & Furby, L. (1970). How should we measure „change“ — or should we? *Psychological Bulletin*, 74, 68–80.
- van den Daele, L. D. (1969). Qualitative models in developmental analysis. *Developmental Psychology*, 1, 303–310.
- Erdfelder, E. (1987). *Die Entwicklung psychometrischer Intelligenz über die Lebensspanne. Aspekte eines allgemeinspsychologischen Zugangs*. Frankfurt: Lang.
- Erdfelder, E. (1990). Deterministic developmental hypotheses, probabilistic rules of manifestation, and the analysis of finite mixture distributions. In A. von Eye (Ed.), *Statistical methods in longitudinal research*, Vol. II (pp. 471–509). New York: Academic Press.
- von Eye, A. & Clogg, C. C. (Eds.) (1994). *Latent variables analysis. Applications for developmental research*. Thousand Oaks: Sage.
- Fienberg, S. E. & Mason, W. M. (1978). Identification and estimation of age-period-cohort models in the analysis of discrete archival data. In K. F. Schuessler (Ed.), *Sociological methodology 1979* (pp. 1–67). San Francisco: Jossey-Bass.

- Formann, A. K. (1994). Measuring change using latent class analysis. In A. von Eye & C. Clogg (Eds.), *Latent variables analysis. Applications for developmental research* (pp. 294–312). Thousand Oaks: Sage.
- Hamerle, A. & Ronning, G. (1995). Panel analysis for qualitative variables. In G. Arminger, C. C. Clogg & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 401–451). New York: Plenum.
- Harris, C. W. (Ed.) (1963). *Problems in measuring change*. Madison: University of Wisconsin Press.
- Heise, D. R. (1975). *Causal analysis*. New York: Wiley.
- Henning, H. J. & Rudinger, G. (1985). Analysis of qualitative data in developmental psychology. In J. R. Nesselroade & A. von Eye (Eds.), *Individual development and social change: Explanatory analysis* (pp. 295–341). Orlando: Academic Press.
- Hertzog, C. & Schaie, K. W. (1986). Stability and change in adult intelligence: 1. Longitudinal covariance structures. *Psychology and Aging, 1*, 159–171.
- Hertzog, C. & Schaie, K. W. (1988). Stability and change in adult intelligence: 2. Simultaneous analysis of longitudinal means and covariance structures. *Psychology and Aging, 3*, 122–130.
- Horn, J. L. (1982). The aging of human abilities. In B. B. Wolman (Ed.), *Handbook of developmental psychology* (pp. 847–870). Chicago: Rand-McNally.
- Horn, J. L. (1991). Comments on „issues in factorial invariance“. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 114–125). Washington: American Psychological Association.
- Kendler, T. S. (1979). Cross-sectional research, longitudinal theory, and a discriminative transfer ontology. *Human Development, 22*, 235–254.
- Kenny, D. A. (1975). Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin, 82*, 887–903.
- Langeheine, R. & van de Pol, F. (1990). Veränderungsmessung bei kategorialen Daten. *Zeitschrift für Sozialpsychologie, 21*, 88–100.
- Lord, F. M. (1963). Elementary models for measuring change. In C. W. Harris (Ed.), *Problems in measuring change* (pp. 21–38). Madison: University of Wisconsin Press.
- Magnusson, D. & Bergman, L. R. (Eds.) (1990). *Data quality in longitudinal research*. Cambridge: Cambridge University Press.
- Magnusson, D., Bergman, L. R., Rudinger, G. & Törestad, B. (Eds.) (1994). *Problems and methods in longitudinal research: Stability and change*. Cambridge: Cambridge University Press.
- Mason, K. O., Mason, W. M., Winsborough, H. H. & Poole, W. K. (1973). Some methodological issues in cohort analysis of archival data. *American Sociological Review, 38*, 242–258.
- McArdle, J. J. & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development, 58*, 110–133.
- Meiser, T. (1993). Loglinear Rasch models: Theory and applications in longitudinal research. *Berichte aus dem Psychologischen Institut der Universität Bonn, 19*(2), 1–37.
- Meiser, T., Hein-Eggers, M., Rompe, P. & Rudinger, G. (in press). Analyzing homogeneity and heterogeneity of change using Rasch and latent class models: A comparative and integrative approach. *Applied Psychological Measurement*.
- Nesselroade, J. R. (1977). Issues in studying developmental change in adults from a multivariate perspective. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 59–69). New York: Van Nostrand Reinhold.
- Nesselroade, J. R. & Baltes, P. B. (Eds.) (1979). *Longitudinal research in the study of behavior and development*. New York: Academic Press.

- Opp, K.-D. & Schmidt, P. (1976). *Einführung in die Mehrvariablenanalyse*. Reinbek: Rowohlt.
- Pedhazur, E. J. (1977). Coding subjects in repeated measures designs. *Psychological Bulletin*, 84, 298–305.
- Rietz, C. (1995). *Faktorielle Invarianz*. Bonn: PACE.
- Rogosa, D. (1980). A critique of cross-lagged correlation. *Psychological Bulletin*, 88, 245–258.
- Rogosa, D. (1988). Myths about longitudinal research. In K. W. Schaie, R. T. Campbell, W. Meredith & S. C. Rawlings (Eds.), *Methodological issues in aging research* (pp. 171–210). New York: Springer.
- Rudinger, G. (1995). Strukturgleichungsmodelle in der Entwicklungspsychologie. In R. Oerter & L. Montada (Hrsg.), *Entwicklungspsychologie* (3. Auflage, S. 1177–1190). Weinheim: Psychologie Verlags Union.
- Rudinger, G., Andres, J. & Rietz, C. (1994). Structural equation models for studying intellectual development. In D. Magnusson, L. R. Bergman, G. Rudinger & B. Törestad (Eds.), *Problems and methods in longitudinal research: Stability and change* (pp. 274–307). Cambridge: Cambridge University Press.
- Rudinger, G. & Bierhoff, H. W. (1980). Quasi-experimentelle Versuchspläne für die Markt- und Kommunikationspsychologie. In K. D. Hartmann & K. F. Köppler (Hrsg.), *Fort-schritte der Marktpsychologie*, Band 2 (S. 135–163). Frankfurt: Fachbuchhandlung für Psychologie.
- Rudinger, G., Chaselon, F., Zimmermann, E. & Henning, H. J. (1985). *Qualitative Daten. Neue Wege sozialwissenschaftlicher Methodik*. München: Urban & Schwarzenberg.
- Rudinger, G. & Rietz, C. (1993). Und mal wieder: Reliabilität und Stabilität in Paneldellen. *ZUMA-Nachrichten*, 32, 60–75.
- Rudinger, G. & Rietz, C. (1995). Intelligenz – Neuere Ergebnisse aus der Bonner Längs-schnittstudie des Alterns (BOLSA). In A. Kruse & R. Schmitz-Scherzer (Hrsg.), *Psy-chologie der Lebensalter* (S. 185–200). Darmstadt: Steinkopff.
- Rudinger, G. & Wood, P. K. (1990). N's, times, and number of variables in longitudinal research. In D. Magnusson & L. R. Bergman (Eds.), *Data quality in longitudinal research* (pp. 157–180). Cambridge: Cambridge University Press.
- Schaie, K. W. (1965). A general model for the study of developmental problems. *Psycho-logical Bulletin*, 64, 92–107.
- Schaie, K. W. (1994). Developmental designs revisited. In S. H. Cohen & H. W. Reese (Eds.), *Life-span developmental psychology: Methodological contributions* (pp. 45–64). Hillsdale: Erlbaum.
- Schaie, K. W. & Hertzog, C. (1982). Longitudinal methods. In B. B. Wolman (Ed.), *Hand-book of developmental psychology* (pp. 91–115). Englewood Cliffs: Prentice Hall.
- Schröder, E., Edelstein, W. & Hoppe-Graff, S. (1994). Qualitative analyses of individual differences in intraindividual change: Examples from cognitive development. In D. Ma-gnusson, L. R. Bergman, G. Rudinger & B. Törestad (Eds.), *Problems and methods in longitudinal research: Stability and change* (pp. 166–189). Cambridge: Cambridge University Press.
- Trautner, H. M. (1992). *Lehrbuch der Entwicklungspsychologie, Band 1: Grundlagen und Methoden* (2. Auflage). Göttingen: Hogrefe.
- Wohlwill, J. F. (1973). *The study of behavioral development*. New York: Academic Press.