

Dr. Klaus-Eckart Rogge

**Angaben zum Autor:**

Dr. rer.nat. Klaus-Eckart Rogge war bis 2005 Akademischer Oberrat am Psychologischen Institut der Universität Heidelberg. Er übte Lehr- und Prüfungstätigkeiten in den Bereichen Forschungsmethoden/Statistik und Biopsychologie aus.

Seine Forschungsschwerpunkte sind Multivariate Statistik, Prognosemodelle und Systemtheorie. Er hat zahlreiche Buch- und Zeitschriftenbeiträge zum Thema veröffentlicht und war mit der Methodenberatung, Organisation, Koordination und Durchführung statistischer Analysen in verschiedenen Forschungsprojekten in Deutschland und den USA befasst.

**Stellungnahme zur Methodik der Studie :  
„Evaluationsstudie zur Sprachförderung von Vorschulkindern (EVAS)“**

Karin Schakib-Ekbatan, Petra Hasselbach, Jeanette Roos, Hermann Schöler, Silvana Poltzeck und Nicole Hofmann

**1. Texte**, auf die sich die Stellungnahme bezieht:

- EVAS-Bericht März 2006
- EVAS-Zwischenbericht April 2007
- EVAS-Zwischenbericht Mai 2007
- EVAS-Zwischenbericht 2008
- EVAS-Vorläufiger Abschlussbericht 16.3.2009

**2. Globale Einschätzung**

Die „Evaluationsstudie zur Sprachförderung von Vorschulkindern (EVAS)“ wird unter eine recht allgemeine interrogative Zielsetzung: „.....ob und in welcher Weise die Sprachfördermaßnahmen, die in den Städten Heidelberg und Mannheim im Laufe des letzten Kindergartenjahres vor der Einschulung eingesetzt und im Rahmen des Programms Sag' mal was - Sprachförderung für Vorschulkinder der Landesstiftung Baden-Württemberg gefördert werden, zu einer Verbesserung der sprachlichen Fähigkeiten der förderbedürftigen Kinder führen und ob sich differenzielle Effekte durch den Einsatz unterschiedlicher Programme zeigen“ (Bericht 2006; S. 12) gestellt und durch zwei Konsekutivziele: die Überprüfung der unmittelbaren Wirksamkeit und der Nachhaltigkeit der Sprachfördermaßnahmen ergänzt. In der Studie werden dann dezidiertere Fragestellungen:

„Führen die speziellen Fördermaßnahmen zu einer Verbesserung der sprachlichen Leistungen der Kinder mit einem Sprachförderbedarf?

(b) Führen die Sprachfördermaßnahmen im Projekt „Sag' mal was“ dazu, dass sich die Bildungschancen von Kindern mit Sprachförderbedarf verbessern?

(c) Ist die Durchführung spezifisch strukturierter Programme durch Förderkräfte effektiver ist als die Bildung<sub>1</sub> von Sprache, wie sie im Kindergarten ohne zusätzliche Maßnahmen und Sprachförderkräfte erfolgt?

(d) Gibt es Unterschiede in der Wirksamkeit der Fördermaßnahmen in Abhängigkeit

von den verschiedenen Sprachförderkonzeptionen, die von den Förderkräften im Rahmen des Projektes „Sag‘ mal was“ in den Kindertageseinrichtungen durchgeführt werden?“ (Abschlussbericht 2009, S. 7)

und Zielsetzungen artikuliert, wobei sich aber nachteilig auswirkt, dass überhaupt keine korrespondierenden Aussagen getroffen werden, auf welcher Datenebene die erhobenen Merkmale der Studie geprüft werden, so dass nicht erkennbar wird, in welchem Bezugsrahmen (z.B. Mittelwerte, Differentialquotienten, Gruppendifferenzen, individuelle Werte) die statistischen Analysen der Studie durchgeführt werden. So wird zum Beispiel auch nicht spezifiziert, ob sich Leistungssteigerungen auf absolute oder relative oder zeitbezogene Wertekonfigurationen beziehen. Die aus den Fragestellungen abgeleiteten Erwartungen (wohl im Sinne von Hypothesen) werden sehr allgemein formuliert und bezüglich der avisierten Daten nicht spezifiziert. Von – wahrscheinlich wesentlichen – Einflussgrößen wie Person-Merkmale oder andere Bedingungsfaktoren (familiärer Sprachhintergrund und sozialer Status) wird nur festgestellt, dass sie zu berücksichtigen seien (Abschlussbericht 2009, S.8) – genauere Hypothesen dazu fehlen (scheinen also a priori nicht prüfungsrelevant zu sein). Dieser Mangel schlägt sich dann in der fehlenden Rückbezüglichkeit der Ergebnisinterpretationen nieder. Außerdem werden die für die Begründung der statistischen Verfahren notwendigen Beziehungen zwischen inhaltlichen und – in der Studie nicht explizierten – statistischen Hypothesen nicht erwähnt.

Das von den Verfassern (Verf.) entwickelte Untersuchungsdesign mit seinen Varianten (Einbezug von Kontrollfaktoren und Kovariaten) dokumentiert das Bemühen um möglichst fundierte, umfängliche und eindeutige Aussagen aus einer quasiexperimentellen Untersuchungsanordnung ziehen zu können. Leider sind solche Bestrebungen in der Prüfung und Mitteilung der Voraussetzungen für die statistischen Analysen nicht zu erkennen. Die Restriktionen bei quasiexperimentellen Designs werden zwar beschrieben und hinreichend begründet (ethische und praktische Vorbehalte), in ihren Auswirkungen (es sind weniger Gruppen als im Solomon-4-Gruppen Versuchsplan gefordert; vgl. Bredenkamp 1969) vor allem auf die Treatment-Effekte und den Geltungsbereich (externe Validität) der Ergebnisse aber nicht charakterisiert oder kritisch diskutiert. Das mindert nicht nur die Einschätzung der Relevanz der erzielten Resultate sondern auch die Möglichkeit zur Bestimmung ihrer Verallgemeinerungsfähigkeit.

Die eingesetzten statistischen Prüfverfahren (Signifikanztests) sind weitgehend indiziert (sofern die dafür notwendigen Voraussetzungen als gegeben oder geprüft angenommen werden können). Kritisch anzumerken bleibt aber, dass über die notwendigen Voraussetzungen für die sensitiven, teststarken Verfahren weder Prüfergebnisse noch Aussagen über ihr Zutreffen gemacht werden. Das gilt insbesondere für die parametergebundenen Verfahren (Varianzanalysen, Scheffe´-Test) und die Einbeziehung von Kovariaten (indiziert und nützlich nur bei hinreichender Kovariation mit der abhängigen Variable). Indizierte alternative Auswertungsstrategien werden leider nicht erwähnt oder diskutiert.

Bedenklich sind ferner die Indikation der inhaltlichen Prüfverfahren (evaluative Erhebungsverfahren für die Effizienz und Nachhaltigkeit von Fördermaßnahmen), der verwendeten Erhebungstechniken der Kontrollfaktoren (z.B. Verständlichkeit des Elternfragebogen) und – in Teilen – die Stichprobenszusammensetzungen der Untersuchungsgruppen.

### 3. Die Kernbereiche der Studie im Einzelnen:

#### **Fragestellung(en) und Hypothesen**

Die von den Verf. in den verschiedenen Zwischenberichten und dem Abschlussbericht formulierten Zielsetzungen bzw. Fragestellungen sind bisher viel zu allgemein formuliert, so dass die in den Untersuchungsdesignvarianten artikulierten Einflussgrößen auf keiner oder nicht hinreichender theoretischer Fundierung oder vorangegangener empirischer Prüfung basieren. Der Mangel an dezidiert ausgearbeiteten Hypothesen betrifft einen Kernbereich wissenschaftlicher Studien (mit Inferenzstatistik), weil nur durch Explikation der (auch statistischen) Hypothesen die theoretischen Begründungszusammenhänge und die Prüfrichtung sowie Prüfmodalität der evozierten Daten ersichtlich und einschätzbar werden. Die im Abschlussbericht (2009) allgemein formulierten Erwartungen (S.7) genügen solchen Ansprüchen nicht.

Ferner wird erst über Hypothesen mit spezifizierten Kennwerten (Parameter) eine Kompatibilität von inhaltlicher Vermutung und statistischer Prüfung sicher gestellt. Die dafür notwendigen Schritte sind einer Aufstellung bei Fahrmeir et al. (2003; S. 410 f) zu entnehmen.

Außerhalb jeder Diskussion bleibt leider auch, ob die ausschließlich auf zentrale Tendenzunterschiede (arithmetisches Mittel) zielende Parametrisierung und die damit verbundenen statistischen Analyseverfahren stets und ausschließlich die einzig indizierte oder alternativlose Methode für die Evaluation sei. Wäre die Hypothesenausrichtung stärker auf individuelle Fortschritte bei den Fördermaßnahmen gerichtet gewesen - statistisch also auf (intraindividuelle) Dispersionstendenzen und individuelle (Modell-)Relationen - dann wären mit den gleichen Daten (!) ganz andere Analysemethoden und Aussagen in den Vordergrund zu rücken oder als zusätzliche Informationen nützlich gewesen. An die Seite der a priori vorgenommenen Gruppeneinteilung (i.e. förderungsbedürftig vs. nicht-förderungsbedürftig) gestellt hätte beispielsweise eine clusteranalytische Präklassifikation (Tabachnik & Fidell 2006) der Kinder mehrere Vorteile erbracht: Zum einen wäre damit ein besserer Homogenisierungsgrad innerhalb der Cluster erreicht worden, d.h. eine viel bessere Vergleichbarkeit der individuellen Ausgangsbedingungen, zum anderen wären individuelle Fortschritte sehr viel exakter auszuweisen gewesen, ferner hätten diskriminante Merkmale gewichtet werden können und nicht zuletzt wären die statistischen Analysen bedeutend günstiger - weil präziser erfasst und mit kleinerem Errorterm belegt - im Hinblick auf die durch die unabhängigen Variablen (Bedingungskonstellationen) hervorgerufenen Effekte zu ermitteln gewesen. Eine detaillierte und Aspekte wechselnde Hypothesenbildung im Sinne einer multivariaten Präklassifikation hätte für die Studie erhebliche Vorteile gehabt; insbesondere bezüglich der Möglichkeit, nicht nur die Präzision der Zentraltendenzparameter zu erhöhen sondern auch über gewichtete (!) individuelle Fortschritte – möglicherweise sogar als Trends auf Zeitverläufe bezogen - Auskunft geben zu können. Abgesehen davon wären die Schwierigkeiten einer Reduktion des Stichprobenumfangs bei Parallelisierung (letztlich auf eine Gruppengröße von N=11, Abschlussbericht 2009, S. 21) zumindest zu mildern gewesen, da bei der clusteranalytischen Präklassifikation umfangstärkere Stichproben (Cluster) zu erwarten gewesen wären. Die ermittelten Probandengruppen hätten dann bezüglich ihrer Frequenzbesetzung bezüglich derjenigen Merkmalen genauer analysiert

werden können (durch Enumeration), die auch zur Parallelisierung herangezogen worden sind. So bleibt die Studie zwar auf einer vertretbaren, aber nicht gerade suffizienten Forschungsstrategie und Kombination statistischer Analyseprozeduren stehen.

### **Untersuchungsdesign**

Von der skizzierten Möglichkeit einer clusteranalytischen Präklassifikation und damit einhergehender Reduktion der Errorterms und der Homogenisierung der Ausgangsbedingungen pro Gruppe (ohne die Notwendigkeit zur Parallelisierung; vgl. Zwischenbericht 2008 und Abschlussbericht 2009, S.19) abgesehen, entspricht das Untersuchungsdesign und die etablierten Extensionen (Klassifikationsfaktoren und Kovariate) den üblichen wissenschaftlichen Standards.

Weit weniger positiv einzuschätzen sind aber bestehende Argumentations- und Diskussionsmängel. So wird die Definition der Studie als quasiexperimentelle Untersuchung zwar einsichtig beschrieben und nachvollziehbar begründet (ethische und praktische Gründe verhindern eine Durchführung experimenteller - oder Gruppendesigns höherer Ordnung; Solomon-Designs), die daraus folgenden Konsequenzen für die Gütekriterien der Untersuchung (i.e.: interne Validität, Geltungsbereich, Verallgemeinerbarkeit, Möglichkeit zur Replikation, statistische Validität) werden jedoch nicht oder nur andeutungsweise angegeben bzw. erörtert. Die Verf. belassen es bei unbefriedigenden, allgemeinen, jedoch nicht lösenden (z.B. die Konsequenzen der fehlenden Zufallszuordnung) und mehrdeutigen Aussagen bewenden wie zum Beispiel: „Das so entstehende quasiexperimentelle Design ermöglicht einen validen Vergleich sowohl der Untersuchungs- als auch der spezifischen Sprachfördergruppen“ (Zwischenbericht 2007, S. 13). In welcher Hinsicht der Vergleich jedoch „valide“ sein soll (interne Validität, externe Validität, statistische Validität, Itemrepräsentanz etc.) wird nicht expliziert und bleibt leider mehrdeutig.

Die Verf. versäumen es fast gänzlich auf die Auswirkungen und Folgen der Probleme bzw. Schwierigkeiten, die bei quasiexperimentellen Untersuchungsdesigns längst bekannt sind (vgl. z.B. Bredenkamp 1969), im Kontext der erzeugten Daten und vor allem den Ergebnissen ihrer Studie hinzuweisen bzw. sie dezidiert zu diskutieren. So hat beispielsweise der vorgenommene Versuch durch Einführung von Kontrollfaktoren und Kovariaten die an sich schwache interne Validität quasiexperimenteller Untersuchungen zu erhöhen durchaus seine Berechtigung, jedoch ist in Folge der wechselseitigen Abhängigkeit der kardinalen Gütekriterien (interne Validität, externe Validität und statistische Validität) eine interrelative Mehrungs-Minderungsproblematik z.B. bei der Betrachtung der internen und der externen Validität zu konstatieren, weil eine Erhöhung der internen Validität oftmals eine Schwächung der externen Validität nach sich zieht. In diesem Sinne argumentieren Bortz & Döring (1995, S. 491) beispielsweise im Hinblick auf die in der EVAS-Studie verwendete (Teil-)Parallelisierung (Zwischenbericht 2008, Abschlussbericht 2009): „Die Parallelisierung geht häufig zu Lasten der externen Validität“. Die externe Validität ist aber bei der vorliegenden Studie von eminenter Wichtigkeit, da der durch sie abgesteckte Geltungsbereich die Verallgemeinerbarkeit der Untersuchungsergebnisse und damit deren praktische Bedeutung (mit)definiert.

Positiv hervorzuheben ist allerdings das Bestreben der Verf., die möglichen Kontrollfaktoren gegebenenfalls auch auf Interaktionswirkungen zu testen, denn oftmals fehlt bei vergleichbaren Studien die Prüfung möglicher Wechselwirkungen und es bleibt bei singulären Testungen der Faktoren, die in der Realität jedoch häufig stark interagieren. Geklärt werden müssten jedoch nach signifikanten Interaktionen die Möglichkeiten der Testung und die Interpretation der Haupteffekte (vgl. Werner 1997).

Unklar bleibt bei der Untersuchungsanlage, ob mögliche – durch die Gruppendifferenzierungen jedoch nicht kontrollierte – Interaktionen von Probanden, Untersuchungsleitern und sprachrelevante Einflussfaktoren außerhalb der Studiensituationen (i.d.R. Kommunikation im sonstigen Alltag) systematische Wirkungen auf die Resultate ausüben konnten. Eine derartige Vermutung wird typischerweise quasiexperimentellen Untersuchungen entgegengesetzt gebracht.

Vor allem für die praktische Nutzung dürfte die im Zwischenbericht (Mai 2007) und im Abschlussbericht (2009, S.22) dokumentierte „Bewertung der Sprachfördermaßnahmen aus der Sicht der Förderkräfte“ von erheblicher Bedeutung sein. Derartige Erfahrungsberichte und numerische Einschätzungen stützen die aus der Studie zu ziehenden wissenschaftlichen Erkenntnisse und vor allem ihre praktischen Umsetzungen erheblich.

### **Stichproben**

Bei der EVAS-Studie handelt es sich in der Terminologie sozialwissenschaftlicher Methodik um eine *anfallende* Stichprobe, d.h. es werden nahezu alle Probanden in die Untersuchung aufgenommen, die verfügbar sind (vgl. Anzahl von Rekrutierungseinheiten in Mannheim und Heidelberg). Zur differenzierteren Beschreibung bzw. zur Kontrolle von Störfaktoren werden dann Quoten bzw. Strata gebildet, die spezifischere Aussagen ermöglichen sollen. Anfallende Stichproben haben in aller Regel einen geringen Geltungsbereich (externe Validität) und weisen damit mitunter einen erheblichen Mangel an Generalisierbarkeit auf. Bereits im Abschnitt „Untersuchungsdesign“ wurde auf Minderungs-faktoren der externen Validität der EVAS-Studie hingewiesen, so dass insgesamt von einer nur sehr eingeschränkten Verallgemeinerungsfähigkeit der Studienergebnisse gesprochen werden kann. Diesbezügliche klare Stellungnahmen der Verf. fehlen jedoch.

Wenn im Abschnitt „Zusammenfassende Wertung und Diskussion“ (Abschlussbericht 2009, S. 65) sachadäquat und dezidiert Faktoren, die das Ausbleiben eines substantiellen Fördereffektes ausmachen könnten (z.B. Kompetenz der Sprachförderkräfte, Entwicklungsalter, Förderbedarf und kultureller Hintergrund der Kinder, Gruppengröße und Förderzeit etc.) zusammengestellt und diskutiert werden, dann verwundert doch das Fehlen der Betrachtung, was die Vielzahl solcher Einflussgrößen denn bei den Daten der vorliegenden Studie angerichtet haben könnte. Wertet man die Kombination der genannten 7 Faktoren als vernetzte zufällige Wirkungskomponente, so resultiert ein hoher Errorterm, der die Geltung der Null-Hypothese (Insignifikanz) favorisieren würde, d.h. eine in dieser Hinsicht ausgelegte Interpretation von Teilresultaten der Studie wäre nahe gelegt.

Wenigstens in der übergeordneten Wertung und Diskussion der Studie oder in der Zusammenfassung (Abschlussbericht 2009, S.1) hätten Hinweise auf die Geltungsbeschränkungen der Resultate der berichteten Studie gegeben werden müssen, damit der Stellenwert, die Valenz und der Geltungsanspruch der Untersuchung klar definiert und abgegrenzt worden wäre. Durch die Allgemeinheit

der Darstellungen im Abschnitt „Zusammenfassende Wertung und Diskussion“ (Abschlussbericht 2009, ab S.65) und das Fehlen eines expliziten Rekurses auf die vorgelegte Studie könnte der Eindruck einer generellen Gültigkeit der erzielten Ergebnisse der EVAS-Studie entstehen – angesichts der in Teilen geringen Fallzahlen der gebildeten Subgruppen [z.B. fast 50% Reduktion der Fallzahlen bei Gruppen mit spezifischer und unspezifischer Förderung am Ende der 2.Klasse (vgl. Abschlussbericht 2009, Abb.5)] und den ja von den Verfassern auch selbst explizierten vielfältigen, unkontrolliert gebliebenen Einflussfaktoren auf die Sprachfördereffekte eine sehr fragwürdige Einschätzung. Die Begrenztheit der Aussagekraft der Untersuchung hätte expliziert werden müssen, um der Gefahr vorschnell gezogener generalisierter Konsequenzen zu begegnen.

Zumindest als Ungereimtheit sind weitere bedenkliche Punkte bei der Stichprobenszusammensetzung einzuschätzen und anzuführen. So muss wohl davon ausgegangen werden, dass es möglicherweise gelegentlich zu Konfundierungen von Fördermaßnahmen gekommen ist. Problematisch ist ferner die Unklarheit, ob bezüglich der Anteile mehrsprachiger vs. einsprachige Kinder immer eine methodologisch unbedenkliche Vergleichsbasis gegeben war.

Ebenfalls in Frage zu stellen ist die soziologische Klassifikation, da ausländische Konditionen (z.B. Stellenwert der Schulabschlüsse) den deutschen nicht immer linear entsprechen, so dass Fehlzuordnungen z.B. in der Bildungsschicht auftreten (z.B. ist eine 10 jährige Schulzeit im Ausland nicht unbedingt gleichzusetzen mit einem deutschen Realschulabschluss).

Ungünstig auf die Parameterschätzungen (via Strichprobenfehler) und damit auch auf die Hypothesenprüfungen wirken sich auch die teils erheblich divergierenden Stichprobenumfänge der gebildeten Subgruppen aus (das gilt z.T. beispielsweise auch für die Eingangsstichprobenumfänge; vgl. Abschlussbericht 2009, Tab.1 und Tab.2, S. 16/17).

Insgesamt beeinflussen die Spezifika der Stichprobenkonstellationen der Studie die erzielten Resultate nicht unerheblich und hätten von den Verf. deutlich kritisch bezüglich ihrer Geltungsrestriktionen diskutiert werden müssen.

### **Statistik**

Formulierungen wie: „Wechselwirkungen zwischen den beiden Einflussgrößen bestehen nicht, d. h. beide Faktoren wirken additiv“ (Bericht 2006, S. 23) sind als Ergebnisinterpretation von statistischen Tests fragwürdige Aussagen, da damit gemeint sein könnte, dass die  $H_0$  (Null-Hypothese) als richtig angesehen würde (Geltung der Additivitätsregel). Schon in der Einführungsliteratur (Rogge 1995) wird aber darauf hingewiesen: „Die Beibehaltung einer  $H_0$  als Folge eines *nicht*signifikanten Prüfergebnisses kann *nicht* als Nachweis dafür gewertet werden, dass die  $H_0$  richtig ist.“

Solche Fehleinschätzungen nähren Zweifel an der Sorgfalt in der Methodik der Studie. Dieser Eindruck verstärkt sich durch das Fehlen von Hinweisen und/oder Darlegungen von Prüfergebnissen zu den Voraussetzungen bei Untersuchungsdesigns mit Messwiederholung. Die Erfüllung bzw. Nicht-Erfüllung der Voraussetzungen hat bekanntlich einen wesentlichen Einfluss auf die Testung der statistischen Hypothesen: „ In Varianzanalysen mit Messwiederholungen müssen die Varianzen unter den einzelnen Faktorstufen und die Korrelationen zwischen den Faktorstufen homogen sein. Eine Verletzung dieser Voraussetzung führt zu progressiven

Entscheidungen“ (Bortz 2005; S.354). Bei Prä- / Post-(2-Zeitpunkt-)Vergleichen ist die Homogenitätsbedingung der Korrelationen logischerweise nicht zu stellen.

Da die Einführung von Kovariaten nur dann sinnförderlich ist, wenn mit ihnen eine hohe Korrelation mit der abhängigen Variable besteht, ist die Prüfung dieser Voraussetzung erforderlich – wird jedoch in den Berichten nicht mitgeteilt. Unerwähnt bleibt ferner die vorteilhafte Möglichkeit, die nominalklassifizierten Variablen in Dummy-Variable zu kodieren, um dann über Algorithmen der Allgemeinen Linearen Modelle (ALM, Werner 1997) rechnen und argumentieren zu können.

Vorzuschlagen wäre ferner der Einbezug von Trendtests, um die Entwicklung vom Prä- über die drei Posttests hinweg auf mathematische Funktionsanpassung und Niveauverlauf in der Zeit prüfen und interpretieren zu können.

### **Tests** zur Erfassung der Sprachkompetenz

Die verwendeten Verfahren zur Erfassung der mündlichen Sprachkompetenz weisen teilweise Mängel auf, weil sie entweder veraltet oder invalide bezüglich Inhalt und/oder Konstruktionsprinzip sind (z.B. HSET; es werden für den Schulerfolg grundlegende Fähigkeiten wie Textverstehen und Textproduktion nicht erhoben).

Ganz ähnliche Einschränkungen sind für die in der Studie verwendete, aber nicht mehr aktuelle Form der Förderkonzeption „Deutsch für den Schulstart“ (Kaltenbacher & Klages) zu konstatieren. Zumindest unklar bleibt auch die Wertigkeit des Elternfragebogens, da darin auftauchende widersprüchliche Ergebnisse den Verdacht nahe legen, der Fragebogeninhalt könnte nicht durchgängig von den Befragten verstanden worden sein.

### Literatur:

Bortz J ( 2005; 6.Aufl.) Statistik für Human- und Sozialwissenschaftler. Berlin: Springer

Bortz J & Döring N (1995): Forschungsmethoden und Evaluation für Sozialwissenschaftler. Berlin: Springer

Bredenkamp J (1969): Experiment und Feldexperiment. In: Graumann C F (Hrsg.): Handbuch der Psychologie. Band 7/1: Sozialpsychologie. Göttingen: Hogrefe

Rogge K E (1995): Methodenatlas. Für Sozialwissenschaftler. Berlin: Springer

Fahrmeir L, Künstler R, Pigeot I & Tutz G (2003; 4.Aufl.): Statistik. Der Weg zur Datenanalyse. Berlin: Springer

Tabachnik B & Fidell L S (2006; 5.Ed.): Using multivariate statistics. Boston: Allyn & Bacon

Werner J (1997): Lineare Statistik. Allgemeines lineares Modell. Weinheim: PVU (Beltz)

Dr. Klaus-Eckart Rogge