

# Die unveränderliche Veränderbarkeit und der unterschiedliche Unterschied – Varianz nachweisen mit Invarianz

Michael Rochnia<sup>1</sup>, Falk Radisch<sup>2</sup>

<sup>1</sup>Bergische Universität Wuppertal, <sup>2</sup>Universität Rostock

## Zusammenfassung

Die Grundidee von Veränderungs- und Unterschiedsmessung ist, dass Mittelwertdifferenzen zwischen Messzeitpunkten oder Gruppen die Variabilität eines Konstrukts abbilden. Dies bedarf der Messinvarianz. Das Fehlen von Messinvarianz kann jedoch selbst Veränderung oder Unterschied anzeigen. Daraus resultieren grundlegende epistemologische Probleme: Die Vorbedingung für die Messung von Veränderung kann selbst auch Veränderung bedeuten. Wir erläutern diese Problematik und erinnern an eine Lösung.

## La variabilité invariante et la différence différente – Prouver la variance avec l'invariance

### Résumé

L'idée de base de la mesure des changements et des différences est que les différences moyennes entre les temps de mesure ou les groupes reflètent la variabilité d'un construit. Cela nécessite une invariance de mesure. Cependant, l'absence d'invariance de la mesure peut elle-même indiquer un changement ou une différence. Il en résulte des problèmes épistémologiques fondamentaux: La condition préalable à la mesure du changement peut elle-même signifier le changement. Nous expliquons ce problème et rappelons une solution.

## The unchangeable changeability and the different difference – Prove variance with invariance

### Abstract

The idea of change- and difference-measurement is that mean-differences between measurement-times, or groups, reflect the variability of a construct. This requires measurement-invariance. The absence of measurement-invariance, however, can mean the change of the construct. Epistemological problems result from this: The precondition for change may imply change, but it also renders it untestable. We explain this problem and remind of a solution.

## 1 Einleitung

Seit der von Roth (1962) geforderten empirischen Wende hat empirisches Arbeiten in der Pädagogik ihr Nischendasein aus den Zeiten von Lay (1908) und Meumann (1911) verlassen. Empirische Forschung im pädagogischen Bereich ist seit mindestens 20 Jahren kein Novum und kein Sonderfall mehr (vgl. Deutsches PISA-Konsortium, 2001). Empirische Zugriffe auf Erziehung- und Bildungsprozesse sind speziell im Rahmen der empirischen Bildungsforschung prominent geworden (McElvany, Gebauer & Gräsel, 2019). Als interdisziplinäres Forschungsfeld hat es sich die empirische Bildungsforschung zur Aufgabe gemacht, praktische Probleme im Bildungssystem zu bearbeiten (Lenzen, 2019). Die Ursache für diese Trends ist nicht zuletzt die zunehmende Ausrichtung des Bildungsbereichs an Evidenz (Bellmann, 2016; Radtke, 2016). Zwar umfasst die empirische Bildungsforschung aufgrund ihrer Interdisziplinarität zahlreiche theoretische und methodische Zugriffe (Tippelt & Schmidt-Hertha, 2018), es dominieren aber quantitative Zugriffe (Keiner, 2011). Im Rahmen quantitativer Zugriffe sind Veränderungs- und Unterschiedsmessungen ein zentraler Bestandteil des Methodenrepertoires der empirischen Bildungsforschung und werden in einschlägiger Methodenliteratur thematisiert (vgl. Backhaus, Erichson, Plinke & Weiber, 2016; Baur & Blasius, 2014; Döring & Bortz, 2016; Bortz & Schuster, 2010; Rasch, Friese, Hofmann & Naumann, 2014a; Rasch, Friese, Hofmann & Naumann, 2014b). Parallel dazu werden Veränderungs- und Unterschiedsmessungen immer wieder zunehmend in Form von längsschnittlich angelegten (Quasi)Experimenten gefordert, da viele Konstrukte und theoretischen Überlegungen im Bildungsbereich gegenstandsbedingt dynamischen oder prozessualen Charakter haben und Wirkungen bestimmt werden sollen (z.B. Klieme & Steinert, 2008; Gräsel, 2011). Insofern sind nicht nur Veränderungs- und Unterschiedsmessung entsprechend der geäußerten Forderungen wichtig, sondern eben auch eine über die Zeit hinweg stabile und trotzdem veränderungs- und unterschiedssensitive Erfassung zentraler Konstrukte.

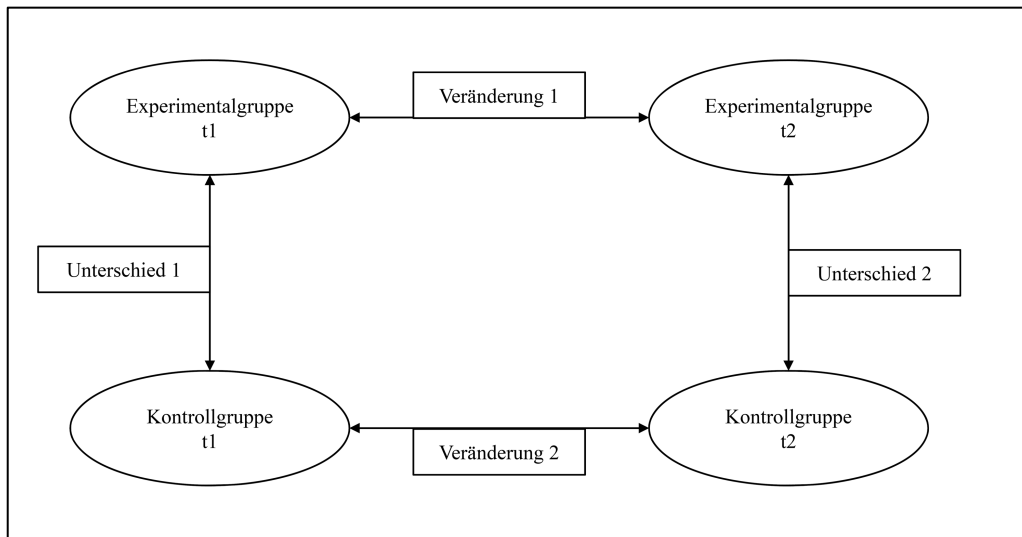
Etablierte Veränderungs- und Unterschiedsmessungen bergen jedoch einige methodische Vorbedingungen, welche grundlegende epistemische Probleme mit sich bringen. Im vorliegenden theoretischen Beitrag widmen wir uns diesem Phänomen, welches bereits in den 1980er Jahren diskutiert wurde (Armenakis & Bedeian, 1982; Bedeian, Armenakis & Gibson, 1980; Terborg, Howard & Maxwell, 1980). Diese Diskussion um Veränderungs- und Unterschiedsmessungen findet in der aktuellen Bildungsforschung unserer Ansicht nach jedoch kaum Berücksichtigung. Dies dient uns als Anlass dazu, die Diskussion erneut aufzugreifen. Der nächste Abschnitt erläutert das Konzept der Veränderungs- und Unterschiedsmessung. Daran anschließend leiten wir epistemische Probleme dieses methodischen Vorgehens ab und greifen zum Abschluss bekannte Möglichkeiten auf, mit der Problematik umzugehen.

## 2 Veränderungs- und Unterschiedsmessungen und ihre Vorbedingungen

Das Interesse der empirischen Bildungsforschung bezieht sich in der Regel auf Konstrukte, also nicht direkt beobachtbare Phänomene (Döring & Bortz, 2016). Darunter fallen beispielsweise Intelligenz sowie Persönlichkeitsmerkmale, aber auch sämtliche Einstellungen, Überzeugungen und Kompetenzen. Allerdings kann anhand beobachtbarer Variablen und Theorien – bei ausreichender Validität – auf die Ausprägung von Konstrukten geschlossen werden (Döring &

Bortz, 2016). Bei der Modellierung eines Konstrukts anhand mehrerer Variablen werden konfirmatorische Faktoranalysen gerechnet. Das klassische Vorgehen dabei ist, manifeste Variablen theoriegeleitet als Faktoren eines latenten Konstrukts zu modellieren (Geiser, 2011).

So lassen zum Beispiel Antworten in einem Motivationsfragebogen Rückschlüsse auf die Motivation der Person zu, welche den Fragebogen beantwortet hat. In ähnlicher Weise kann anhand der korrekt gelösten Aufgaben in einem Mathematiktest auf die mathematische Kompetenz geschlossen werden. Erkenntnispotential entfalten solche Ansätze jedoch nur, wenn sie theoriegeleitet vorgehen und auf dieser Grundlage Annahmen in Hypothesen überführt und getestet werden (Popper, 1932). Die Veränderungen von Konstrukten werden dabei anhand von Veränderungshypothesen geprüft (Döring & Bortz, 2016). Analog dazu testen Unterschiedshypothesen Unterschiede (Döring & Bortz, 2016). In der Forschungspraxis verschwimmt die Differenzierung zwischen Veränderung und Unterschied jedoch – Veränderungen sind stets auch Unterschiede und vice versa. Dies zeigt sich am Beispiel eines idealtypischen (Quasi)Experiments mit zwei Gruppen und zwei Messzeitpunkten (siehe Abbildung 1).



**Abb. 1:** Veränderungen und Unterschiede (in Anlehnung an Reinders, 2006)

Unterschied 1 und 2 beschreiben, ob beide Gruppen sich hinsichtlich des untersuchten Konstrukts unterscheiden. Dabei gilt eine signifikante Mittelwertdifferenz als Indiz für die unterschiedliche Ausprägung des Konstrukts bei beiden Gruppen. Analog dazu geben Veränderung 1 und 2 an, inwiefern sich das jeweilige Konstrukt innerhalb der Gruppe verändert hat. An dieser Stelle wird eine signifikante Mittelwertdifferenz als Veränderung des Konstrukts von t1 zu t2 interpretiert. Solche und ähnliche Designs kommen in zahlreichen Studien der empirischen Bildungsforschung zum Einsatz (vgl. Bortz & Schuster, 2010). Dies verdeutlicht ein Blick auf eine beispielhafte Untersuchung von Rochnia, Trempler und Schellenbach-Zell (2019). Die Autor\*innen untersuchten die Entwicklung der Forschungsorientierung im Lehramt- und Medizinstudium. Dabei wurde betrachtet, wie sich das Konstrukt der Forschungsorientierung innerhalb der Studiengänge im Zeitraum vom ersten zum fünften Semester entwickelt und ob sich Lehramt- und Medizinstudium hinsichtlich ihrer Forschungsorientierung unterscheiden. Beide Hypothesen wurden anhand unterschiedlicher Mittelwerte in der Forschungsorientierung untersucht. Die Interpretation von Mittelwertdifferenzen ist jedoch nur eine Möglichkeit, bestimmte Unterschiede und Veränderungen zu beobachten und basiert auf einer grundlegenden

genden Vorbedingung: Die sogenannte Konstruktäquivalenz (Rippl & Seipel, 2008). Konstruktäquivalenz bedeutet je nach Hypothese, dass ein Konstrukt in verschiedenen Gruppen oder an verschiedenen Messzeitpunkten das gleiche Phänomen beschreiben muss. Für die Unterschieds- und Veränderungsmessungen im obigen Studienbeispiel ist demnach zentral, dass Forschungsorientierung im Lehramt- und Medizinstudium, sowie im ersten und fünften Semester, das Gleiche bedeutet. Das heißt, bereits im ersten Semester müssen die grundlegende Struktur, das grundlegende Verständnis des Konstruktes und die Bedeutsamkeit der einzelnen Indikatoren bei den Studierenden gleich ausgeprägt sein wie in späteren Semestern, um überhaupt eine Analyse der Veränderung durchführen zu können.

### **3 Epistemische Probleme von Veränderungs- und Unterschiedsmessungen**

Das Verständnis von Konstruktäquivalenz als Vorbedingung für Veränderungs- und Unterschiedsmessungen ist also nicht unproblematisch. Die grundsätzliche Problematik liegt darin, dass die Messung von Unterschiedlichkeit und Veränderbarkeit das Gleichbleiben des Konstrukts auf messtheoretischer Ebene voraussetzt (Rippl & Seipel, 2008). Auf theoretischer Ebene können die Indikatoren (manifesten, direkt mess- oder beobachtbaren Variablen) häufig je unterschiedliche Relevanz für ein Konstrukt haben. Eine Verschiebung ihrer Relevanz verändert notwendigerweise die Bedeutung des Konstruktes, kann also in diesem Sinne auch Gegenstand von theoretisch plausibler und hochrelevanter Veränderung oder von interessanten Unterschieden sein. Beispielsweise kann die Thematisierung von aktueller Forschung für das Konstrukt Forschungsorientierung relevanter sein als Methodenwissen. Diese Bedeutsamkeit kann sich aber im Verlauf eines Studiums systematisch verschieben. Methodisch zeigt sich dies an unterschiedlichen Faktorladungen der manifesten Variablen auf das Konstrukt und Alpha-Reliabilitäten (Brown, 2006). Auf der Ebene eines Mittelwertvergleichs für das latente Konstrukt ist diese unterschiedliche Bedeutung für das Konstrukt jedoch nicht relevant und auch nicht ersichtlich. Eine Untersuchung von Veränderung über Mittelwertvergleiche hinterlässt hier also blinde Flecken.

Ein Blick auf das Prinzip der Messinvarianz als Nachweisverfahren verdeutlicht wie Konstruktäquivalenz bestimmt wird. Messinvarianz ist hierarchisch gestuft und bezieht sich auf die Äquivalenz von verschiedenen faktorenanalytischen Parametern zwischen Gruppen oder Messzeitpunkten (siehe Tabelle 1). Für den Vergleich von Mittelwerten eines Konstrukts zwischen Gruppen oder Messzeitpunkten sollte zumindest skalare Messinvarianz vorliegen (Christ & Schlüter, 2012). Die Abgrenzung der verschiedenen Stufen von Messinvarianz, welche anhand von statistischen Parametern erfolgt, ist dabei häufig nicht völlig eindeutig und bedarf der Interpretation (vgl. Chen, 2007; Cheung & Rensvold, 2002).

**Tabelle 1: Merkmale von Messinvarianz**

Konfigurale Messinvarianz	Gleiche Faktorenanzahl und identisches Ladungsmuster des Konstrukts bei den Vergleichsgruppen
Metrische Messinvarianz	Wie bei konfiguraler Messinvarianz + gleiche Faktorladungen
Skalare Messinvarianz	Wie bei metrischer Messinvarianz + identische Intercepts
Strikte Messinvarianz	Wie bei skalarer Messinvarianz + gleiche Residualvarianzen

Gerade die Forderung nach skalarer Messinvarianz ist als problematisch einzustufen. Interessiert man sich wie bei dem obigen Studienbeispiel von Rochnia, Trempler und Schellenbach-Zell (2019) für die Veränderung der Forschungsorientierung im Laufe des Lehramtsstudiums und den Unterschied zum Medizinstudium, verursacht skalare Messinvarianz ein Dilemma: Der Nachweis von Veränderung des Konstrukts von t1 zu t2, sowie des Unterschieds des Konstrukts zwischen beiden Studiengängen erfolgt über die Mittelwerte. Damit diese allerdings überhaupt verglichen und ggf. als Veränderung oder Unterschied des gleichen Konstruktes interpretiert werden können, bedarf es skalarer Messinvarianz. Das Fehlen von skalarer Messinvarianz selbst ist jedoch ein theoretisch hochplausibler Hinweis darauf, dass es Veränderung zwischen zwei Messzeitpunkten gegeben hat oder dass es hochplausible Unterschiede zwischen verschiedenen Gruppen gibt. Chen (2008) merkt dazu an, dass ein unterschiedliches Verständnis eines Konstrukts vorherrscht, wenn Messinvarianz verfehlt wird. Es ist jedoch eine hochplausible und sogar theoretisch angelegte Veränderung, wenn Studierende unter dem Konstrukt Forschungsorientierung im fünften Semester etwas Anderes verstehen als im ersten Semester. Ursache hierfür sind neben dem Erwerb forschungsaffiner Kompetenzen (vgl. Groß, Ophoff & Rott, 2017) auch elaboriertere epistemologische Überzeugungen, die sich im Studienverlauf herausbilden (vgl. Kienhues, Bromme & Stahl, 2008). Analog dazu kann und muss Forschungsorientierung disziplinbedingt in der Medizin etwas Anderes als im Lehramtsstudium bedeuten (vgl. Berliner, 2002) – das Konstrukt muss also plausibler Weise zwischen den beiden Studiengängen variieren. In beiden Fällen kann und muss also bereits fehlende Messinvarianz als Nachweis von Veränderung oder Unterschied interpretiert werden. Ein Vergleich der Mittelwerte wäre strenggenommen gar nicht zulässig. Prinzipiell gibt es zwei problematische und zwei unproblematische Möglichkeiten (siehe Tabelle 2). Wenn Konstruktäquivalenz gegeben ist, können Mittelwertdifferenzen eindeutig als Veränderungen oder Unterschiede interpretiert werden. Sollten keine Mittelwertdifferenzen beobachtet worden sein, sind vice versa weder Veränderungen noch Unterschiede anzunehmen. Problematisch wird die Interpretation der Sachverhalte, wenn die Konstruktäquivalenz nicht gegeben ist. Unterschiede und Veränderungen können dann nicht wie üblich über die Mittelwertdifferenz beobachtet werden, obwohl bereits auf der Ebene des Messmodells theoretisch oftmals plausible Variabilität – also Unterschiede und Veränderungen – existiert und nachgewiesen wurde (siehe Tabelle 2).

**Tabelle 2:** Epistemische Probleme von Veränderungs- und Unterschiedsmessungen

		Konstruktäquivalenz	
		nein	ja
Mittelwert- differenz	nein	Epistemisches Problem 1: Latente(r) Veränderung/Unterschied Mittelwertdifferenz irrelevant	Keine Veränderung Kein Unterschied
	ja	Epistemisches Problem 2: Latente(r) Veränderung/Unterschied Mittelwertdifferenz uninterpretierbar	Veränderung Unterschied

Die Forderung nach Konstruktäquivalenz führt somit dazu, dass eine Vorbedingung von Veränderungs- und Unterschiedsmessungen selbst bereits Veränderung und Unterschied belegt – und zwar paradoxerweise dadurch, dass diese Forderung verletzt wird.

#### 4 ABG-Typologie als mögliche Problemlösung

Eine Möglichkeit mit den epistemischen Problemen aus Tabelle 2 umzugehen, ist den Verlust von Messinvarianz direkt als Hypothesentest für Unterschiede oder Veränderungen zu implementieren. Genau diesen Weg beschreitet die ABG-Typologie von Golembiewski, Billingsley und Yeager (1976). Die Autoren unterscheiden in Alpha-, Beta- und Gamma-Change. Unter Alpha-Change kann eine traditionelle Mittelwertdifferenz in einem *t*-Test verstanden werden (Riordan, Richardson, Schaffer & Vandenberg (2001). Beta-Change bezeichnet den Wandel in der Interpretation von Antwortkategorien, die Bedeutung der Ausprägung von manifesten Variablen variiert dadurch (Riordan et al., 2001). Es ist zum Beispiel möglich, dass Proband\*innen nach einer Intervention höhere Anforderungen an Forschungsorientierung stellen und daher niedrigere Werte vergeben, weil sich ihr Verständnis von Mittelmäßigkeit geändert hat. Methodisch bedeutet dies, dass die Skalierung der Items des Konstrukts sich in der Interpretation durch die Proband\*innen neu organisiert hat (Millsap & Hartog, 1988). Dies bedeutet, dass eine Intervention eine Skala rekali­brieren kann und Prä-Post-Messungen in einem solchen Fall keine Aussage über die Wirkung der Intervention ermöglichen (Terborg et al., 1980). Je nach Verteilung ist auch zu erwarten, dass eine konfirmatorische Faktoranalyse variante Faktorladungen und die Intercepts zeigt. Gamma-Change ist die Stufe, bei der Wandel die Interpretation des gesamten Konstrukts verändert (Bedeian et al., 1980; Millsap & Hartog, 1988). Daraus ergibt sich per se ein neues Ladungsmuster und damit eine variante Faktorstruktur in konfirmatorischen Faktoranalysen – also prinzipiell fehlende Messinvarianz. Beispielsweise kann das Lesen von Ratgeberliteratur nach einer entsprechenden Intervention nicht mehr Teil des Konstrukts Forschungsorientierung sein, da diese Quellen fortan nicht mehr als wissenschaftlich gelten. Terborg et al., 1980 geben einen praktischen Vorschlag für die Erfassung von Alpha-Change: Then-Measure, welcher quasi als *t*<sub>3</sub> direkt auf *t*<sub>2</sub> (Post-Measure) folgt. Dabei wird aber gefragt, wie sie sich jetzt zu *t*<sub>1</sub> (Pre-Measure) einschätzen würden. Die Idee dahinter ist, da Then-Measure sehr nah am Post-Measure liegt, beide Messungen mit einer rekali­brierten Skala erfolgten – sprich frei von Beta-Change sind und daher Alpha-Change beobachtet werden kann. Die Bilanzierung von Hypothesen erfolgt anhand einer „doppelten“ Testung. Wenn keine Unterschiede zwischen Pre- und Then-Measure gefunden werden, fand kein Beta-Change statt und daher kann zwischen Pre- und Post-Measure, aber auch Then-Measure, Alpha-Change beobachtet werden (Terborg et al., 1980). Ein Vorschlag zur Identifikation von

Beta-Change ist bei Bedeian et al. (1980) zu finden. Dabei findet eine numerische Anpassung von Skalenrekalibrierung zwischen aktuellen und idealen Maßen statt (Bedeian et al., 1980). Dies setzt jedoch voraus, dass es vier Datenpunkte gibt: Aktuelle und ideale Maße zu t1 und zu t2 (siehe Bedeian et al., 1980, Figure 1). Für Gamma-Change schlagen Terborg et al. (1980) die Betrachtung von Korrelationen vor: Wenn Gamma-Change stattfand, dann sollten die Korrelationen zwischen Post- und Then-Measures größer sein als die Zusammenhänge zwischen Pre- und Post-Measures sowie Pre- und Then-Measures. Armenakis und Bedeian (1982) kritisieren die Idee des Then-Measures jedoch, da es sich dabei um eine retrospektive Erfassung handelt. Es ist fraglich, ob Proband\*innen sich zuverlässig an Ereignisse in der Vergangenheit erinnern können. Somit ist auch nicht klar, inwiefern auf dieser Grundlage belastbare Analysen durchgeführt werden können.

Klopp und Stark (2019) gehen einen anderen Weg: Alpha-, Beta- und Gamma-Change können über die Analyse der Messinvarianz geprüft werden. Klopp und Stark (2019) ordnen der ABG-Typologie die entsprechende Stufe von Messinvarianz zu (siehe Tabelle 3). Die Prüfung von Messinvarianz wird dabei nicht mehr nur als Prüfung von notwendigen Bedingungen von Mittelwertvergleichen angesehen, sondern als integraler Bestandteil der Analyse von Unterschiedlichkeit oder Veränderung verstanden. Aus dieser Typologie heraus wäre Gamma-Change die radikale Veränderung oder der deutlichste Unterschied. Die Variabilität (als Unterschiedlichkeit oder Veränderung) ist so deutlich, dass sie statistische Unvergleichbarkeit bedeutet. Gängige Veränderungs- und Unterschiedsmessungen, die sich auf Alpha-Change beschränken, wären dafür blind.

**Tabelle 3:** Veränderungsprozesse, Messinvarianz und Parameter (Klopp & Stark, 2019)

Change	Messinvarianz	Parameter
Alpha $\alpha$	Skalar	Mittelwerte
Beta $\beta$	Metrisch/konfigural	Faktorladungen/Intercepts
Gamma $\gamma$	Keine	Faktorenanzahl/Ladungsmuster

## 5 Schlussfolgerung

Die in diesem Beitrag dargestellte Problematik erfährt in der empirischen Bildungsforschung bisher wenig Aufmerksamkeit (vgl. Schwab & Helm, 2015). Ein Grund hierfür ist auch, dass Konstruktäquivalenz selten geprüft wird – das Problem wird somit weitestgehend ignoriert bzw. kann gar nicht thematisiert werden (vgl. z.B. Schulte et al., 2013, S. 101f; Schwab & Helm, 2015, S. 176f). Ein weiteres Problem ist, dass strukturelle Variabilität häufig theoretisch plausibel ist, aber unter einem blinden Fleck liegt, wenn Unterschiede und Veränderungen lediglich über Mittelwerte beobachtet werden. Es ist theoretisch plausibel, dass Lehramtsstudierende im ersten und fünften Semester über alles hinweg ähnliche Mittelwerte für die Forschungsorientierung ihres Studiums angeben, auf Ebene des Messmodells das Konstrukt aber Veränderungen unterlegen ist. Forschungsorientierung kann später im Studium etwas Anderes bedeuten als in der Studieneingangsphase – vielleicht sollte es dies sogar. Diese Differenz würde jedoch bei der Fokussierung auf Mittelwertdifferenzen nicht sichtbar werden können – Veränderungen und Unterscheide könnten so gar nicht näher untersucht und thematisiert werden.

Eine Möglichkeit diesem Problem zu begegnen ist die ABG-Typologie (Golembiewski et al., 1976). Trotz des Alters des Ansatzes ist dieser in der aktuellen empirischen Bildungsforschung – ebenso wie das Problembewusstsein für Veränderungs- und Unterschiedsmessungen – wenig rezipiert. Die elaborierte Diskussion wie bereits in 1980er Jahren, finden aktuell nicht wirklich statt (vgl. Armenakis et al., 1982; Bedeian et al., 1980; Terborg et al., 1980). Klopp und Stark (2019) verwendeten den Ansatz zu Analyse von Veränderungen in den epistemologischen Überzeugungen von Lehramtsstudierenden. Es ist durch die Einbeziehung der Messinvarianzprüfung in die eigentliche Analyse von Unterschieden und Veränderungen durchaus möglich, Veränderungen und Unterschiede im Bildungsbereich multiperspektivisch zu beobachten und gewinnbringende Implikationen zu ziehen. Das eröffnet ganz neue Möglichkeiten in der empirischen Bildungsforschung. Es ist aus der Perspektive des Conceptual Change durchaus zu erwarten, dass sich nicht nur Ausprägungen von stabilen Konstrukten verändern, sondern dass die Konzepte selbst, welche durch die Konstrukte abgebildet werden, einem gewollten und theoretisch begründeten Wandel unterliegen. Diesen Wandel aber blendet man systematisch aus, wenn man versucht Veränderungen und Unterschiede anhand von statistischer Vergleichbarkeit zu theoretisieren. Gerade Unvergleichbarkeit heißt aber oftmals sogar gewollte oder plausible Veränderung und Andersartigkeit. Fortan wäre es zu überdenken, Messinvarianzanalysen in ihrer Bedeutsamkeit stärker wahrzunehmen und die Unvergleichbarkeit – Gamma-Change – im Sinne der ABG-Typologie als empirisch und theoretisch interessanten Unterschied oder Veränderung in den Blick zu nehmen, gerade weil ein Konstrukt unvergleichbar geworden ist.

## Danksagung

Wir danken Eric Klopp für die Bekanntmachung mit der ABG-Typologie. Ohne diesen Impuls wäre die Idee für dieses Papier vermutlich nie entstanden.

## Literaturverzeichnis

- Armenakis, A. A. & Bedeian, A. G. (1982). On the Measurement and Control of Beta Change: A Reply to Terborg, Maxwell, and Howard. *Academy of Management Review*, 7(2), 296-299.
- Backhaus, K., Erichson, B., Plinke, W. & Weiber, R. (2016). *Multivariate Analysemethoden*. Springer Gabler.
- Baur, N. & Blasius, J. (2014) (Hrsg.). *Handbuch Methoden der empirischen Sozialforschung*. VS Verlag.
- Bedeian, A. G., Armenakis, A. A. & Gibson, R. W. (1980) The Measurement and Control of Beta Change. *Academy of Management Review*, 5(4), 561-566.
- Bellmann, J. (2016). Datengetrieben und/oder evidenzbasiert? Wirkungsmechanismen bildungspolitischer Steuerungsansätze. *Zeitschrift für Erziehungswissenschaft*, 19(Supplement 1), 147-161.
- Berliner, D. C. (2002). Educational Research: The Hardest Science of All. *Educational Researcher*, 31(8), 18-20.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. VS Verlag.
- Brown, T. A. (2006). *Confirmatory Factor Analysis*. The Guilford Press.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464-504.
- Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, 95, 1005-1018.



- Cheung, G. W. & Rensvold, R. B. (2002). Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling*, 9, 233-255.
- Christ, O. & Schlüter, E. (2012). *Strukturgleichungsmodelle mit Mplus – Eine praktische Einführung*. Oldenbourg Verlag.
- Deutsches PISA-Konsortium (2001) (Hrsg.). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Leske + Budrich.
- Döring, N. & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. VS Verlag.
- Geiser, C. (2011). *Datenanalyse mit Mplus*. VS Verlag.
- Golembiewski, R. T., Billingsley, K., & Yeager, S. (1976). Measuring change and persistence in human affairs: Types of change generated by OD designs. *Journal of Applied Behavioral Science*, 12, 133-157.
- Gräsel, C. (2011). Was ist Empirische Bildungsforschung? In H. Reinders, H. Ditton, C. Gräsel & B. Gniewosz (Hrsg.), *Empirische Bildungsforschung* (S. 13-28). VS Verlag.
- Groß Ophoff, J. & Rott, B. (2017). Educational research literacy. *Journal for educational research online*, 9(2), 5-10.
- Keiner, E. (2011). Evidenzbasierte Pädagogik ohne historische und vergleichende Kontexte? In J. Bellmann & T. Müller (Hrsg.), *Wissen, was wirkt* (S. 217-234). VS Verlag.
- Kienhues, D., Bromme, R., & Stahl, E. (2008). Changing epistemological beliefs: the unexpected impact of a short-term intervention. *The British journal of educational psychology*, 78(4), 545-565.
- Klieme, E. & Steinert, B. (2008). Schulentwicklung im Längsschnitt. Ein Forschungsprogramm und erste explorative Analysen. In M. Prenzel & J. Baumert (Hrsg.), *Vertiefende Analysen zu PISA 2006* (S. 221-238). VS Verlag.
- Klopp, E. & Stark, R. (2019, Februar). Die Entwicklung epistemischer Überzeugungen –Veränderung der Ausprägung oder epistemischer Konzeptwandel? 7. Tagung der Gesellschaft für empirische Bildungsforschung, Köln.
- Lay, W. A. (1908). *Experimentelle Pädagogik mit besonderer Rücksicht auf die Erziehung durch die Tat*. B.G. Teubner.
- Lenzen, D. (2019). Empirische Bildungsforschung in Deutschland. *Journal for educational research online*, 11(1), 9-15.
- McElvany, N., Gebauer, M. & Gräsel, C. (2019). Empirische Bildungsforschung – eine Standortbestimmung. *Journal for educational research online*, 11(1), 5-8.
- Meumann, E. (1911). *Vorlesungen zur Einführung in die Experimentelle Pädagogik und ihre psychologischen Grundlagen*. Engelmann.
- Millsap, R. E. & Hartog, S. B. (1988). Alpha, beta, and gamma change in evaluation research: A structural equation approach. *Journal of Applied Psychology*, 73, 574-584.
- Popper, K. (1932). Zuschriften an die Herausgeber. Ein Kriterium des empirischen Charakters theoretischer Systeme. *Erkenntnis*, 3, 426-428.
- Radtke, F.-O. (2016). Konditionierte Strukturverbesserung. *Zeitschrift für Pädagogik*, 62(5), 707-731.
- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2014a). *Quantitative Methoden 1*. VS Verlag.
- Rasch, B., Friese, M., Hofmann, W. & Naumann, E. (2014b). *Quantitative Methoden 2*. VS Verlag.
- Reinders, H. (2006). Kausalanalysen in der Längsschnittforschung. Das Crossed-Lagged-Panel Design. *Diskurs Kindheits- und Jugendforschung*, 1(4), 569-587.
- Riordan, C., Richardson, H. A., Schaffer, B. S. & Vandenberg, R. J. (2001). Alpha, beta, and gamma change: A review of past research with recommendations for new directions. In C. A. Schriesheim & L. L. Neider (Hrsg.), *Equivalence in Measurement* (S. 51-97). Information Age Publishing.
- Rippl, S. & Seipel, C. (2008). *Methoden kulturvergleichender Sozialforschung*. VS Verlag.
- Rochnia, M., Trempler, K. & Schellenbach-Zell, J. (2019). Vergleich der Forschungs- sowie Praxisorientierung zwischen Lehramts- und Medizinstudium. *Zeitschrift für empirische Hochschulforschung*, 3(2), 123-138.
- Roth, H. (1962). Die realistische Wendung in der pädagogischen Forschung. In H. Becker, E. Blochmann, O. F. Bollnow, E. Heimpel, M. Wagenschein (Hrsg.), *Neue Sammlung. Göttinger Blätter für Kultur und Erziehung* (S. 481-490). Vandenhoeck & Ruprecht.

- Schulte, K., Nonte, S. & Schwippert, K. (2013). Die Überprüfung von Messinvarianz in international vergleichenden Schulleistungsstudien am Beispiel der Studie PIRLS. *Zeitschrift für Bildungsforschung*, 3(2), 99-118.
- Schwab, S. & Helm, C. (2015). Überprüfung von Messinvarianz mittels CFA und DIF-Analysen. *Empirische Sonderpädagogik*, 7(3), 175-193.
- Terborg, J. R., Howard, G. S. & Maxwell, S. E. (1980). Evaluating Planned Organizational Change: A Method for Assessing Alpha, Beta, and Gamma Change. *Academy of Management Review*, 5(1), 109-121.
- Tippelt, R. & Schmidt-Hertha, B. (2018) (Hrsg.). *Handbuch Bildungsforschung*. VS Verlag.