

Differentielle Effekte von Feedback in Intelligenztests

In a u g u r a l - D i s s e r t a t i o n

zur

Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Nadin Beckmann

aus Merseburg, Sachsen-Anhalt

2004

Gedruckt mit der Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

Referent(in): Prof. Dr. Axel Buchner

Korreferent(in): Prof. Dr. Martin Heil

Tag(e) der mündlichen Prüfung: 31. Januar 2005

Vorwort und Danksagung

Die vorliegende Arbeit hat zum Ziel, einen Beitrag zur Aufklärung des Zusammenhangs von Persönlichkeit und Leistung im Kontext psychologischer Testung zu leisten. Damit ordnet sich die Arbeit in das Spannungsfeld von differentieller Persönlichkeitsforschung und Leistungs-, speziell Intelligenzdiagnostik ein. Mein Interesse an differentialpsychologischen Fragestellungen wurde während meiner Studienzeit an der Universität Leipzig vor allem durch Professor JÜRGEN GUTHKE angeregt. Unter seiner Leitung habe ich die vorliegende Arbeit an den Leipziger Psychologischen Instituten begonnen. Seine Vorlesungen aber auch die wissenschaftlichen Diskussionen in den Arbeitskreisen haben mein Verständnis von Differentieller Psychologie und Psychologischer Diagnostik wesentlich geprägt und meine Faszination für die psychologische Forschung geweckt. Ich möchte ganz besonders ihm für seine Unterstützung und Ermutigung danken, ein Promotionsthema in diesem Bereich aufzunehmen. Trotz seiner schweren Erkrankung blieb er stets an der Fortführung dieser Forschungsarbeit interessiert. Sein viel zu früher Tod im April 2004 bleibt unfassbar.

Mein besonderer Dank gilt darüber hinaus Professor AXEL BUCHNER (Institut für Experimentelle Psychologie der Heinrich-Heine-Universität Düsseldorf) für sein Interesse an diesem differentialpsychologischen Forschungsprojekt, was dankenswerterweise in der Übernahme der Betreuung meines Promotionsvorhabens mündete. Seine vielfältigen Anregungen und hilfreichen Feedbacks haben wichtige Impulse für die Strukturierung der Arbeit gegeben. Seine Offenheit und wissenschaftliche Neugier haben es dabei gleichzeitig ermöglicht, das Betreuungsverhältnis auch über die weite Distanz zwischen New Haven und Düsseldorf aufrechtzuerhalten. Außerdem danke ich Professor JULIAN G. ELLIOTT (University of Durham, UK) für die Unterstützung bei der Organisation der Datenerhebung und für die Einführung in die „britische Kultur“. Vor allem durch ihn habe ich gelernt, stets die praktische Relevanz psychologischer Forschung im Auge zu behalten. Für die freundschaftliche Unterstützung und fachliche Beratung möchte ich mich außerdem bei Dr. rer. nat. BARBARA B. SEIWALD bedanken.

Die finanzielle Förderung meines Promotionsstudiums wurde durch die STUDIENSTIFTUNG DES DEUTSCHEN VOLKES geleistet. Als besonders wertvoll und anre-

gend habe ich den interdisziplinären Diskurs auf Tagungen und Doktoranden-Foren der Studienstiftung erlebt. Allen Organisatoren und Teilnehmern dieser exzellenten Veranstaltungen sei herzlichst gedankt.

Mein innigster Dank gilt jedoch meinem lieben Mann für seine Liebe und für die stete Unterstützung meines Promotionsvorhabens. Die vergangenen, sehr arbeitsreichen Jahre gemeinsam mit Dir waren vor allem sehr glückliche Jahre.

Nadin Beckmann

New Haven, CT im November 2004

Inhaltsverzeichnis

1	EINLEITUNG UND ZIELSTELLUNG	1
2	FEEDBACK-PHÄNOMENE IM LEISTUNGSKONTEXT.....	4
2.1	ZUM BEGRIFF FEEDBACK.....	4
2.1.1	Das operante Modell – Feedback als Verstärker.....	4
2.1.2	Das kognitive Modell – Feedback als Information.....	7
2.1.2.1	Das konstruktivistische Modell – Feedback als Informationsangebot.....	10
2.1.2.2	Das sozialpsychologische Modell – Feedback als Kommunikation	12
2.1.3	Schlussfolgerungen.....	13
2.2	LEISTUNGSFEEDBACK – EINE SINNVOLLE INTERVENTION?	15
2.2.1	Gegenwärtige Befundlage zur Effektivität von Leistungsfeedback.....	15
2.2.2	Die Feedback-Interventions-Theorie	18
2.2.2.1	Grundannahmen	20
2.2.2.2	Verhaltensregulation	20
2.2.2.3	Feedback und Affekt.....	21
2.2.2.4	Feedback und Leistung	22
2.2.3	Empirische Evidenz für die Effektivität verschiedener Feedback-Strategien.....	24
	Feedback in Intelligenztests	27
2.2.4	Implikationen für eine Leistungsdiagnostik mit Feedback	28
2.2.5	Motivationale Orientierung.....	31
	Lern- versus Leistungszielorientierung	33
2.2.6	Testangst	34
2.2.7	Schlussfolgerungen.....	36
2.3	ZUSAMMENFASSUNG ZU KAPITEL 2	37
3	DYNAMISCHES TESTEN ALS BESONDERE VARIANTE EINER LEISTUNGSMESSUNG MIT FEEDBACK	40
3.1	ZUM TESTKONZEPT	40
3.1.1	Das Lerntest-Konzept.....	43
3.1.2	Feedback in dynamischen Tests.....	44
	Performanzbezogene Feedback-Effekte beim Dynamischen Testen	46
3.1.3	Schlussfolgerungen.....	47
3.2	DIFFERENTIELLE EFFEKTE BEIM DYNAMISCHEN TESTEN	48
3.2.1	Das Kompensationsmodell.....	48
3.2.1.1	Kognitive Impulsivität.....	48
3.2.1.2	Problemlöseverhalten	49
3.2.1.3	Neurotizismus	50
3.2.1.4	Leistungsmotivation	50

3.2.1.5	Stresssensibilität	50
3.2.1.6	Leistungsangst.....	51
3.2.1.7	Schlussfolgerungen	53
3.2.2	Das Modifikationsmodell	56
	Schlussfolgerungen	58
3.3	ZUSAMMENFASSUNG ZU KAPITEL 3	59
4	PROBLEMSTELLUNG	62
	FRAGESTELLUNG	67
5	METHODIK	70
5.1	VERSUCHSPLAN.....	70
5.2	STICHPROBE.....	72
5.3	PROZEDUR.....	76
	Zusätzliche Daten	77
5.4	MESSINSTRUMENTE.....	78
5.4.1	Leistungsdiagnostik.....	78
5.4.1.1	Figurenfolgen.....	78
5.4.1.2	Zahlenfolgen.....	80
5.4.2	Persönlichkeitsdiagnostik	82
5.4.2.1	Zielorientierung.....	84
5.4.2.2	Attribution von Erfolg/Misserfolg	94
5.4.2.3	Begabungskonzept.....	97
5.4.2.4	Testangst.....	99
5.4.2.5	Schulische Motivation	103
5.4.3	Zusammenfassung zu den psychometrischen Eigenschaften der eingesetzten Verfahren	104
6	ERGEBNISSE	106
6.1	GENERELLE FEEDBACK-EFFEKTE	106
6.1.1	Numerische Reasoning-Testleistung.....	106
6.1.2	Figurale Reasoning-Testleistung.....	108
6.2	DIFFERENTIELLE FEEDBACK-EFFEKTE.....	111
6.2.1	Feedback-Effekte und intellektuelle Leistungsfähigkeit.....	113
6.2.1.1	Numerische Reasoning-Testleistung	113
6.2.1.2	Figurale Reasoning-Testleistung.....	115
6.2.2	Feedback-Effekte und außerintellektuelle Personmerkmale	119
6.2.2.1	Zielorientierung.....	120
6.2.2.2	Anstrengungsvermeidung.....	134
6.2.2.3	Attribution von Erfolg/Misserfolg	143
6.2.2.4	Testangst.....	144

6.2.3	Zum Informationsgehalt von Richtig/Falsch-Feedback.....	165
7	ZUSAMMENFASSUNG UND DISKUSSION DER BEFUNDE.....	168
7.1	GENERELLE FEEDBACK-HYPOTHESE	168
7.2	DIFFERENTIELLE FEEDBACK-HYPOTHESE.....	170
7.2.1	Intellektuelle Leistungsfähigkeit.....	170
7.2.2	Außerintellektuelle Personmerkmale.....	171
7.2.2.1	Zielorientierung.....	172
7.2.2.2	Testangst.....	179
7.3	ZUM INFORMATIONSGEHALT VON FEEDBACK	181
7.4	IMPLIKATIONEN FÜR DAS DYNAMISCHE TESTEN	186
7.5	AUSBlick	189
8	ZUSAMMENFASSUNG.....	190
9	LITERATUR.....	191
10	ANHANG	I

Vorwort und Danksagung

Die vorliegende Arbeit hat zum Ziel, einen Beitrag zur Aufklärung des Zusammenhangs von Persönlichkeit und Leistung im Kontext psychologischer Testung zu leisten. Damit ordnet sich die Arbeit in das Spannungsfeld von differentieller Persönlichkeitsforschung und Leistungs-, speziell Intelligenzdiagnostik ein. Mein Interesse an differentialpsychologischen Fragestellungen wurde während meiner Studienzeit an der Universität Leipzig vor allem durch Professor JÜRGEN GUTHKE angeregt. Unter seiner Leitung habe ich die vorliegende Arbeit an den Leipziger Psychologischen Instituten begonnen. Seine Vorlesungen aber auch die wissenschaftlichen Diskussionen in den Arbeitskreisen haben mein Verständnis von Differentieller Psychologie und Psychologischer Diagnostik wesentlich geprägt und meine Faszination für die psychologische Forschung geweckt. Ich möchte ganz besonders ihm für seine Unterstützung und Ermutigung danken, ein Promotionsthema in diesem Bereich aufzunehmen. Trotz seiner schweren Erkrankung blieb er stets an der Fortführung dieser Forschungsarbeit interessiert. Sein viel zu früher Tod im April 2004 bleibt unfassbar.

Mein besonderer Dank gilt darüber hinaus Professor AXEL BUCHNER (Institut für Experimentelle Psychologie der Heinrich-Heine-Universität Düsseldorf) für sein Interesse an diesem differentialpsychologischen Forschungsprojekt, was dankenswerterweise in der Übernahme der Betreuung meines Promotionsvorhabens mündete. Seine vielfältigen Anregungen und hilfreichen Feedbacks haben wichtige Impulse für die Strukturierung der Arbeit gegeben. Seine Offenheit und wissenschaftliche Neugier haben es dabei gleichzeitig ermöglicht, das Betreuungsverhältnis auch über die weite Distanz zwischen New Haven und Düsseldorf aufrechtzuerhalten. Außerdem danke ich Professor JULIAN G. ELLIOTT (University of Durham, UK) für die Unterstützung bei der Organisation der Datenerhebung und für die Einführung in die „britische Kultur“. Vor allem durch ihn habe ich gelernt, stets die praktische Relevanz psychologischer Forschung im Auge zu behalten. Für die freundschaftliche Unterstützung und fachliche Beratung möchte ich mich außerdem bei Dr. rer. nat. BARBARA B. SEIWALD bedanken.

Die finanzielle Förderung meines Promotionsstudiums wurde durch die STUDIENSTIFTUNG DES DEUTSCHEN VOLKES geleistet. Als besonders wertvoll und anre-

gend habe ich den interdisziplinären Diskurs auf Tagungen und Doktoranden-Foren der Studienstiftung erlebt. Allen Organisatoren und Teilnehmern dieser exzellenten Veranstaltungen sei herzlichst gedankt.

Mein innigster Dank gilt jedoch meinem lieben Mann für seine Liebe und für die stete Unterstützung meines Promotionsvorhabens. Die vergangenen, sehr arbeitsreichen Jahre gemeinsam mit Dir waren vor allem sehr glückliche Jahre.

Nadin Beckmann

New Haven, CT im November 2004

Inhaltsverzeichnis

1	EINLEITUNG UND ZIELSTELLUNG	1
2	FEEDBACK-PHÄNOMENE IM LEISTUNGSKONTEXT.....	4
2.1	ZUM BEGRIFF FEEDBACK.....	4
2.1.1	Das operante Modell – Feedback als Verstärker.....	4
2.1.2	Das kognitive Modell – Feedback als Information.....	7
2.1.2.1	Das konstruktivistische Modell – Feedback als Informationsangebot.....	10
2.1.2.2	Das sozialpsychologische Modell – Feedback als Kommunikation	12
2.1.3	Schlussfolgerungen.....	13
2.2	LEISTUNGSFEEDBACK – EINE SINNVOLLE INTERVENTION?	15
2.2.1	Gegenwärtige Befundlage zur Effektivität von Leistungsfeedback.....	15
2.2.2	Die Feedback-Interventions-Theorie	18
2.2.2.1	Grundannahmen	20
2.2.2.2	Verhaltensregulation	20
2.2.2.3	Feedback und Affekt.....	21
2.2.2.4	Feedback und Leistung	22
2.2.3	Empirische Evidenz für die Effektivität verschiedener Feedback-Strategien.....	24
	Feedback in Intelligenztests	27
2.2.4	Implikationen für eine Leistungsdiagnostik mit Feedback	28
2.2.5	Motivationale Orientierung.....	31
	Lern- versus Leistungszielorientierung	33
2.2.6	Testangst	34
2.2.7	Schlussfolgerungen.....	36
2.3	ZUSAMMENFASSUNG ZU KAPITEL 2	37
3	DYNAMISCHES TESTEN ALS BESONDERE VARIANTE EINER LEISTUNGSMESSUNG MIT FEEDBACK	40
3.1	ZUM TESTKONZEPT	40
3.1.1	Das Lerntest-Konzept.....	43
3.1.2	Feedback in dynamischen Tests.....	44
	Performanzbezogene Feedback-Effekte beim Dynamischen Testen	46
3.1.3	Schlussfolgerungen.....	47
3.2	DIFFERENTIELLE EFFEKTE BEIM DYNAMISCHEN TESTEN	48
3.2.1	Das Kompensationsmodell.....	48
3.2.1.1	Kognitive Impulsivität.....	48
3.2.1.2	Problemlöseverhalten	49
3.2.1.3	Neurotizismus	50
3.2.1.4	Leistungsmotivation	50

3.2.1.5	Stresssensibilität	50
3.2.1.6	Leistungsangst.....	51
3.2.1.7	Schlussfolgerungen	53
3.2.2	Das Modifikationsmodell	56
	Schlussfolgerungen	58
3.3	ZUSAMMENFASSUNG ZU KAPITEL 3	59
4	PROBLEMSTELLUNG	62
	FRAGESTELLUNG	67
5	METHODIK	70
5.1	VERSUCHSPLAN.....	70
5.2	STICHPROBE.....	72
5.3	PROZEDUR.....	76
	Zusätzliche Daten	77
5.4	MESSINSTRUMENTE.....	78
5.4.1	Leistungsdiagnostik.....	78
5.4.1.1	Figurenfolgen.....	78
5.4.1.2	Zahlenfolgen.....	80
5.4.2	Persönlichkeitsdiagnostik	82
5.4.2.1	Zielorientierung.....	84
5.4.2.2	Attribution von Erfolg/Misserfolg	94
5.4.2.3	Begabungskonzept.....	97
5.4.2.4	Testangst.....	99
5.4.2.5	Schulische Motivation	103
5.4.3	Zusammenfassung zu den psychometrischen Eigenschaften der eingesetzten Verfahren	104
6	ERGEBNISSE	106
6.1	GENERELLE FEEDBACK-EFFEKTE	106
6.1.1	Numerische Reasoning-Testleistung.....	106
6.1.2	Figurale Reasoning-Testleistung.....	108
6.2	DIFFERENTIELLE FEEDBACK-EFFEKTE.....	111
6.2.1	Feedback-Effekte und intellektuelle Leistungsfähigkeit.....	113
6.2.1.1	Numerische Reasoning-Testleistung	113
6.2.1.2	Figurale Reasoning-Testleistung.....	115
6.2.2	Feedback-Effekte und außerintellektuelle Personmerkmale	119
6.2.2.1	Zielorientierung.....	120
6.2.2.2	Anstrengungsvermeidung.....	134
6.2.2.3	Attribution von Erfolg/Misserfolg	143
6.2.2.4	Testangst.....	144

6.2.3	Zum Informationsgehalt von Richtig/Falsch-Feedback.....	165
7	ZUSAMMENFASSUNG UND DISKUSSION DER BEFUNDE.....	168
7.1	GENERELLE FEEDBACK-HYPOTHESE	168
7.2	DIFFERENTIELLE FEEDBACK-HYPOTHESE.....	170
7.2.1	Intellektuelle Leistungsfähigkeit.....	170
7.2.2	Außerintellektuelle Personmerkmale.....	171
7.2.2.1	Zielorientierung.....	172
7.2.2.2	Testangst.....	179
7.3	ZUM INFORMATIONSGEHALT VON FEEDBACK	181
7.4	IMPLIKATIONEN FÜR DAS DYNAMISCHE TESTEN	186
7.5	AUSBlick	189
8	ZUSAMMENFASSUNG.....	190
9	LITERATUR.....	191
10	ANHANG	I

1 Einleitung und Zielstellung

Feedback wird in Lerntheorien und darauf aufbauenden instruktionalen Modellen als zentraler Wirkmechanismus verstanden, der die Adaptivität des Lernprozesses gewährleistet. Der Erwerb von Wissen, Fertigkeiten und Fähigkeiten ohne Rückkopplungsschleifen im Lernprozess ist nicht vorstellbar. Feedback als Intervention wird folglich in all jenen Anwendungsfeldern der Psychologie diskutiert, in denen Lernprozesse initiiert werden sollen, so in der Schule (z.B. Gestaltung von Lehrerfeedback, Konstruktion von Lehr- und Lernprogrammen), im Arbeitsumfeld (z.B. Gestaltung von Mitarbeiterfeedback), sowie in der Psychodiagnostik (z.B. Intelligenztests mit Feedback, Lerntests) und Psychotherapie (z.B. Verhaltenstherapie, Biofeedback).

Die durchaus plausible und weit vertretene Annahme, Leistungsfeedback sei eine effektive Interventionsstrategie zur *Förderung* von Lernprozessen, findet jedoch nicht immer ihre empirische Entsprechung. Zahlreiche Gegenbefunde, vor allem die Beobachtung von Leistungseinbußen nach Feedback, haben dazu geführt, dass Feedback-Interventionen kritisch, z.B. als „*Double-edged Sword*“ bewertet wurden (Kluger & DeNisi, 1998, S. 67). Die Erweiterung des vorherrschenden kognitiven Feedback-Modells um nicht-kognitive Variablen Ende der 90er Jahre stellt einen Versuch dar, das teilweise erwartungswidrige Befundmuster zur Wirksamkeit von Feedback näher aufzuklären und Richtlinien für die Gestaltung effektiver Feedback-Strategien abzuleiten. Vor diesem Hintergrund soll die vorliegende Arbeit einen Beitrag zur Klärung des Zusammenhangs von Persönlichkeit und Leistung unter Feedback-Bedingungen leisten.

Erkenntnisse der differentialpsychologisch ausgerichteten Feedbackforschung finden auch ihre *Anwendung* bei der Entwicklung von *psychodiagnostischen* Methoden. So werden zunehmend Feedback-Strategien genutzt (Kyllonen, 1991; Lawrence, 2004), um validere diagnostische Aussagen über intellektuelle Leistungsvoraussetzungen zu treffen. Die Idee einer diagnostischen Nutzung von Feedback ist jedoch keineswegs neu. Bereits in den 70er Jahren wurden Intelligenztests entwickelt, die explizite Rückmeldungen an den Testanden vorsahen (Carlson & Wiedl, 1976; Guthke, 1972). Dieser diagnostische Zugang gilt als wesentlicher Bestandteil des *Dynamischen Testens* (Guthke & Wiedl, 1996).

Unter dem Begriff Dynamisches Testen werden eine Reihe von Testansätzen subsumiert, die neben der Erfassung des aktuellen *Status* einer Eigenschaft, z.B. der Eigenschaft Intelligenz, auch die Erfassung der *Veränderbarkeit* dieser Eigenschaft anstreben. Am Beispiel der Intelligenztestung, dem Hauteinsatzgebiet dynamischer Tests, soll dieses Testkonzept kurz umrissen werden. Autoren dynamischer Intelligenztests, z.B. *Kurzzeit-Lerntests* (Guthke & Wiedl, 1996), reicht die Kenntnis der zum Zeitpunkt der Intelligenztestung entwickelten Leistungsfähigkeit eines Kindes nicht aus, um valide Aussagen über seine intellektuelle Begabung zu treffen. Sie interessieren sich vor allem für die Fähigkeit des Kindes, auf Lernangebote zu reagieren. Durch die gezielte Schaffung von Lernsituationen im Testprozess selbst, z.B. durch implementierte Feedback-Schleifen, sollen in Kurzzeit-Lerntests Indikatoren für das intellektuelle *Potential* des Kindes abgeleitet werden. In diesem Sinne wird die im Testprozess auslösbare Veränderbarkeit des zu messenden intellektuellen Leistungsverhaltens erfasst. Traditionelle Intelligenztest-Werte bilden dagegen den zum Testzeitpunkt bereits entwickelten Fähigkeits-*Status* ab, der neben intellektuellen Leistungsvoraussetzungen wesentlich durch die soziale und materielle Lernumwelt des Kindes bestimmt ist. Bei defizitären Lernumwelten stellen diese Testwerte allerdings keine validen Indikatoren für die Intelligenz des Kindes dar. Unterdurchschnittliche Testleistungen dürfen dann nicht auf eine mangelnde Begabung zurückgeführt werden. Im Sinne einer fairen Testung simulieren dynamische Tests dagegen neue Lernsituationen, die Aussagen über das bisher eventuell verborgen gebliebene intellektuelle Leistungspotential des Kindes zulassen sollen. Leistungen in dynamischen Tests werden folglich als angemessene Indikatoren für zukünftiges Lernen bewertet. Im Testprozess implementiertes Feedback, wie z.B. in Kurzzeit-Lerntests, stellt dabei eine mögliche Intervention dar, die einen dynamischen Testprozess ausmacht.

Ein weiteres Ziel der Arbeit besteht nun darin, die umfängliche Wissensbasis der Forschungstradition zu Feedback-Phänomenen im Leistungsbereich für eine Leistungsdiagnostik mit Feedback nutzbar zu machen. Hier gilt es naturgemäß, differentielle Effekte von Feedback, die sich aus dem Zusammenspiel von intellektuellen und eben auch außerintellektuellen Personvariablen ergeben, in die Betrachtung einzubeziehen. Die vorliegende Arbeit stellt somit gleichzeitig einen Beitrag zur Klärung der differentiellen Validität (Jäger, 1978) von Testverfahren mit Feedback dar. Zudem

sollen Erkenntnisse zum Aufbau von geeigneten Feedback-Strukturen in dynamischen Tests abgeleitet werden.

Die im Folgenden vorgestellte experimentelle Studie dient der Analyse einer potentiellen Moderation von Leistungsverhalten unter Feedback-Bedingungen durch außerintellektuelle Personmerkmale. Dazu ist zunächst zu klären, welche Wirkmechanismen von Feedback in Lern- und Leistungssituationen angenommen werden können (Kapitel 2). Kapitel 3 ist der Frage gewidmet, welche Feedback-Strategien speziell in dynamischen Tests verwendet werden und inwiefern diese effektiv sind. Außerdem wird die bisherige Befundlage zu differentiellen Effekten bei der Bearbeitung dynamischer Tests zusammengefasst. Die Kapitel 4 bis 7 beinhalten die Planung, Durchführung und Befundinterpretation eines umfangreichen Experiments ($N = 249$) zur Untersuchung potentieller Moderator-effekte außerintellektueller Personvariablen beim Einsatz von computergestützten Reasoning-Tests für Schüler der Sekundarstufe 1 und 2.

2 Feedback-Phänomene im Leistungskontext

2.1 Zum Begriff Feedback

Im folgenden Kapitel soll der Begriff „Feedback“ konkretisiert werden. Zu diesem Zweck werden zwei historisch aufeinander folgende theoretische Zugänge vorgestellt, und ihre Anwendung im Rahmen der Entwicklung der Lehrprogramme¹ wird skizziert.

2.1.1 Das operante Modell – Feedback als Verstärker

Systematische wissenschaftliche Untersuchungen zum Lernen am Erfolg wurden erstmals von Thorndike (1898) vorgelegt. Er entdeckte im Tierversuch, dass unter verschiedenen, nach Versuch und Irrtum gezeigten Reaktionen diejenigen stärker mit der Situation verknüpft werden, die von einem befriedigenden Zustand begleitet oder gefolgt werden. Die sich dadurch einstellende Erhöhung der Auftrittswahrscheinlichkeit von Verhalten in Abhängigkeit von positiven Verhaltenskonsequenzen bezeichnete Thorndike (1911, 1932) als *Gesetz des Effekts*. Den Grundgedanken des Lernens anhand von Konsequenzen führte Skinner mit seinen Arbeiten zum *operanten Konditionieren* weiter, unter anderem mit dem Ziel, die Mechanismen der Manipulierbarkeit der Auftrittswahrscheinlichkeit operanten Verhaltens (spontan, auf die Umwelt wirkend) durch dessen Konsequenzen (Reize) aufzuzeigen.

Anfang der 60er Jahre erlangte Skinners *programmierter Unterricht* (PU, auch programmierte Unterweisung) hohe Popularität, in dessen Rahmen Grundprinzipien des operanten Konditionierens auf menschliches Lernen angewendet werden sollten (Ewert & Thomas, 1996). Danach lernt eine Person nur dann, wenn sie aktiv auf Reize reagiert, wobei Lernen hier primär als eine Veränderung von Verhaltensweisen verstanden wird. Der Lernprozess sollte dabei so geplant werden, dass er über alle Annäherungsschritte bis zum Ziel möglichst fehlerfrei verläuft, da nur durch Verstärkung des richtigen Verhaltens gelernt wird. Dieses Konzept wurde auch als Prinzip des *fehlerlosen* Lernens bekannt. Damit der Effekt einer Handlung mit den auslösenden Bedingungen verknüpft wird, müssen Rückmeldungen zudem sofort erfolgen. Hier wird ein weiteres Prinzip

¹ Der Begriff *Lehrprogramme* soll hier als übergeordneter Begriff verwendet werden, der sowohl schriftlich dargebotene programmierte Instruktionstexte (Buchprogramme) als auch die Lehrmaschinen und computergestützt dargebotene Lehrsysteme einschließt. Die Gemeinsamkeit besteht darin, dass es sich um unterschiedlich gestaltete Programme zur Vermittlung von Wissen handelt.

operanten Konditionierens deutlich, nämlich das Prinzip der unmittelbaren *Rückmeldung*.

Ein auf Basis des operanten Konditionierens entwickeltes Lehrprogramm bot dem Lernenden zunächst eine Information. Es schloss sich eine Frage an, die durch eine Aktivität meist schriftlich beantwortet werden sollte. Die Rückmeldung erfolgte sofort. Daraufhin ging der Lernende zum nächsten Lernschritt über und erhielt zunächst wieder eine Information. Der Lehrstoff war also insgesamt in kleine Einheiten zerlegt. Auf jeden Lernschritt folgte ein weiterer, der auf den vorausgehenden aufbaute (auch als lineares Programm bezeichnet, vgl. Leutner, 1992). Neben gedruckten Lehrprogrammen oder Buchprogrammen wurden so genannte Lehrmaschinen² von Skinner und seinen Mitarbeitern entwickelt. Letztere zeichneten sich dadurch aus, dass ein mechanisches Gerät die Lehrfunktionen übernahm. Beispiele für Skinners *Teaching Machines* werden bei Benjamin (1988) referiert.

Am Konzept des programmierten Unterrichts, vor allem an den Lehrmaschinen und den auf dieser Basis entwickelten frühen (ab den 60er Jahren) computergestützten Lehrsystemen, dem *computerunterstützten Unterricht* (CUU oder *Computer-aided Instruction* [CAI]) ist vielfältig Kritik geübt worden. Einen Überblick dazu gibt Fischer (1985; siehe auch Clauß, Conrad, Knöchel & Lohse, 1974; Fischer & Mandl, 1988; Fischer, Mandl, Frey, Jeuck, Schröder & Ackermann, 1988; Musch, 1999).

Die *Rückmeldungen* der Skinnerschen Lehrmaschinen und der darauf basierenden frühen CAI waren konsequenterweise ausschließlich auf positive Rückmeldung („Trefffer“) beschränkt. Auf diese Weise sollte die Rückmeldung motivieren und vor allem als Verstärker wirken, d.h. zu einer erhöhten Auftrittswahrscheinlichkeit des Zielverhaltens führen. Rückmeldungen falscher Antworten wurden möglichst vermieden, zum einen um nicht selbst gelernt zu werden bzw. einen Markierungseffekt zu verhindern, zum anderen sollten keine motivationsbeeinträchtigenden Bestrafungen („falsch“) erfolgen. Letztlich führte diese Herangehensweise dazu, dass die frühen CAI aus simplen, extrem portionierten Aufgaben bestanden, die nahezu mit Sicherheit richtig beantwortet wurden. Mit diesen Aufgaben konnten nur sehr kleine Wissensbestandteile vermittelt wer-

² Lehrmaschinen gab es allerdings schon vor den Arbeiten von B.F. Skinner. So wurde bereits 1866 eine Buchstabiermaschine durch H. Skinner konstruiert, die als Unterrichtsmittel zur Entlastung des Lehrers gedacht war. Als Ausgangspunkt für die Entwicklung von Lehrmaschinen gelten aber die von Pressey 1915 (vgl. Pressey, 1946) entwickelten Maschinen zur „Intelligenzprüfung und Informationsvermittlung“, die erstmals 1926 der Fachöffentlichkeit vorgestellt wurden (vgl. Benjamin, 1988; Leutner, 1992).

den, aus denen durch *Shaping* (Verhaltensformung) größere Komplexe entstehen sollten (vgl. Fischer, 1985; Fischer et al., 1988). Später wurden Fehler-Rückmeldungen eingeführt, um komplexe Gegenstände vermitteln zu können, die die Fehlerwahrscheinlichkeit erhöhen mussten. Der gewünschte Erfolg blieb aber aus. Fischer (1985) beklagt, dass der *Informationswert* der Rückmeldung auf diese Weise nicht erhöht werden konnte, sondern lediglich der Anteil an Bestrafung. Er sieht darin einen wesentlichen Grund für das Scheitern der frühen CAI.

Neben dem Ansatz nach Skinner muss noch auf einen zeitgleichen zweiten, den *reizzentrierten Ansatz nach Crowder* (1959) hingewiesen werden. Im Gegensatz zum reaktionszentrierten Ansatz nach Skinner werden beim reizzentrierten Ansatz falsche Antworten als förderlich betrachtet, da sie Defizite aufdecken. Bei einem Fehler verzweigt das Lehrprogramm dann auch zu einer helfenden Lehreinheit (gleiches gilt auch für die in Abschnitt 3.1.1 vorgestellten Kurzzeit-Lerntests). Zudem wird entgegen der Annahmen nach Skinner davon ausgegangen, dass Fehler motivieren können, treten sie nicht zu häufig auf. Falsche Antworten bzw. Fehler sollen hier nicht grundsätzlich vermieden werden, denn sie steuern den Lernprozess. Entsprechend ist es möglich, auch komplexe Aufgaben zu implementieren (vgl. Leutner, 1992). Als theoretischer Hintergrund dient die kognitive Lerntheorie nach Tolman (1932). Entscheidend ist, dass Lernen hier als eine Organisation oder Reorganisation kognitiver Strukturen angesehen wird und nicht entsprechend Skinner als eine Erhöhung von Verhaltenswahrscheinlichkeiten.

Zusammenfassend beinhaltet Lernen durch Verstärkung folgende Grundkonzepte: Auf ein Verhalten in einer bestimmten Situation muss eine positive Konsequenz folgen, damit es in Form einer Reiz-Reaktionsverbindung gelernt wird. Lernen impliziert dabei die Erhöhung der Wahrscheinlichkeit, dass dieses Verhalten in Zukunft auftritt. Durch *Feedback* wird gelernt, da Feedback als Verstärker wirkt, also die Auftretenswahrscheinlichkeit des erwünschten Verhaltens erhöht.

Die Annahme einer automatischen Wirkung von Verstärkung gemäß Thorndike (1932) gilt inzwischen allerdings als widerlegt (beispielsweise Herrmann & Stapf, 1973; Mix, 1981; Stapf & Degner, 1981; Stapf, Fischer & Degner, 1986). Thorndike postulierte mit seinem revidierten Effektesetz eine asymmetrische Wirksamkeit von Belohnung und Bestrafung in Lernsituationen zugunsten von Belohnung. Er interpretierte diese asymmetrische Wirkung mit einer universelleren, direkteren und damit sichereren Be-

einflussung des Lernens durch Belohnung und zwar aufgrund von Lust- versus Unlustempfindungen (vgl. Mix, 1981). Diese Position ist vielfältig kritisiert worden. Thorndikes Untersuchungen wird beispielsweise ein Isolationseffekt zugunsten der Belohnung angelastet, d.h. Belohnungen wurden durch ihr selteneres Auftreten im Vergleich zu Bestrafungen artifiziell akzentuiert. Somit war der beobachtete Lerngewinn bei Belohnungen höher (Zirkle, 1946). Herrmann und Stapf stellen bereits 1973 nach Durchsicht der umfangreichen Literatur fest, dass die Frage nach der *relativen* Wirksamkeit von Belohnung und Bestrafung nicht pauschal zugunsten einer von beiden Modalitäten beantwortet werden kann und schlussfolgern:

„Die Rückmeldungswirkung ist also weitgehend durch vermittelnde kognitive Prozesse bestimmt. Von einer gewissermaßen *automatischen* Wirkung positiver, negativer oder fehlender Rückmeldung, wie sie etwa der frühe Behaviourismus (vgl. Thondikes Effektgesetz) unterstellt, kann keine Rede sein“ (S. 122, Hervorhebung zugefügt).

Kluger und DeNisi (1996) fassen viel später zusammen: „Yet, despite all logical and empirical shortcomings of the law of effect, it had a substantial influence on FI [Feedback Intervention, d.V.] researchers. Therefore the law of effect was blamed by some for hindering FI research“ (S. 259).

2.1.2 Das kognitive Modell – Feedback als Information

Da eine ausschließliche Konzeptualisierung von Feedback als Verstärker wenig erfolgreich war, wurde ab den 70er Jahren im Kontext der kognitiven Wende in der Psychologie versucht, die Wirksamkeit von Feedback mit Hilfe von kognitiven Variablen zu erklären. Dafür wurden Erkenntnisse aus der Kybernetik (Wiener, 1948, 1952) herangezogen (vgl. Musch, 1999), wobei die Feststellung eines *Regelungs- und Rückmeldungsprinzips* in der Funktionsweise biologischer und technischer Systeme grundlegend war. Demnach wird ein stabiles Systemverhalten bei zufälligen oder dynamischen Veränderungen bzw. Störungen der Eingangsgrößen dann am besten erreicht, wenn der aktuelle Fehler zurückgemeldet und zur Fehlerbehebung verwendet wird.

Ein Rückkopplungssystem, bei dem kognitive Prozesse beteiligt sind (im Gegensatz zu technischen Rückkopplungssystemen), das *TOTE-Modell*, wurde von Miller, Galanter und Pribram bereits 1960 konzipiert. Als Grundeinheit des Verhaltens wird eine Folge von Prüfphasen und Handlungsphasen (Test-Operate-Test-Exit) vorgeschla-

gen, die bis zur Zielerreichung durchlaufen werden (vgl. auch Dunckel, 1986). Unter Beachtung der aktuellen *Ist-Soll-Differenz* wird eine Regelgröße (Ist-Wert) solange verändert, bis die Differenz verschwindet und damit das Ziel (Soll) erreicht ist. Der Rückfluss von Ergebnisdaten (= Rückmeldung im Sinne einer Information) wird dabei vorausgesetzt. In Anlehnung an das TOTE-Modell wurden im Rahmen der psychologischen Handlungstheorien die zyklische Einheit nach Volpert (1982) und die Vergleichs-Veränderungs-Rückkopplungseinheit nach Hacker (1978) als Grundelemente von *Handlungen* entwickelt. Eine Handlung wird somit als Regelvorgang verstanden. Diese Konzeptualisierung von Handlungen wurde deshalb als Handlungsregulation bekannt.

Mit der Steuerung und Regelung speziell von *Lernprozessen* hat sich Leutner (1992) beschäftigt und auf Konzepte der Regelungstechnik zur Optimierung von Lernprozessen, speziell zur Entwicklung von Lehrprogrammen hingewiesen. Es soll hier nur auf zwei Grundbegriffe eingegangen werden: In der Regelungstechnik wird streng unterschieden zwischen *Regelung* und *Steuerung* (beispielsweise Müller-Schwarz, 1969). Demnach liegt eine Steuerung eines Prozesses dann vor, wenn ein offener Wirkungsweg gegeben ist, bei dem keine Kenntnisse über den Ist-Zustand des gesteuerten Prozesses gewonnen werden bzw. vorhandene Kenntnisse nicht berücksichtigt werden. Von Regelung eines Prozesses (hier Lernprozess) spricht man dagegen dann, wenn ein geschlossener Wirkungsweg gegeben ist. Die Kenntnis des Ist-Zustandes des gesteuerten Prozesses wird herangezogen, um Eingriffe in den Prozess vorzunehmen. Bei einer kybernetischen Betrachtungsweise von Lehr- und Lernprozessen werden nun zwei Systeme unterschieden, die Informationen austauschen: ein lehrendes und ein lernendes System. *Lehren* dient dann zunächst einer Steuerung von Lernprozessen. Eine Regelung liegt dann vor, wenn das lehrende und das lernende System nicht nur einen offenen Wirkungsweg bilden (beispielsweise bei einer Vorlesung), sondern einen geschlossenen Wirkungsweg darstellen, der einen Austausch von Information ermöglicht, d.h. eine Interaktion (z.B. bei einem Seminar). Dies wird als Grundlage für eine dynamische Anpassung der Lernumgebung an die aktuellen Bedürfnisse des Lernenden verstanden (Leutner, 1992). Rückmeldungssysteme sind demzufolge immer Regelungssysteme (vgl. auch Fischer, 1985). Folgen von Veränderungen durch das System wirken auf das System zurück. Es ist damit in der Lage, sich selbst zu steuern. Somit stellen auch Testverfahren mit Feedback (z.B. Kurzzeit-Lerntests, siehe Abschnitt 3.1.1) Regelungssysteme dar.

Auf der Basis eines kognitiven Feedback-Modells wurden ab den 70er Jahren beginnend mit Carbonell (1970) mit Hilfe der Künstlichen Intelligenz neuartige nämlich „intelligente“ computergestützte Lehrsysteme entwickelt. Diese wurden oft als *Intelligent Computer-aided Instruction (ICAI)* oder *Intelligent Tutoring Systems (ITS)* bezeichnet. Grundsätzlich werden drei notwendige Bestandteile eines ICAI-Systems unterschieden: ein kommunizierbares Modell des Gegenstandsbereichs (Soll-Zustand des Lernenden), ein dynamisch sich veränderndes Wissensmodell des Lernenden (Ist-Zustand des Lernenden) und Kenntnisse darüber, wie eine Diskrepanz zwischen Soll-Zustand und Ist-Zustand behoben werden kann. Bezüglich der *Rückmeldungen*, die innerhalb der ICAI-Systeme dem Lernenden angeboten werden, ist festzuhalten: Durch den Vergleich zwischen dem Expertenwissen (Soll-Zustand) und dem diagnostizierten Kenntnisstand des Lernenden (Ist-Zustand) können in einem ICAI-System sehr gezielt fehlerabhängige Rückmeldungen gegeben werden, mit deren Hilfe die Person ihren Kenntnisstand erweitern und sich so dem Soll-Zustand annähern soll (vgl. auch Fischer, 1985; Musch, 1999). Einen Überblick über die Kritik an den ICAI-Systemen gibt Leutner, der 1998 die praktische Bedeutsamkeit der ICAI bereits als gering einschätzt. Die wichtigste Kritik aus Sicht der Psychologie besteht wohl darin, dass die implementierten Lehrstrategien bzw. Lehrmethoden nur selten instruktionspsychologisch begründet oder experimentell überprüft, sondern eher intuitiv durch die Systementwickler erstellt wurden (Leutner, 1998; vgl. auch beispielsweise Musch 1999; einen ersten Überblick geben beispielsweise Mandl & Lesgold, 1988).

Zusammenfassend kann Folgendes festgehalten werden: Innerhalb des kognitiven Modells hat Feedback vor allem eine informative Funktion. Negatives Feedback ist besonders deshalb nützlich, weil es auf Defizite im Sinne von Ist-Soll-Diskrepanzen hinweist und somit hilft, diese zu überwinden (siehe auch Kluger & DeNisi, 1996). Diese Idee findet sich schon bei Crowder (1959). Verzweigte Programme nach Crowder werden von Leutner (1992) deshalb bereits den Regelungssystemen zugeordnet. Ebenso sind Testverfahren mit implementiertem Feedback als Regelungssysteme konzipiert. So erhalten Testanden in Kurzzeit-Lerntests als eine Form dynamischer Intelligenztests zu jeder ihrer Antworten fehlerspezifische Rückmeldungen. Das Ziel besteht darin, neben dem Leistungsstatus des Testanden auch das intellektuelle Potential – das so genannte intellektuelle Veränderungspotential – zu erfassen.

Es wird zudem von einem geschlossenen Wirkungsweg, oft als Rückkopplungsschleife bezeichnet, ausgegangen, so dass der Lernende den Lernprozess selbst steuern kann. Man spricht hier auch von der Selbstregelung bzw. Selbstregulation (vgl. Musch, 1999). Gleichzeitig wird angenommen, dass der Lernende die *Rückmeldungen* ausschließlich als Information interpretiert, verarbeitet und nutzt. Feedback geht demnach als Information in einem zwischen Reiz und Reaktion geschalteten Informationsverarbeitungsprozess ein und beeinflusst dadurch das darauf folgende Verhalten in gewünschter Richtung.

Die Lehrmaschinen der ersten Hälfte des 20. Jahrhunderts und früher, die überwiegend auf Skinner und Crowder zurückzuführen sind, wurden in den 60er und 70er Jahren durch die CAI-Systeme ersetzt (siehe Abschnitt 2.1.1 zum operanten Feedback-Modell); in den 70er kamen die auf Modellen der Künstlichen Intelligenz basierenden ICAI-Systeme hinzu (siehe Abschnitt 2.1.2 zum kognitiven Feedback-Modell). Hier beginnt auch die Entwicklung der dynamischen Tests, die zunächst als Buchprogramme und ab Ende der 80er Jahre auch als computergestützte („Lern“)Programme entwickelt wurden, z.B. 1987 eine erste Urform der Adaptiven Computergestützten Intelligenz-Lerntest-Batterie (ACIL, siehe Abschnitt 3.1.1). In den 90er Jahren erfolgte eine Erweiterung durch die im Folgenden kurz vorzustellenden multimedialen Lernumgebungen (z.B. Gerstenmaier & Mandl, 1995; Mandl, Gruber & Renkl, 1997a).

2.1.2.1 Das konstruktivistische Modell – Feedback als Informationsangebot

Konstruktivistische Instruktionsansätze, die auch ein informationstheoretisches Feedback-Modell vertreten, streben eine *Förderung* des Lernens anstelle einer Steuerung des Lernens an. Dem Lernenden soll im Sinne des entdeckenden Lernens viel stärker die Möglichkeit für eine *selbstgesteuerte* Systemexploration gegeben werden.

Eine konstruktivistische Auffassung des Wissenserwerbs geht nach Gerstenmaier und Mandl (1995) zum einen von einer *Wissenskonstruktion* des Lernenden aus und fordert zum anderen die enge Verbindung von Wissenserwerb und *Wissensanwendung*. Da also nach konstruktivistischem Verständnis Wissen nicht einfach vom Lehrenden zum Lernenden transportiert, sondern individuell im Lernprozess konstruiert wird, spielt die *Situation*, in der der Lernprozess stattfindet, eine zentrale Rolle. Für dieses Konzept wurde deshalb der Begriff des „sitierten Lernens“ eingeführt. Da Wissen und damit auch Lernen demnach situiert bzw. kontextbezogen sind, sollen Lern- und Anwen-

dungssituationen möglichst ähnlich gestaltet sein, um einen Wissenstransfer zu erleichtern (Mandl et al., 1997b). Die Theorie des Situierten Lernens kann als Synthese von kognitiven Theorien und Verhaltenstheorien angesehen werden (Greeno, 1992, zitiert nach Gerstenmeier & Mandl, 1995). Gerstenmaier und Mandl (1995) halten fest, dass eine konstruktivistische Lernumgebung dem Lernenden Freiheitsgrade bieten muss: neue Inhalte dürfen nicht als fertiges System präsentiert werden. Allerdings muss der Lernende die bestehenden Freiheitsgrade der Lernumgebung auch erkennen und nutzen können, d.h. der subjektiv wahrgenommene Handlungsspielraum ist entscheidend. Die Gestaltung solcher komplexer Lernumgebungen wird vorrangig mit dem Einsatz von *Multimedia* versucht.

Zur Veranschaulichung ein Beispiel: Das Computerprogramm „Thyroidea“ (Fischer, Gräsel, Mandl, Gärtner & Scriba, 1994, zitiert nach Mandl et al., 1997b) ist ein Lehrprogramm, dessen didaktische Konzeption sich explizit an den Instruktionsmodellen zum situierten Lernen orientiert. Mit diesem Programm sollen angehende Ärzte durch die Bearbeitung computersimulierter Fälle, die mit visuellen Medien (beispielsweise einer Videoeinspielung) veranschaulicht sind, anwendungsorientiertes Wissen zur Diagnose und Therapie von Schilddrüsenerkrankungen erwerben. Der dargebotene Fall enthält ein authentisches Problem aus der klinischen Praxis. Die Fallbearbeitung ist am Ablauf der Patientenversorgung in der Klinik orientiert. Durch interaktive Multimedia-Elemente soll der Lernende sein Wissen aktiv auf das Problem beziehen und somit Zusammenhangswissen und Handlungswissen erwerben. Zudem erhält er Unterstützung durch einen Experten, der seine Vorgehensweise erklärt.

Bezüglich konstruktivistischer Lernumgebungen wird der hohe Aufwand kritisiert, der mit dem Erstellen der entsprechenden Lernmedien verbunden ist. Auch gibt es kaum Beschränkungen für die Gestaltung, so dass die Gefahr der Herstellung „ineffektive(r) oder gar sinnlose(r) Produkte“ groß ist (Mandl et al., 1997b, S.176). Zum anderen profitieren auch nicht alle Lernenden gleich von der Verwendung multimedialer Lernumgebungen (Aptitude-Treatment-Effekte, Snow, 1989). Mandl et al. (1997b) berichten Befunde, die nahe legen, dass insbesondere bei guten Lernvoraussetzungen, wie ein hohes Vorwissen und/oder hohe Intelligenz, konstruktivistisch orientierte Lehrprogramme Vorteile gegenüber traditionellen Lehrprogrammen bieten.

Zusammenfassend kann festgehalten werden: Konstruktivistische Instruktionsmodelle gehen von einem aktiven, selbstgesteuerten Lernenden aus. Der Lehrende soll weniger Vermittler oder Präsentator von Wissen sein, sondern vielmehr ein Mitgestalter von Lernumgebungen und Unterstützer von Lernprozessen (vgl. Gerstenmaier & Mandl, 1995). Mit den Lehrprogrammen im Sinne multimedialer Lernumgebungen soll nach konstruktivistischen Lernprinzipien entdeckendes und kooperatives Lernen er-

möglichst werden. *Feedback* wird dabei als ein (Informations-)Angebot an den Lernenden verstanden. Ob und inwieweit er darauf zurückgreift, bleibt ihm überlassen.

2.1.2.2 Das sozialpsychologische Modell – Feedback als Kommunikation

Das Phänomen Feedback lässt sich auch aus sozialpsychologischer Perspektive betrachten. Hier werden Rückmeldungen als Bestandteil sozialer Interaktionen aufgefasst. Sie stellen den Rückfluss von Information von einem Empfänger zu einem Sender dar (siehe Kommunikationsmodelle, z.B. nach Frey & Greif, 1994; Shannon & Weaver, 1949). Beachtet man die von Frey und Greif (1994) postulierte Doppelfunktion menschlicher Kommunikation, wonach eine Person gleichzeitig Sender und Empfänger ist, ist jede Informationsübertragung als eine Rückmeldung infolge einer vorhergehenden Informationsübertragung aufzufassen. Insofern lässt sich Feedback im weitesten Sinne auch als Kommunikation verstehen. Dann stellt Feedback eine gerichtete Informationsübertragung dar und beinhaltet den *Einfluss* von einem Sender-System auf ein Empfänger-System (vgl. Definition von Kommunikation nach Frey und Greif, 1994).

Die übermittelte Information kann jedoch verfälscht werden. Eine wesentliche Fehlerquelle wird in der Encodierung und Decodierung der Information, also der *Informationsverarbeitung* durch den Sender bzw. Empfänger gesehen. Im Rahmen einer sozialpsychologischen Perspektive geht *Feedback* demnach als Information in einen Informationsverarbeitungsprozess ein und beeinflusst dadurch das darauf folgende Verhalten in gewünschter Richtung, vorausgesetzt die Information wird adäquat wahrgenommen (decodiert), verarbeitet und genutzt (encodiert).

Die Durchführung eines psychologischen Tests kann durchaus als eine Form von sozialer Interaktion aufgefasst werden. Es lässt sich somit auch argumentieren, dass durch die besondere Gestaltung eines dynamischen Tests (Test mit Feedback, siehe Kapitel 3) im Vergleich zum traditionellen Statustest eine Interaktionsform zwischen Testleiter und Testand hergestellt wird, die den dynamischen Test als solchen erst ausmacht (vgl. auch hinsichtlich einer nicht-standardisierten dynamischen Teststrategie Haywood & Tzuril, 1992). Das reine „Frage-Antwort-Spiel“ der traditionellen Testung soll dabei durch eine stärker auf die Zusammenarbeit orientierende Interaktion zwischen Testleiter und Testand durch die im Testprozess implementierten Interventionen, z.B. standardisiertes Feedback, abgelöst werden (Guthke, Beckmann & Wiedl, 2003, S. 227). Eine Testdurchführung, die einen Rückfluss an Information über die erzielte Leistung zulässt oder gerade anzielt, sollte ausgleichend wirken auf die traditionell am ehesten durch Asymmetrie zu Ungunsten des Testanden bestimmte Interaktionsform zwischen Testleiter und Testand. Eine unmittelbare Leistungsrückmeldung im Test selbst würde allerdings gegen das Standardisierungsgebot in klassischen Leistungstests verstoßen. Die Option einer testandenangemessenen Leistungsrückmeldung nach dem Test wird in der psychologischen Praxis eher selten genutzt.

Die kommunikationstheoretische Konzeptualisierung des Feedback-Begriffs wird vor allem in der arbeits- und organisationspsychologischen Forschung sowie in angrenzenden Wissenschaften und ihren Anwendungen, beispielsweise im Rahmen der Perso-

alentwicklung und im Verkaufstraining (Vilsmeier, 2000) bevorzugt (z.B. Farr, 1991; Fenger, 1998; Kolleker, 1999; Slembek & Geißner, 1998; von Hornstein & von Rosenstiel, 2000). Von Hornstein und von Rosenstiel (2000) definieren Feedback auch wie folgt: „Feedback ist eine Rückmeldung über die Wahrnehmung einer Person durch eine andere und stellt damit die Basis für die Veränderung des eigenen Verhaltens dar“ (S. 209). Hierzu werden zahlreiche Regeln vorgeschlagen, die diesen Prozess unterstützen sollen. Zum Beispiel sollte Feedback eher beschreibend als bewertend, eher konkret als allgemein, eher verhaltensbezogen als charakterbezogen, eher sofort und situationsbezogen als verzögert gestaltet sein. Anwendungen im Rahmen der klinisch-psychologischen Praxis finden sich in Encounter- und Selbsterfahrungsgruppen, aber auch innerhalb der verschiedenen psychotherapeutischen Techniken (z.B. Biofeedback, Verhaltenstherapie).

2.1.3 Schlussfolgerungen

In Abschnitt 2.1 wurde eine konzeptuelle Charakterisierung von „Feedback“ versucht. Dabei zeigte sich, dass die anfänglich postulierte Gleichsetzung von Feedback und Verstärkung durch Vertreter der Feedback-Forschung kritisch beurteilt und stattdessen ein informationstheoretisches Verständnis favorisiert wird. Folgende vier Konzeptionen des Begriffes „Feedback“ lassen sich aus der obigen Darstellung ableiten:

- *Feedback als Verstärkung:* Durch Feedback wird gelernt, weil Feedback als Verstärker wirkt, also die Auftrittswahrscheinlichkeit des erwünschten Verhaltens erhöht.
- *Feedback als Information:* Feedback geht als Information in einem zwischen Reiz und Reaktion geschalteten Informationsverarbeitungsprozess ein und beeinflusst dadurch das darauf folgende Verhalten in gewünschter Richtung.
- *Feedback als Informationsangebot:* Feedback beinhaltet ein (Informations-)Angebot an den Lernenden. Ob und inwieweit er darauf zurückgreift, bleibt ihm überlassen.
- *Feedback als Kommunikation:* Feedback stellt im Sinne einer gezielten Einflussnahme die gerichtete Informationsübertragung von einem Sender-System zu einem Empfänger-System dar, vorausgesetzt die Information wird adäquat wahrgenommen, verarbeitet und genutzt.

Die beiden zuletzt vorgestellten Ansätze stellen dabei eine Erweiterung oder Ergänzung des informationstheoretischen Gedankens dar, die in der besonderen Betonung der Informationsnutzung und adäquaten Informationsverarbeitung liegt. Die von den

verschiedenen Autoren innerhalb der Forschungstradition zu Feedback-Phänomenen vertretenen Auffassungen zum Feedback-Begriff lassen sich im Allgemeinen einer dieser vier Konzeptionen zuordnen.

Bei der Erklärung der Feedback-Wirkung lässt sich also wie ausgeführt unterscheiden, ob diese eher motivational (verstärkend, belohnend) oder eher informativ interpretiert wird. Eine Hervorhebung des Informationswertes von Feedback schließt streng genommen Überlegungen zu motivationalen und affektiven Wirkungen von Rückmeldungen aus, da sie mit einer Vorstellung von Regelung prinzipiell unvereinbar sind (vgl. auch Fischer, 1985). Mandl, Fischer, Frey und Jeuck halten aber bereits 1985 fest:

„Rückmeldung, d.h. Information über den Erfolg eigenen Lernens oder Tuns ist nicht neutral und ausschließlich auf die Sachinformation Richtig/Falsch oder Fehler/Treffer beschränkt, sondern hat auch eine belohnende oder bestrafende Komponente affektiver Art“ (S. 180).

Auch Stapf et al. (1986) merken an, dass ein

„Außerachtlassen motivationaler Anteile der Rückmeldung unbefriedigend (ist), auch wenn vorwiegend von einer kognitiven Betrachtungsweise ausgegangen wird, da die Einwirkung der Rückmeldung auf das Verhältnis von gebotener und genutzter Information geklärt werden muss“ (S.222).

Fehler werden demnach nicht nur als Information, dass der gewünschte Erfolg nicht eingetreten ist, verarbeitet, sondern durchaus als Misserfolg erlebt. Jegliche einseitige Betrachtung von Feedback, also motivierend oder informierend scheint demnach unangemessen. Vielmehr ist es nahe liegend anzunehmen, dass Feedback informiert *und* belohnt oder bestraft, was zusätzliche kognitive und aber auch affektive Verarbeitungsprozesse anregt. Das wirft die spannende Frage auf, unter welchen Bedingungen oder bei welchen Personen Feedback Motivations- und/oder Leistungsgewinne oder -einbrüche bewirkt und warum? Diese Fragen sollen im nächsten Abschnitt erörtert werden.

Unter *Feedback* (Rückmeldung) soll im Folgenden eine leistungsbezogene *Information* verstanden werden, die eine Person innerhalb einer Lern- und Leistungssituation erhält. Neben einer vorrangig *informativen* Funktion wird auch eine *affektive Funktion*

von Feedback angenommen. Dabei wird keine automatische Reaktion seitens des Lernenden erwartet, sondern es ist diesem freigestellt, die zusätzlich angebotene Information im Lernprozess auch zu verarbeiten und zu nutzen.

2.2 Leistungsfeedback – eine sinnvolle Intervention?

Im vorangegangenen Abschnitt wurde versucht, den Feedback-Begriff im Rahmen der psychologischen Forschung zu konkretisieren und die Entwicklungslinien innerhalb der Forschungstradition zu Feedback-Phänomenen abzustecken. Dabei wurde bereits auf die kontroverse Diskussion zur Effektivität von Feedback im Leistungsbereich hingewiesen. Das folgende Kapitel dient der Zusammenfassung der in der Literatur dargelegten aktuellen Erkenntnisse zu generellen Effekten von Leistungsfeedback.

2.2.1 Gegenwärtige Befundlage zur Effektivität von Leistungsfeedback

Feedback, d.h. die Rückmeldung einer Ist-Soll-Diskrepanz, wird, wie in Abschnitt 2.1 beschrieben, in Theorien zum Lernen und darauf aufbauenden instruktionalen Modellen als grundlegender Mechanismus verstanden, der zur geeigneten Adaption des Lernprozesses und in dessen Folge zu verbesserten Leistungen führen soll. Der Anwendungsbereich von Feedback-Interventionen ist weit gefächert und spiegelt sich in der umfangreichen Literatur zu Feedback-Phänomenen in den Anwendungsgebieten der Psychologie wider, so im:

- pädagogischen Kontext (z.B. Gestaltung von Lehrerfeedback, Konstruktion von Lehr- und Lernprogrammen, Untersuchung so genannter paradoxer Phänomene von Lob und Tadel, Rheinberg & Weich, 1988; Rheinberg, 1998),
- arbeits- und organisationspsychologischen Kontext (z.B. Gestaltung von Vorgesetzten- oder Mitarbeiterfeedback) als auch im
- klinischen Kontext (z.B. Feedback im psychotherapeutischen Setting, Biofeedback).

Kluger und DeNisi (1998) bewerten leistungsbezogenes Feedback dementsprechend als „one of the mostly widely applied psychological interventions“ (S. 67). Die ursprünglich aus dem Effektgesetz nach Thorndike (1932, siehe auch Abschnitt 2.1.1) abgeleitete und weit verbreitete Annahme, dass Feedback generell positive Leistungseffekte bewirkt, hielt der empirischen Prüfung jedoch nicht stand. Während Ammons (1956) in seinem viel zitierten und später kritisierten (vor allem Kluger & DeNisi, 1996; 1998) Beitrag nach zusammenfassender Betrachtung der bis dahin 50jährigen For-

schungstradition zu Feedback-Phänomenen noch schlussfolgerte, dass Feedback im Sinne von *Knowledge of Results* Lernen und Motivation signifikant verbessert, belegen Metaanalysen in den 90er Jahren (Bangert-Drowns, Kulik, Kulik & Morgan, 1991; Kluger & DeNisi, 1996) vor allem die große Bandbreite der Effekte von Feedback und stellen insbesondere die bis dahin ignorierten negativen Feedback-Effekte zur Diskussion. So arbeiten Kluger und DeNisi (1996) heraus: „... when Ammons (1956) reviewed the FI [Feedback Intervention, d.V.] (or KR [Knowledge of Results, d.V.]) literature, an uncritical view of the effectiveness of KR had already begun to develop, despite considerable evidence to the contrary.“ (S. 256) und bewerten seine viel zitierte Arbeit als „unfortunate milestone in FI research“ (S.256). In der Folge finden sich in der Literatur einige Arbeiten, die vor allem „unerwartete“, negative oder fehlende Effekte von Leistungsfeedback besprechen (z.B. bezogen auf Feedback in Intelligenztests Delgado & Prieto, 2003; Rousseau & McKelvie, 2000; Stankov & Crawford, 1997). Necka und Stettner (2002) überschreiben ihren Artikel sogar mit „No Feedback, Please!“ (S.31). Insbesondere Kluger und Mitarbeiter (1996, 1998, 2001, 2004) aber haben seither einen wesentlichen Beitrag zur Aufklärung von leistungsbezogenen Feedback-Effekten geleistet, vor allem indem sie versucht haben, ein theoretisches Modell anzubieten – die Feedback-Interventions-Theorie – (Kluger & DeNisi, 1996), das in der Lage sein soll, auch bisher unerwartete Effekte von Leistungsfeedback zu erklären und vorherzusagen. Sie fassen zusammen:

„We argue that a considerable body of evidence suggesting that feedback intervention (FI) effects on performance are quite variable has been historically disregarded by most FI researchers. This disregard had led to a widely shared assumption that FIs consistently improve performance. Fortunately, several FI researchers have recently recognized that FIs have highly variable effects on performance, such that in some conditions FIs improve performance, in other conditions FIs have no apparent effects on performance, and yet others debilitate performance.“ (S. 254).

Die reiche Wissensbasis der Forschungstradition zu Feedback-Phänomenen, insbesondere die zu den unerwartet ausbleibenden oder gar negativen Feedback-Effekten, ist bisher in der Forschung zum dynamischen Testen (siehe Kapitel 3) weitestgehend ignoriert geblieben. Ebenso sind auch umgekehrt bis auf eine Arbeit von Bethge,

Carlson und Wiedl (1982) keine Studien zur psychometrischen Qualität von Testverfahren mit Feedback in die genannten Metaanalysen zu leistungsbezogenen Feedback-Effekten eingegangen. Umso Gewinn bringender scheint eine Integration der Erkenntnisse beider Forschungstraditionen, die mit dieser Arbeit angestrebt wird. Unter anderem sollen wesentliche Erkenntnisse der Feedback-Forschung zur Gestaltung von Feedback-Strategien aufgearbeitet werden, um Schlussfolgerungen für einen Testansatz mit Leistungsrückmeldungen zu ermöglichen (siehe vor allem Abschnitt 2.2.3).

Der Umfang der Forschungslage zu leistungsbezogenen Feedback-Effekten zeigt sich in der von Kluger und DeNisi (1996) in ihrer Metaanalyse durchgesehenen Literatur. Sie finden bereits 1992 etwa 3000 englischsprachige Arbeiten, die sich mit Effekten von Feedback auf Leistungsvariablen beschäftigen. Aber auch zahlreiche aktuelle Studien belegen das gegenwärtig bestehende Interesse an psychologischen Feedback-Prozessen. So wurden in den ersten 7 Monaten des Jahres 2004 immerhin 252 psychologische Arbeiten zu Feedback-Phänomenen publiziert (Quelle: PsycINFO), davon 48 zu Feedback-Effekten im Leistungsbereich.

Allerdings bezogen Kluger und DeNisi (1996) aufgrund methodischer Mängel der zitierten Studien, z.B. das Nichtvorhandensein einer Kontrollgruppe bei 37% der Untersuchungen, nur einen sehr umgrenzten Teil (5%) in ihre Metaanalyse ein. Auf der Basis von 131 berücksichtigten Studien wurden 607 Effektgrößen berechnet, die auf einer Stichprobe von 12652 Personen beruhen. Der durchschnittliche an der Stichprobengröße gewichtete Effekt von leistungsbezogenem Feedback betrug weniger als eine halbe Standardabweichung (Cohens $d = .41$). Erwartungswidrig waren zudem 38% der gefundenen Effekte negativ. Bangert-Drowns et al., (1991) bestimmen in ihrer Metaanalyse einen nur schwachen, generellen Feedback-Effekt (Cohens $d = .26$), und berichten ebenfalls von 18 (von 58) negativen Effektgrößen (31%) aus 40 einbezogenen Studien. Sie folgern: „ ... one is struck by the generally weak effects of feedback on achievement. ... Surely, attributing negative effects to feedback is counterintuitive.“ (S. 224).

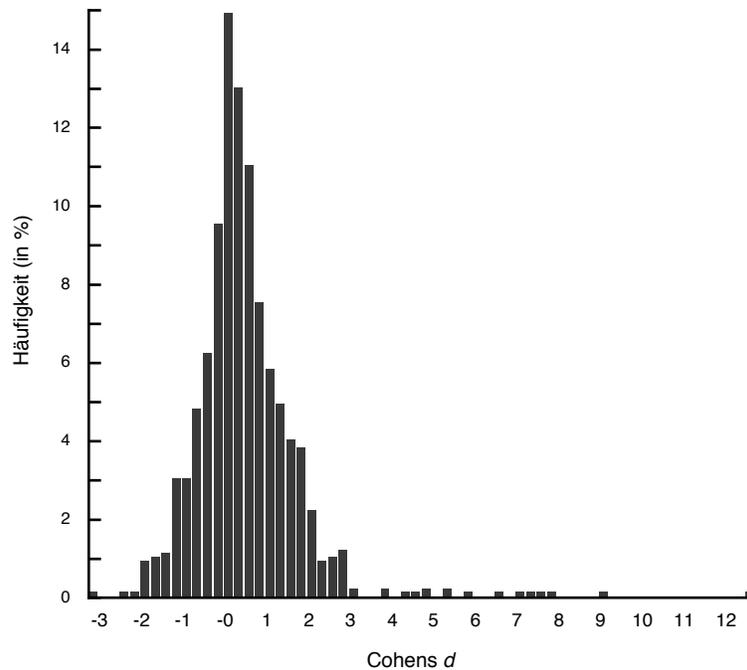


Abbildung 2.1 Verteilung von 607 Effektgrößen von Feedback-Interventionen auf die Leistung (nach Kluger & DeNisi, 1996, S. 258).

Abbildung 2.1 zeigt die Verteilung der innerhalb der Metaanalyse von Kluger und DeNisi (1996) betrachteten Effektgrößen, wobei neben den negativen Effekten die Varianz der Feedback-Effekte auffällt. Es sei auch angemerkt, dass innerhalb dieser Metaanalyse keine Effekte von Feedback auf Latenzzeiten und Reaktionszeiten als weitere Leistungsindikatoren ausgemacht werden konnten. Die dargestellten Befunde sehen Kluger und DeNisi (1996) als Nachweis dafür an, dass geringe oder negative Feedback-Effekte nicht als singuläre Artefakte zu verstehen sind und fordern konsequenterweise die Suche nach *Moderatoren* von leistungsbezogenen Feedback-Effekten. Dazu stellen sie selbst ein theoretisches Modell vor, das unter Berücksichtigung von außerintellektuellen Variablen auch negative Feedback-Effekte erklären und vorhersagen soll.

2.2.2 Die Feedback-Interventions-Theorie

Trotz der umfänglichen Kritik dominierte das revidierte Effektgesetz nach Thorndike (1932) über 70 Jahre Forschungsarbeit zu Feedback-Phänomenen (siehe Kapitel 2.1). Bei Betrachtung der Literatur fällt jedoch auf, dass es nach nunmehr etwa 100 Jahren Feedback-Forschung an alternativen theoretischen Ansätzen fehlt, die über die reine Feststellung eines informationstheoretischen Verständnisses von Feedback in Ablösung eines verstärkerorientierten Modells hinausgehen. So lassen sich strenggenommen inner-

halb eines informationstheoretischen Feedback-Modells Befunde (z.B. Bangert-Drowns, et al., 1991) nicht erklären, die zeigen, dass zusätzliche dosierte Information die Leistung beeinträchtigen kann. Nach einem kognitiv-informationstheoretischen Verständnis von Feedback dürfte nur ein Zuviel an Information Leistungseinbußen bedingen, da die Verarbeitungskapazität des Arbeitsgedächtnisses begrenzt ist. Leistungsverschlechterungen nach Richtig/Falsch-Feedback sind aber innerhalb eines rein kognitiven Modells nicht erklärbar. Werden jedoch andere, *nicht-kognitive* Erklärungsmodelle in der Betrachtung zugelassen, finden sich einige Forschungsparadigmen, z.B. die Theorie der gelernten Hilflosigkeit (Seligman, 1975), die die Wirkung von Leistungsrückmeldungen einschließlich möglicher Moderatoren unter bestimmten Bedingungen (hier Misserfolgsbedingungen) erklären können. Andere Ansätze, die zur Erklärung und Vorhersage von kognitiven, affektiven und behaviouralen Reaktionen auf Erfolg und Misserfolg herangezogen werden sollten, sind die *Attributionstheorie*, welche kognitive Reaktionen auf Erfolge und Misserfolge erklärt und Leistungsvorhersagen trifft (Weiner, 1986), das Modell des *Begabungskonzepts* als zentraler Moderator in Leistungssituationen (Meyer, 1984), das Konzept der *Selbstwirksamkeit* (Bandura, 1977) als Erwartungstendenz in Leistungssituationen, *Motivationstheorien* (z.B. Heckhausen, 1963, 1989) und *Zieltheorien* (siehe Dweck & Leggett, 1988; Dweck, 2000; Nicholls, 1984; Nicholls, Cobb, Wood, Yackel & Patashnick, 1990).

Die Situation lässt sich zusammenfassend folgendermaßen charakterisieren. Es gibt sehr wohl zahlreiche Persönlichkeitstheorien, die sich auf die eine oder andere Weise auf Feedback und dessen differentielle Verarbeitung beziehen lassen, eine reine Feedback-Theorie hingegen existiert nicht. Um der Theorienarmut Abhilfe zu schaffen, haben Kluger und DeNisi (1996) ein explizites und somit bisher einziges umfassendes Modell zu Feedback-Phänomenen im Leistungskontext vorgeschlagen. Ihre Leistung besteht meines Erachtens vor allem in der Integration der vielfältig vorhandenen etablierten Theorien zu einem in sich schlüssigen Feedback-Modell und besticht durch den Einbezug auch nicht-intellektueller Variablen in die Betrachtung von leistungsbezogenen Feedback-Effekten. Die Feedback-Interventions-Theorie stellt somit konsequenterweise keine allein kognitive Theorie der Feedback-Verarbeitung dar, widerspricht aber gleichzeitig nicht einem informationstheoretischen Verständnis von Feedback. Theoretische Ansätze die Berücksichtigung fanden sind unter anderem: die Kontrolltheorie oder Kybernetik (z.B. Annett, 1969, Carver & Scheier, 1981; siehe auch Abschnitt 2.1.2), die

Zielsetzungstheorie (Locke & Latham, 1990), die Sozial-kognitive Lerntheorie (Bandura, 1991) und ein Ansatz aus der Forschung zur gelernten Hilflosigkeit (Mikulincer, 1994).

2.2.2.1 Grundannahmen

Im Zentrum der Feedback-Interventions-Theorie steht der *Wirkungsmechanismus* von Feedback, der als *Aufmerksamkeitslenkung* verstanden wird. Der theoretische Zugewinn gegenüber einem rein kognitiv-informationstheoretischen Feedback-Modell besteht zunächst vor allem in der Betonung der Notwendigkeit einer solchen Aufmerksamkeitslenkung. Es werden fünf Grundannahmen gemacht:

- Verhalten ist durch den *Vergleich* von Feedback mit Zielen oder Standards (auch Soll) reguliert (siehe auch Kybernetik, Abschnitt 2.1.2).
- Ziele oder Standards sind *hierarchisch* organisiert. Diese Konzeptualisierung ist weit verbreitet und findet sich zum Beispiel auch in der Handlungsregulationstheorie, beispielsweise in der Arbeitspsychologie (Hacker, 1978; Volpert, 1982).
- Die *Aufmerksamkeitsfähigkeit* ist begrenzt. Deshalb wird eine Diskrepanz zwischen Feedback und einem Zielzustand nur dann verhaltenswirksam, wenn ihr Aufmerksamkeit zuteil wird.
- Die Aufmerksamkeit ist normalerweise auf den mittleren Bereich der Zielhierarchie gerichtet, nicht auf höhere Ziele des Selbst (z.B. Berufserfolg) und nicht auf einzelne automatisierte Mikro-Komponenten der gerade ablaufenden Aktivität.
- Feedback-Interventionen verändern den *Focus* der Aufmerksamkeit.

2.2.2.2 Verhaltensregulation

Verhalten wird reguliert durch den Vergleich des Ist-Zustandes mit einem Soll-Zustand (siehe schon TOTE-Modell, Miller et al., 1960, und Abschnitt 2.1.2). Im Falle einer Ist-Soll-Diskrepanz werden vier Coping-Mechanismen angenommen:

- Erhöhung oder Verringerung der investierten Anstrengung,
- Ablehnung des angestrebten Soll-Zustandes oder Ziels (z.B. Hinwendung zu einer neuen Aufgabe),
- Veränderung des angestrebten Soll-Zustandes oder Ziels (z.B. Herabsetzen des angestrebten Leistungsniveaus),
- Ablehnung des Feedback-Inhalts (z.B. Infragestellen der Rückmeldung des Lehrers).

Für einen Ist-Soll-Vergleich ist trivialerweise die Erkenntnis eines Soll-Zustandes, d.h. des Ziels notwendig. Ziele werden deshalb als Verhaltensregulatoren verstanden. Feedback kann mit einer Norm, mit Erwartungen, vorangegangenen Leistungen, Leistungen bestimmter anderer sozialer Gruppen oder Ideal-Zuständen verglichen werden. Eine Erweiterung soll hier aber in der Feststellung gesehen werden, dass diese mitunter multiplen Ziele nicht immer präsent sein müssen, z.B. bei Vorgabe einer neuen, unbekannteren Aufgabe. Hier hat der Rezipient unter Umständen noch keine Vorstellung von einem Soll-Zustand gebildet, mit dem er das erhaltende Feedback vergleichen kann. Nach Kluger und DeNisi (1996) ist es deshalb nicht verwunderlich, dass Feedback in einem solchen Fall keine erhöhte Anstrengung und damit Leistungsverbesserung hervorruft, es sei denn es werden zusätzlich Ziele vorgegeben.

2.2.2.3 Feedback und Affekt

Eine andere interessante Erweiterung der Feedback-Interventions-Theorie gegenüber dem bisher vorherrschenden rein kognitiv-informationstheoretischen Feedback-Modell besteht in der Analyse der affektiven Wirkung von Feedback. Demnach beeinflusst Feedback sowohl die Stimmung, z.B. Freude, als auch die Aktiviertheit, welche als zwei dominante Dimensionen verstanden werden, die die gefühlsmäßige Erfahrung bestimmen. Bekannt ist, dass positives Feedback erwartungsgemäß eine positive Stimmung hervorruft, während negatives Feedback eine negative Stimmung bedingt. Hingegen legt erst die Größe der Ist-Soll-Diskrepanz die Stärke der Aktiviertheit fest, und zwar so, dass extrem positives und extrem negatives Feedback hohe Aktiviertheit bedingen. Moderates oder fehlendes Feedback führt entsprechend zu geringerer Aktiviertheit. Aus der Forschung zum Zusammenhang von Stimmung und Leistung ist bekannt, dass sowohl eine positive oder negative Stimmung als auch der Grad der Aktiviertheit Leistungen wesentlich determinieren. Beispielsweise hemmen negative Stimmungslagen kreative Leistungen. Genauso kann durch Feedback induzierte Aktiviertheit die Leistung beeinflussen. So führt eine hohe Aktivierung bei einfachen Vigilanzaufgaben zu besseren Leistungen, während kreative Leistungen beeinträchtigt, und etwa Intelligenztest-Leistungen nicht beeinflusst werden. Aus dem Yerkes-Dodson-Gesetz (1908) lässt sich allgemein ableiten, dass maximale Leistungen unter einer mittleren Aktivierung zu erwarten sind. Feedback kann also affektive Reaktionen hervorrufen, die wiederum die Leistung (mit)determinieren.

2.2.2.4 Feedback und Leistung

Entscheidend ist, dass Feedback nicht ignoriert, sondern aufgrund seines potentiell selbstwertbedrohlichen Inhalts generell Aufmerksamkeit zugeordnet wird. Dabei ist nicht die Frage gestellt, *ob* überhaupt sondern *wohin* die Aufmerksamkeit gelenkt wird. Feedback beeinflusst also den Lernprozess, weil es die Aufmerksamkeit auf eine Ist-Soll-Diskrepanz richtet. Auf welche Weise soll Feedback nun zu einer *Leistungsverbesserung* führen? Die hier von Kluger und DeNisi (1996) angebotenen Erklärungen sind unspezifisch:

- Zunächst wird davon ausgegangen, dass auf Misserfolg verweisendes Feedback zu vermehrter *Anstrengung* führt, also die *Aufgabenmotivation* erhöht. Wenn Anstrengung nicht zu Erfolg führt, werden im Allgemeinen neue Strategien generiert, um die Leistung zu verbessern.
- Feedback soll auf diese Weise zu einer tieferen kognitiven Verarbeitung führen (siehe auch Bangert-Drowns et al., 1991).
- Der Lernprozess kann auch direkt beeinflusst werden, in dem Hilfen, d.h. elaboriertes Feedback, bestimmte fehlerhafte Aufgabenkomponenten direkt ansprechen und korrigieren (siehe auch Kulhavy, 1977).

Interessanterweise werden aber einige Gründe aufgeführt, warum Feedback andererseits auch *Leistungseinbußen* hervorrufen kann, nämlich dann

- wenn Feedback keine konkreten Hinweise enthält, fehlerhafte Hypothesen abzulehnen. Der Rezipient wird eventuell in inkonsistenter Weise verschiedene Hypothesen generieren und prüfen, also mit bisher verwendeten Strategien experimentieren. Dieses Verhalten birgt zunächst die Gefahr eines Leistungsabfalls in sich, da die verwendeten Strategien eventuell unangemessen sind oder aufgrund mangelnder Automatisierung fehlerhaft umgesetzt werden und darüber hinaus kognitive Ressourcen binden.
- wenn Feedback keine zusätzliche Information enthält bzw. aufgrund von Vorwissen redundant ist (siehe auch Bangert-Drowns et al., 1991). Informationsarmes Feedback klärt den Rezipienten nicht über erfolgreiche Aufgabenstrategien auf und kann deshalb nicht zu einer geeigneten Verhaltensänderung auf Seiten des Rezipienten führen.

- wenn Feedback die Aufmerksamkeit aufwärts in der Hierarchie der Ziele, also näher zu selbstrelevanten Zielen richtet, z.B. einen guten Eindruck zu hinterlassen, statt auf untere Ebenen, d.h. auf einzelne Aufgabenkomponenten, z.B. eine Teilaufgabe zu lösen. Diese Aufmerksamkeitsverlagerung kann unter anderem dazu führen, dass der Rezipient
 - versucht, die Ist-Soll-Diskrepanz zu verringern (z.B. durch *Anstrengungsreduktion*, wie der Zuwendung zu einer anderen, selbstwertdienlicheren Aufgabe),
 - die Aufmerksamkeit auf das *Selbst* richtet (z.B. bei normativem Feedback Metakognitionen bezogen auf die selbstwertdienliche Nützlichkeit einer Aufgabe ausgebildet, beispielsweise für das Ziel, einen guten Eindruck zu hinterlassen),
 - *affektive* Reaktionen zeigt (durch kognitive Bewertung des Feedbacks bezogen auf selbstrelevante Ziele, z.B. in einer Testsituation Intelligenz zu demonstrieren, Stimmung und Aktiviertheit verändert),
 - *kognitive Ressourcen* für die Aufgabenbearbeitung nicht mehr zur Verfügung hat (durch meta-kognitive Prozesse, z.B. Worry-Gedanken, insbesondere bei nicht-automatisierten Aufgaben).

Zusammenfassung: Kluger und DeNisi (1996) gehen davon aus, dass Feedback die Aufmerksamkeit auf Ziele richtet, die die Aufgaben*motivation* beeinflussen. Zeigt Feedback einen selbstwertschädigenden Misserfolg an, z.B. Versagen in einem Intelligenztest, kann diese Ist (Feedback) – Soll (Ziel: Intelligenzdemonstration) Diskrepanz durch *Anstrengungsreduktion*, z.B. Ablehnung der Art der Aufgabe/des Tests, geschehen. Selbst wenn der Rezipient in der Lage ist, die Aufmerksamkeit wieder zurück auf die Aufgabenebene, d.h. weg vom Selbst, zu richten, kann seine Leistungsfähigkeit durch die vorangegangene Aufmerksamkeitsablenkung auf selbstrelevante Themen, durch die damit verbundene Ressourcenbeanspruchung gefolgt von affektiven Reaktionen (Stimmungsabfall, Überaktivierung) bereits beeinträchtigt sein. Eine Aktivierung von auf das Selbst bezogenen Feedback-Schleifen und dessen Folgen ist dabei durch andere, in der Persönlichkeit liegende Variablen moderiert. Hierzu gehören etwa die *Selbstwirksamkeitserwartung* und *Ängstlichkeit*.

Moderatoren von leistungsbezogenen Feedback-Effekten

Als Variablen, die die Effektivität von Feedback bestimmen, werden der *Feedback-Inhalt*, *Aufgabenmerkmale* sowie *situationale* und *personale* Merkmale angenommen. Letztere

bestimmen, wie der Rezipient mit einer Ist-Soll-Diskrepanz umgeht, nachdem seine Aufmerksamkeit durch Feedback auf diese gelenkt wurde. Auf personale Merkmale soll kurz eingegangen werden:

Selbstwertgefühl, Selbstwirksamkeit, Kontrollüberzeugungen, Ängstlichkeit oder Altruismus werden als Moderatorvariablen von leistungsbezogenen Feedback-Effekten verstanden. Persönlichkeitsvariablen bestimmen subjektive Zielstrukturen, beeinflussen inwiefern ein Aufmerksamkeitsshift tatsächlich auf selbstrelevante Ziele erfolgt und auf welche Weise eine wahrgenommene Ist-Soll-Diskrepanz aufgelöst wird. So sollten vor allem Personen mit niedrigem Selbstwertgefühl und/oder hoher Ängstlichkeit das Ziel verfolgen, negatives Feedback zu vermeiden. Auf Misserfolg verweisendes Feedback richtet bei diesen Personengruppen die Aufmerksamkeit auf das Selbst bzw. löst metakognitive Prozesse aus (auch als Worry-Kognitionen bezeichnet) und beeinträchtigt in der Folge ihre Leistung. Kluger und DeNisi (1996, 1998) schränken an dieser Stelle ein, dass in der Mehrzahl der Studien zu Feedback Persönlichkeitsmerkmale als Moderatoren nicht erfasst wurden und machen einen Mangel an entsprechenden Studien aus.

2.2.3 Empirische Evidenz für die Effektivität verschiedener Feedback-Strategien

Lassen sich nun empirische Belege für die Feedback-Interventions-Theorie finden? Kluger und DeNisi (1996) konnten innerhalb einer von ihnen vorgenommenen Metaanalyse erste Evidenz für ihr Feedback-Modell ableiten. Die Schlussfolgerungen der Autoren sind:

1. Feedback, das die Aufmerksamkeit auf Meta-Prozesse (das Selbst) richtet, vermindert positive Leistungseffekte, während Feedback, welches die Aufgaben*motivation* unterstützt oder die Aufmerksamkeit auf Lernprozesse lenkt, Leistungsverbesserungen hervorruft. Diese Befunde werden als Nachweis für die Grundannahme der Feedback-Interventions-Theorie gesehen, nämlich dass Feedback die Aufmerksamkeitsausrichtung verändert. Je eher es gelingt, diese auf möglichst niedrige Hierarchieebenen zu lenken, desto größere Leistungseffekte sind zu erwarten. Stützende Befunde sind (am Stichprobenumfang gewichtete Effektgrößen):
 - Lob ($d = .09$) und Tadel ($d = -.14$) haben keine Effekte oder stören die Leistung, da sie die Aufmerksamkeit auf das Selbst lenken.
 - Personenvermitteltes Feedback ($d = .23$) bewirkt geringere Leistungseffekte als computergestütztes Feedback ($d = .41$). Die Leistungen bei personenvermittel-

tem Feedback sind wahrscheinlich deshalb reduziert, weil die Anwesenheit einer Person meta-kognitive Prozesse fördert (Kluger & Adler, 1993).

- Häufiges Feedback ($d = .55$) führt zu Leistungsverbesserungen, da eine präzise Kenntnis der Ist-Soll-Diskrepanz unterstützt und ein intraindividueller Vergleich ermöglicht werden.
- Über die korrekte Lösung informierendes Feedback ($d = .43$) fördert Leistungen.

2. Die Autoren finden erste, aber insgesamt recht vage Hinweise auf *Aufgabenmerkmale* als Moderatoren:

- Bei wenig komplexen Aufgaben ($d = .55$) ist der Feedback-Gewinn höher als bei komplexen Aufgaben ($d = .03$).
- Feedback beeinträchtigt bei neuen und unbekanntem Aufgaben die Leistung, wenn sie über einen kurzen Zeitraum erfasst wurde. Allerdings sind die Befunde bezüglich Geübtheit und Aufgabenkomplexität hier nicht ganz eindeutig.
- Feedback ist weniger effektiv bei Regel-Finde-Aufgaben ($d = .19$; siehe die Kurzzeit-Lerntests), dagegen sehr effektiv bei Gedächtnisaufgaben ($d = .69$).
- Feedback-Effekte sind immer dann vorhanden, wenn *Ziele* vorgegeben werden ($d = .51$), besonders wenn Feedback schwer zu evaluierende Information enthält. Die Explikation von Zielen erhöht Feedback-Effekte.

Eine Ist-Soll-Diskrepanz kann in verschiedener Art und Weise aufgelöst werden. Welche Art der Diskrepanzreduktion gewählt wird, kann in der Umwelt (Aufgabenmerkmale, Feedback-Merkmale) oder in der Person (charakterliche Voraussetzungen, z.B. Selbstwirksamkeit) liegen. Entscheidend ist die Fähigkeit des (Zurück)Lenkens der Aufmerksamkeit zur Aufgabe. Untersuchungen zu relevanten Personmerkmalen als Moderatoren von Feedback-Effekten stehen weitestgehend noch aus. Einige Arbeiten unterstützen aber die Annahmen zur besonderen Rolle verschiedener selbstrelevanter Konstrukte im Kontext von Feedback und Leistung, z.B. des *Selbstkonzepts* (Kluger & Adler, 1993; Meyer & Starke, 1981/1982; Stiensmeier-Pelster, Balke & Schlangen, 1996), des *Selbstwertgefühls* (Kluger & Adler, 1993; Shrauger & Rosenberg, 1970; Stake, 1982), der *Selbstwirksamkeit* (Cervone & Wood, 1995; Johnson, Perlow & Pieper, 1993; Narciss, 2001; VandeWalle et al., 2001) oder allgemeiner der *Selbstregulation* (Chen-Idson & Higgins, 2000; Förster, Grant, Idson & Higgins, 2001).

Die überwiegende Zahl dieser Studien zu außerintellektuellen, differentiellen Effekten von Feedback verwendet allerdings *nicht-kontingentes* also „falsches“ Feedback, welches das Leistungsniveau der Versuchspersonen nicht real widerspiegelt. Ziel der Autoren ist hier, Erfolgs- oder Misserfolgserfahrungen auf Seiten der Versuchspersonen kontrolliert zu induzieren, um deren Wirkung auf das Leistungsverhalten zu analysieren. So liegen zahlreiche Befunde zu differentiellen Effekten von nicht-kontingenten Leistungsrückmeldungen aus Forschungsarbeiten zur *gelernten Hilflosigkeit* (Seligman, 1975; z.B. Brunstein, 1989; Kirchhof, 1995a, 1995b; Kirchhof & Stiensmeier-Pelster, 1999; Lauterbach, Stecker & Partl, 1988; Mikulincer, 1988; Schmalt, 1994; Winefield & Rourke, 1991), der *Stressforschung* (z.B. Jerusalem, 1991, 1992; Hock, Hardt & Krohne, 1990) sowie aus den Forschungstraditionen zum *Begabungskonzept* (Meyer, 1984; z.B. Bossong, 1982; Försterling & Schuster, 1987; Försterling & Schöler, 1984; Köller & Möller, 1995; Schwarzer & Jerusalem, 1982), zur *Zielorientierung* (Dweck & Leggett, 1988; Nicholls, 1984; z.B. Balke & Stiensmeier-Pelster, 1995; Harackiewicz & Elliott, 1993; Fodor & Carver, 2000; Spinath & Stiensmeier-Pelster, 2000; Stiensmeier-Pelster & Schlangen, 1996; Stiensmeier-Pelster et al., 1996) und *Leistungsmotivation* (Latta, 1978; Schneider, 1978) oder zur *Leistungsangst* vor (Frey, Stahlberg & Fries, 1986; Lugt-Tappeser, Schäfer, Scheiblechner & Franzen, 1994; Quast, 1987; Schwarzer & Jerusalem, 1982). Aufgrund der dort diskutierten Befunde lässt sich dennoch schlussfolgern, dass leistungsbezogenes (hier nicht-kontingentes) Feedback sehr wohl außerintellektuelle differentielle Effekte bedingen kann und nicht ausschließlich allgemein kognitiv-intellektuell verarbeitet wird. Eine genauere Analyse der Befundlage wird hier jedoch nicht vorgenommen, da sich die vorliegende Studie auf kontingentes, den Leistungsstand des Testanden tatsächlich widerspiegelndes Feedback konzentriert.

In einer Metaanalyse nach Bangert-Drowns et al. (1991) ließen sich 50% der Varianz der Feedback-Effektgrößen durch die zwei zuerst genannten der folgenden Moderatoren regressionsanalytisch vorhersagen:

- *Presearch Availability* (Kulhavy, 1977): Feedback-Effekte werden verhindert, wenn Rezipienten die korrekten Antworten direkt aus dem Feedback ableiten können bzw. schon vor ihrer Antwortgabe Feedback abrufen. Die Möglichkeit, Feedback schon vor Beantwortung der Items zu erhalten, besteht in manchen instruktionalen Programmen (z.B. im programmierten Unterricht oder heute in computergestützt vor-

gegebenen Instruktionsprogrammen, siehe Abschnitt 2.1.1), also in Programmen zur Vermittlung von deklarativem Wissen. Unter Kontrolle dieses Problems ergab sich ein etwas größerer leistungsbezogener Feedback-Effekt ($d = .39$), der auch den Befunden nach Kluger und DeNisi (1996) entspricht.

- *Feedback-Art*: Feedback, das „nur“ über die Korrektheit einer Antwort informiert (Richtig/Falsch-Feedback) ergab keine Effekte ($d = -.08$). Effektives Feedback enthielt zusätzliche Informationen, z.B. in Form von fehlerspezifischen Hilfen ($d = .53$) bis hin zur Vorgabe der Lösung ($d = .22$). Informationsarmes Feedback wird als zweite Hauptursache für negative Feedback-Effekte ausgemacht. Bei Beachtung beider Variablen Presearch Availability und Feedback-Art fand sich ein moderater Feedback-Effekt ($d = .58$).
- *Art der Instruktion*: Im Programmierten Unterricht implementiertes Feedback erbrachte die geringsten Leistungseffekte ($d = -.04$), computerbasierte Instruktionsprogramme ergaben mittlere Effekte ($d = .22$), die höchsten Effekte ergaben sich für auf *Testleistungen* bezogenes Feedback ($d = .63$), wobei hier nicht allein Intelligenztests gemeint sind.

Bangert-Drowns und Mitarbeiter (1991) fassen zusammen, dass effektives Feedback Fehler korrigiert, und zwar dann, wenn es eine vertiefende (auch elaborierte) Verarbeitung fördert (siehe auch Kluger & DeNisi, 1996), also die Reorganisation von fehlerhaftem Wissen und fehlerhaften Handlungsstrategien unterstützt. Problematisch sind Desinteresse und Langeweile, aber auch ungeeignete Erwartungen der Rezipienten. Schon Kulhavy (1977) hat festgestellt, dass insbesondere erwartungswidrigem Feedback hohe Aufmerksamkeit zugeteilt wird. Rezipienten nutzten Feedback vor allem dann, wenn sie auf eine falsche Aufgabenlösung hingewiesen wurden, obwohl sie subjektiv überzeugt waren, die Aufgabe richtig gelöst zu haben, also hohe Antwortsicherheit angaben; auch Hancock, Stock, Kulhavy, 1992; Kröner & Leutner, 2001; Kulhavy & Stock, 1989; Webb, Pridemore, Stock, Kulhavy & Henning, 1997).

Feedback in Intelligenztests

Feedback ist ein typisches Merkmal dynamischen Testens. Zahlreiche Validitätsstudien führten den Nachweis der psychometrischen Qualität dieser Verfahrensgruppe (siehe Beckmann, 2001; Meijer & Elshout, 2001). Aber auch außerhalb dieser Forschungstradition wurden Testverfahren mit Feedback-Angebot konstruiert. So fordert z.B. Kyllonen

(1991) als eines seiner Prinzipien zur Konstruktion von leistungsdiagnostischen, computerisierten Testbatterien, über die Korrektheit der Itemantwort informierendes Feedback zu integrieren: „I suspect, admittedly without any solid supporting data, that providing accuracy feedback following every item is almost always a good thing to do (in the sense of resulting in a more valid score).“ (S. 11). Dass seine Annahme zu optimistisch, auf jeden Fall zu undifferenziert war, werden die in dieser Arbeit vorgelegten empirischen Daten belegen. Darüber hinaus finden Delgado und Prieto (2003) bei Studenten höhere Fehlerzahlen bei itemspezifischem (auditivem) Richtig/Falsch-Feedback in Intelligenztests (*Primary Mental Ability Verbal Test, Rotation of Solid Figures Test*, Thurstone, 1939; Thurstone & Thurstone, 1949) als unter feedbackfreier Testbedingung. Auch bei Kontrolle des Fähigkeitsniveaus der Testanden blieb dieser negative Feedback-Effekt bestehen. Ebenso konnten Stankov und Crawford (1997) keine Effekte von itemspezifischem Richtig/Falsch-Feedback auf Intelligenztest-Leistungen (*Raven's Progressive Matrices*) ausmachen. Sie zeigten jedoch, dass unter Feedback korrekte Antworten schneller gegeben wurden als in der feedbackfreien Bedingung. Dieser Effekt auf das Zeitverhalten könnte als Effizienzsteigerung unter Feedback interpretiert werden. Stankov und Crawford (1997) schlussfolgern aber: „Feedback after providing answers to psychological test items is not effective at all“ (S.104). Auch Roos, Wise und Plake (1997) stellen die sich in den publizierten Befunden zu Feedback-Effekten widerspiegelnde Unklarheit darüber heraus, ob und wie sich itemspezifisches Feedback innerhalb computergestützter adaptiver Teststrategien auf Leistungen (und Testangst) auswirkt. Aufgrund der Mehrdeutigkeit der Befundlage bemerken sie: „... hence it is difficult to draw a general conclusion regarding the desirability of feedback in testing.“ (Roos, Wise & Plake, S.87). Eine Diskussion ihrer Befunde unter Beachtung der Erkenntnisse der Forschungstradition zu dynamischen Testverfahren wird von den Autoren leider nicht vorgenommen.

2.2.4 Implikationen für eine Leistungsdiagnostik mit Feedback

Kluger und Mitarbeiter stellen mit ihren Arbeiten die allgemein anerkannte Annahme, Feedback bringe Leistungsverbesserungen, theoretisch wie empirisch in Frage. Damit wird auch eine Prämisse angezweifelt, die bei der Konstruktion von dynamischen Testverfahren als selbstverständlich vorausgesetzt wird. Grundsätzlich gehen Vertreter des Dynamischen Testens davon aus, dass Rückmeldungen in Tests Lernmöglichkeiten darstellen und in Abhängigkeit der Lernfähigkeit des Testanden somit differentiell leis-

tungsförderlich sind. Es werden positive Feedback-Effekte unterstellt, bzw. keine negativen Effekte durch Feedback erwartet. Die Fähigkeit, von Feedback zu profitieren, wird dabei als adäquater Indikator für die *intellektuelle Leistungsfähigkeit* des Testanden interpretiert, der traditionellen Statustest-Werten vorzuziehen ist (siehe Abschnitt 3.1). Andere als intellektuelle Faktoren werden im Zusammenhang mit einem Versagen in standardisierten dynamischen Verfahren, also Testverfahren mit Feedback, nicht angenommen. Analysen bezüglich der Wirkung außerintellektueller Faktoren auf die Testleistung existieren zwar, beziehen sich aber auf den Nachweis einer kompensatorischen Wirkung von dynamischen Testprozeduren bei ungünstigen außerintellektuellen Leistungsvoraussetzungen, wie z.B. Impulsivität (Carlson & Wiedl, 1979). Die Befundlage ist hier aber widersprüchlich (siehe Abschnitt 3.2).

Welche Schlüsse lassen sich bisher aus den Ergebnissen der Forschung zu Feedback-Phänomenen ableiten, um die Frage nach geeigneten Feedback-Strukturen in Intelligenztests (Dynamisches Testen) zu beantworten? Feedback in dynamischen Verfahren sollte:

- die korrekte Lösung enthalten,
- fehlerhafte Hypothesen korrigieren (elaboriertes Feedback),
- auf Richtig/Falsch-Informationen verzichten,
- eher computergestützt als durch einen Testleiter vorgegeben werden,
- kein Lob oder Tadel enthalten (ist bei der Formulierung von Feedback zu beachten, z.B. Vermeiden von „Das hast Du gut gemacht“),
- einen intraindividuellen Vergleich zulassen (Veränderung eines Testanden anzeigen),
- keine interindividuellen Vergleiche vornehmen (keine normbezogene Information anbieten; sondern z.B. geblocktes Feedback, siehe Kyllonen, 1991, *Block Level Feedback*, z.B. graphische Darstellung des Leistungsniveaus nach jeder 5. Aufgabe),
- aufgabenbezogene Ziele vorgeben (da die Zielstrukturen des Testanden nicht klar sind).

Teilweise werden die genannten Anforderungen an Feedback-Strukturen in dynamischen Tests bereits verwirklicht. Ein Testverfahren, das nach dieser Analyse ein effektives Rückmeldesystem umsetzt, ist die Adaptive Computergestützte Intelligenz-Lerntest-Batterie, ACIL, Guthke et al., 1995). Innerhalb dieser computerbasierten Kurz-

zeit-Lerntest-Batterie wird *computergetütztes* Feedback angeboten, welches *fehlerbezogene* Hilfen (elaboriertes, fehlerkorrigierendes Feedback) bis hin zur *korrekten* Lösung beinhaltet.

Aufgrund der bisher referierten Befundlage sind relativ geringe leistungsbezogene Effekte von Feedback in Intelligenztests zu erwarten, bei Verwendung von:

- *komplexen Aufgaben*: Intelligenztest-Aufgaben sind aber im Allgemeinen komplexe Aufgaben, die vor allem von kognitiven Prozessen abhängen (sollen). Eine Steigerung der *Motivation* durch Feedback muss deshalb nicht unbedingt zur Leistungsverbesserung führen. Dieser Prozess ist innerhalb einer Intelligenzmessung aber nützlich, da kognitive und nicht motivationale Prozesse erfasst werden sollen. Gleiches gilt für:
- *Regel-Finde-Aufgaben*: Intelligenztests enthalten häufig Reasoning-Aufgaben (siehe auch Kurzzeit-Lerntests, z.B. Carlson & Wiedl, 1976; Guthke et al., 1995; Frohriep, 1978), die als g-Faktor geladene Aufgaben gelten und deshalb als valide Indikatoren für Intelligenz anerkannt sind. Die Befunde zeigen aber, dass geringere Effekte von Feedback auf die Lösungsqualität von Regel-Finde-Aufgaben im Gegensatz z.B. zu reinen Gedächtnisaufgaben zu erwarten sind.
- *Neuartigen Aufgaben*: Innerhalb eines Intelligenztests werden üblicherweise neuartige, wenig geübte Aufgaben angeboten, da bei der Erfassung von Intelligenz Vorwissen nicht interessiert und kontrolliert werden soll. Die Leistung in diesen neuartigen Aufgaben wird zudem über einen kurzen Zeitraum erfasst. Um Feedback-Effekte zu fördern, wäre jedoch eine ausgedehnte Übungsphase oder Wiederholungsmessung (siehe z.B. auch Befunde nach Meijer, 2001) notwendig, da positive Feedback-Effekte bei neuen, wenig geübten Aufgaben, aufgrund neuer zunächst potentiell fehlerhafter Hypothesenbildung nicht zu erwarten sind.

Entscheidend ist jedoch die aus der Feedback-Interventions-Theorie abzuleitende Erkenntnis, dass Leistungsrückmeldungen während einer Intelligenztestung vom Testanden potentiell als *selbstwertbedrohlich* wahrgenommen werden. Der Selbstwert bzw. das Selbstwertgefühl stellt dabei die evaluative Dimension, d.h. gefühlsmäßige Beurteilung des *Selbstkonzepts* einer Person dar. Das Selbstkonzept wird als selbstbezogene Disposition verstanden und beinhaltet das *Wissen* über die eigene Person

(Asendorpf, 1999). Dem lassen sich verschiedene bereichsspezifische Selbstkonzepte, beispielsweise das *Selbstkonzept eigener Fähigkeiten* (auch Begabungskonzept) unterordnen. Spezifische Selbstkonzepte (Fähigkeitskonzepte) sind für nachfolgendes Verhalten bedeutsam, da sie indirekt Auskunft darüber geben, was sich eine Person generell oder in einem bestimmten Bereich, z.B. bei Bearbeitung eines Intelligenztests, zutraut (häufig auch als Vertrauen in die eigenen Fähigkeiten konzipiert, vgl. Mielke, 1996).

Das Selbstkonzept einer Person stellt demnach in Leistungssituationen mit Feedback-Angebot ein zentrales Konstrukt dar, das im Rahmen einer Analyse des Zusammenhangs von Persönlichkeit und Leistung besonders berücksichtigt werden sollte. Bewährte persönlichkeits-theoretische Konzepte, die Aufklärung bei der Analyse von Leistungsverhalten unter Feedback-Bedingungen versprechen sind deshalb:

- das Konzept der motivationalen Orientierung in Lern- und Leistungssituationen (Dweck & Leggett, 1988; Dweck, 2000; Nicholls, 1984, 1990) und
- das Konzept der Testangst (Morris, Davis & Hutchings, 1981; Meijer, 1996, 2001).

Beide Theorien machen Aussagen zur Rolle des *Selbstkonzepts eigener Fähigkeiten* in Leistungssituationen, zum einen im Zusammenhang mit den Zielstrukturen einer Person, zum anderen unter Berücksichtigung der Testängstlichkeit. Die Zielstrukturen einer Person als auch das Konstrukt Ängstlichkeit werden innerhalb der Feedback-Interventions-Theorie als wesentliche Merkmale einer Person verstanden, die die Verarbeitung und Nutzung von Feedback (mit)bestimmen. Beide Ansätze sollen deshalb in der vorliegenden Studie zur Prüfung des Zusammenhangs von Persönlichkeit und Leistung unter Feedback-Bedingungen herangezogen werden und werden im Folgenden kurz umrissen.

2.2.5 Motivationale Orientierung

Im Bereich der Lern- und der Leistungsmotivationsforschung unterscheiden einige Autoren zwei Arten von *Zielen*, die beim Lernen verfolgt werden: zum einen das Ziel, Kompetenz zu erwerben und zum anderen das Ziel, anderen gegenüber hohe Kompetenz zu demonstrieren bzw. geringe Kompetenz zu verbergen. Folgende Begriffe werden zur Konzeptualisierung dieser Zielkategorien verwendet (vgl. z.B. Köller & Schiefele, 2001; Stiensmeier-Pelster & Schlangen, 1996; Spinath & Stiensmeier-Pelster, 2000; Ziegler, 2001): *intrinsische* versus *extrinsische Motivation* (Deci, 1975), *Aufgaben-* versus *Ichorientierung* (Nicholls, 1984), *Lern-* versus *Leistungszielorientierung* (Dweck &

Leggett, 1988), *Bewältigungs-* versus *Leistungsziele* (Ames & Ames, 1984), *individuelle* versus *soziale Bezugsnormorientierung* (Heckhausen, 1974; Rheinberg, 1980), oder auch *Leistungs-* versus *Machtmotiv* (Murray, 1938; zu den dennoch bestehenden Unterschieden zwischen einigen Konstrukten siehe Spinath & Stiensmeier-Pelster, 2000). Beide Zielkategorien sind sowohl als überdauernde Disposition als auch als situativ anregbarer Zustand konzeptualisiert. Im Allgemeinen wird schließlich davon ausgegangen, dass das Ziel, hohe Kompetenz zu erwerben, Motivation und somit mittelbar die Leistung fördert (z.B. das Lernziel, Dweck & Leggett, 1988). Das Ziel, Kompetenz zu demonstrieren bzw. Inkompetenz zu verbergen, beeinträchtigt dagegen Motivation und Leistung (z.B. das Leistungsziel, Dweck & Leggett, 1988). Allerdings räumen Dweck und Leggett (1988) sowie Nicholls (1984) ein, dass sich nicht eine der beiden Zielorientierungen generell negativ auf Motivation und Leistung auswirkt, vielmehr ist das *Begabungskonzept* (Selbstkonzept eigener akademischer Fähigkeiten) entscheidend für die Manifestation von Leistungsdefiziten. Erst ein Zusammentreffen von Leistungszielorientierung (das Ziel: Kompetenz zu demonstrieren) und geringem Begabungskonzept führt zu hilflosem Verhalten und infolgedessen zu Leistungsdefiziten unter Misserfolgsbedingungen (Dweck & Leggett, 1988; Elliott & Dweck, 1988). Zahlreiche Studien unterstützen die Relevanz von motivationalen Zielen zur Verhaltensklärung- und vorhersage im Leistungskontext (z.B. Dweck, 2000; für einen Überblick siehe auch Köller & Schiefele, 2001; Rawsthorne & Elliot, 1999; Spinath & Stiensmeier-Pelster, 2000; Stiensmeier-Pelster et al., 1996, Utman, 1997). Utman (1997) konnte im Rahmen einer Metaanalyse zeigen, dass Lernziele in verschiedenen Lern- und Leistungssituationen mit höheren Leistungen einhergingen als Leistungsziele. Dieser Effekt wurde auch als Lernzielvorteil bekannt. Untersuchungen zu Leistungsunterschieden nach Erfolgs- bzw. Misserfolgsfeedback in Abhängigkeit der Zielstrukturen einer Person, sind in verschiedenen Kontexten unternommen wurden, z.B. im Arbeitssetting (z.B. Brunstein, 1995; Tuckey, Brewer & Williamson, 2002; Korsgaard & Diddams, 1996), an der Universität (z.B. Fodor & Carver, 2000; VandeWalle & Cummings, 1997; VandeWalle, Cron & Slocum, 2001) und in der Schule (z.B. Broome, 1998, 2001; Butler, 1987; Elliott & Dweck, 1988; Ziegler & Schober, 1999). Die Befunde unterstützten im Wesentlichen den genannten Leistungsvorteil durch Lernziele gegenüber Leistungszielen.

Die vorliegende Arbeit fokussiert auf die Ansätze nach Dweck (Dweck & Leggett, 1988) sowie nach Nicholls (1984). Während Nicholls davon ausgeht, dass Lern- und

Leistungsziele gleichzeitig verfolgt werden können, versteht Dweck die Zielorientierung einer Person als ein bipolares, eindimensionales Konstrukt. Im Sinne einer besseren Generalisierbarkeit der Befunde werden in dieser Arbeit beide Ansätze berücksichtigt.

Lern- versus Leistungszielorientierung

Nach der Dweckschen Theorie der Leistungsmotivation (Dweck, 1986, 2000; Dweck & Leggett, 1988) verstehen Individuen mit einer *Lernzielorientierung* die Beschäftigung mit einer Aufgabe vor allem als Möglichkeit zum Erwerb neuer Fähigkeiten. Ihre Aufmerksamkeit ist auf Hinweise gerichtet, die anzeigen, wie und wo etwas gelernt werden kann oder auf Hinweise, die die Angemessenheit des eigenen Lern- und Leistungsverhaltens betreffen. *Rückmeldungen* werden als Informationen über den Stand der eigenen Fähigkeiten oder über die Angemessenheit des eigenen Lern- und Leistungsverhaltens angesehen. Als Bezugsgrößen für die Einschätzung der eigenen Fähigkeiten werden frühere Leistungen oder in der Aufgabe selbst liegende Kriterien herangezogen. Personen mit einer *Leistungszielorientierung* erleben Lernsituationen dagegen vor allem als Anlässe zur Bewertung der eigenen Fähigkeiten. Ihr Ziel besteht darin, hohe Fähigkeiten zu zeigen, um positive Bewertungen zu erhalten, oder einen Mangel an Fähigkeiten zu verbergen, um negative Bewertungen zu vermeiden. Die Aufmerksamkeit ist auf Hinweise gerichtet, die Rückschlüsse auf die Bewertung der eigenen Person und der eigenen Leistung durch andere ermöglichen. *Rückmeldungen* werden als positive oder negative Bewertungen der eigenen Fähigkeit bzw. der eigenen Person interpretiert. Bezugsgrößen für die Einschätzung der eigenen Begabung sind vor allem die Leistungen anderer.

Nach Misserfolg bzw. Misserfolg anzeigendem Feedback bedingt die Zielorientierung eines von zwei Verhaltensmustern, und zwar so genanntes hilfloses versus meisterndes Verhalten. Das Konzept der Lern- versus Leistungszielorientierung ist hier an Arbeiten zur Theorie der gelernten Hilflosigkeit angelehnt (Seligman, 1975). Hilfloses Verhalten äußert sich in der Vermeidung von potentiellen Herausforderungen, sowie in Leistungsdefiziten bei Konfrontation mit Hindernissen bei der Zielerreichung. Meisterndes Verhalten ist dagegen durch die Suche nach Herausforderung gekennzeichnet. Bei Störungen werden effektive Handlungsstrategien aufrechterhalten. Das Verhalten in Leistungssituationen mit Erfolgs-/Misserfolgspotential wird neben der Zielorientierung wesentlich durch das *Begabungskonzept* einer Person bestimmt. Entscheidend ist zudem die Art der *Ursachenerklärung* von Misserfolg und die daraus resultierende Aufrechterhaltung oder Verringerung der *Anstrengung*, je nachdem, ob das angestrebte Ziel noch

erreichbar scheint. So macht es für leistungszielorientierte Personen, die Misserfolge auf ihre eigenen mangelnden Fähigkeiten zurückführen, wenig Sinn, ihre Anstrengung zu erhöhen, sobald sie Misserfolg erleben. Im Gegensatz zu lernzielorientierten Personen interpretieren leistungszielorientierte Personen ein hohes Ausmaß an investierter Anstrengung sogar als Indikator für mangelnde Fähigkeiten. Auch nach Nicholls (Nicholls et al., 1990) sind Zielorientierungen eng mit Ursachenerklärungen für Erfolg und Misserfolg verknüpft, die hier als subjektive Erfolgstheorien bezeichnet werden. Aufgabenorientierte (nach Dweck: lernzielorientierte) Schüler glauben, schulische Erfolge seien vor allem durch Interesse, Anstrengung, Kooperation mit anderen und verstehendes Lernen zu erreichen. Ichorientierte (nach Dweck: leistungszielorientierte) Schüler glauben dagegen, dass vor allem der Wettstreit mit Mitschülern zu schulischen Erfolgen führt.

Insgesamt schützt eine Lernzielorientierung sowie ein hohes Begabungskonzept bzw. Vertrauen in die eigenen Fähigkeiten eine Person davor, hilfloses Verhalten nach Misserfolg zu zeigen. Hilfloses Verhalten kann sich in ungünstigen selbstbezogenen Kognitionen, wie z.B. die Attribution von Misserfolg auf mangelnde Fähigkeiten anstatt auf Anstrengung, und negativen Stimmungen äußern, sowie in Leistungsdefiziten bei auftretenden Hindernissen im Problemlöseprozess, z.B. nach negativem Feedback während der Bearbeitung von Intelligenztest-Aufgaben.

2.2.6 Testangst

Testangst ist als spezifische Angst in Test- oder Prüfungssituationen konzeptualisiert (Morris et al., 1981). Liebert und Morris (1967) haben ein Zwei-Komponenten-Modell der Testangst vorgeschlagen. Danach besteht die Angsterfahrung aus zwei Anteilen: einer Worry-Komponente und einer Emotionality-Komponente. Die Worry-Komponente beinhaltet die kognitiven Elemente der Angsterfahrung, vor allem negative Erwartungen und Befürchtungen bezogen auf die eigene Person, die Situation einschließlich möglicher zukünftiger Konsequenzen. Die emotionale Komponente der Testangst beinhaltet die wahrgenommenen physiologisch-affektiven Elemente der Angsterfahrung, wie Arousal und unangenehme Gefühle, wie Nervosität und Anspannung. Nach Morris et al. (1981) sind beide Komponenten der Angst konzeptionell unabhängig. Zum einen sollen sie durch unterschiedliche situationale Reize ausgelöst und aufrechterhalten werden. Die Befundlage ist hier jedoch nicht eindeutig, z.B. ließ sich experimentell die emotionale Komponente der Angst nicht unabhängig von Worry-Kognitionen vari-

ieren (für einen Überblick zu frühen Arbeiten, siehe Morris et al., 1981). Zum anderen stehen seiner Meinung nach beide Komponenten in unterschiedlichem Zusammenhang zur Leistung. Dabei ist insbesondere die Worry-Komponente konsistent negativ mit akademischen Leistungen korreliert (Morris et al., 1981). Als Wirkmechanismus wird allgemein angenommen, dass aufgabenirrelevante Kognitionen die Aufmerksamkeitsausrichtung auf die Aufgabe stören (z.B. Deffenbacher, 1978) und auf diese Weise kognitive Ressourcen beanspruchen, die ansonsten für die Lösung der Aufgabe zur Verfügung stünden. Die emotionale Komponente der Angst wird dabei eher als Folge oder Nebenwirkung des Worrying-Prozesses verstanden (vgl. auch Meijer, 2001). Hier besteht Übereinstimmung zum Feedback-Modell nach Kluger & DeNisi (1996), wonach aufgabenirrelevante Kognitionen, z.B. bezogen auf die Erreichung selbstrelevanter Ziele, ebenfalls als wichtige Ursache für ausbleibende oder negative leistungsbezogene Feedback-Effekte ausgemacht werden. Speziell zum Leistungsverhalten unter Feedback-Bedingungen berichten Thompson, Webber und Montgomery (2002) Befunde, wonach „Worrier“ (Studenten) nach nicht-kontingentem Misserfolgsfeedback schlechtere Leistungen in Diskriminations- und Anagrammaufgaben erzielten, sowie höhere Angst und kognitive Interferenzen berichteten als „Non-Worrier“. Alternativ ist jedoch auch denkbar, dass Worry-Kognitionen lediglich die Lernerfahrung widerspiegeln und weniger Ursache für Leistungsdefizite sind. Neben der Worry-Komponente sowie der Emotionality-Komponente stellt Meijer (2001) den Mangel an Selbstvertrauen in die eigenen Fähigkeiten (*Lack of Self-Confidence*, siehe auch Meijer & Elshout, 2001) als eine entscheidende Komponente der Testangst heraus. Nach Sarason und Spielberger (1975) tendieren vor allem ängstliche Personen dazu, Reize in Bewertungssituationen als Bedrohung ihres *Selbstwertes* wahrzunehmen.

Zusammengefasst stört Testangst erwartungsgemäß die Leistung, da ängstliche Personen in Testsituationen evaluative Reize als Bedrohung ihres Selbstwertes wahrnehmen und möglicherweise infolgedessen zu Worry-Kognitionen (auch sorgenvollem Grübeln) neigen, die die Aufmerksamkeit von der zu lösenden Aufgabe abziehen. Aufgrund der Worry-Kognitionen wird ein Anteil der begrenzt vorhandenen kognitiven Ressourcen beansprucht und steht für die Aufgabenbearbeitung nicht mehr zur Verfügung. Basierend auf diesen Befunden ist die Dimension Testängstlichkeit in die Untersuchung der differentiellen Wirkung von Feedback auf die Testperformanz einzubeziehen.

2.2.7 Schlussfolgerungen

Feedback-Interventionen zählen zu den am häufigsten eingesetzten Interventionsformen innerhalb der angewandten Psychologie. Dennoch ist die Wirkungsweise der vielfältig verwendeten Feedback-Strategien bisher nicht zufrieden stellend aufgeklärt. Die mit einem kognitiven Feedback-Modell nicht vereinbare, häufige Beobachtung von fehlenden oder negativen Effekten von Leistungsfeedback führte zur Forderung, auch nicht-kognitive Personmerkmale bei der Analyse von Feedback-Effekten zu berücksichtigen. Kluger und DeNisi haben ein erstes Modell vorgestellt, das einen theoretischen Rahmen für entsprechende Forschungsarbeiten bietet. Empirische Evidenz steht aber weitestgehend noch aus. Hier besteht weiterer Forschungsbedarf.

Als ein Anwendungsgebiet von Feedback-Interventionen werden innerhalb der psychologischen Diagnostik, z.B. innerhalb der Intelligenzdiagnostik (Dynamisches Testen), Feedback-Strategien genutzt, um mit Hilfe einer auf Veränderung abzielenden Teststrategie validere Aussagen über Personmerkmale, z.B. Intelligenz, zu treffen (Abschnitt 3.1). Erkenntnisse aus der Forschungstradition zu Feedback-Phänomenen sind für die Entwicklung dynamischer Tests nützlich, um Gestaltungsprinzipien für die im Testprozess implementierten Feedback-Strategien abzuleiten und theoretisch zu begründen. Zudem ist auch für Intelligenztests mit Feedback nicht auszuschließen, dass andere als intellektuelle Personmerkmale die Leistung moderieren. Die Bedrohung des Selbstwertgefühls und/oder die Beeinträchtigung der Motivation durch Leistungsfeedback im Testprozess können zu Leistungsdefiziten führen. Eine sich ausschliesslich auf intellektuelle Kapazitätsdefizite beziehende Interpretation dieser Testleistungen wäre hier natürlich unangemessen. Es sollte deshalb untersucht werden, ob die Bewältigung von Intelligenztests mit Feedback (Dynamisches Testen) durch andere als intellektuelle Merkmale bestimmt ist.

Werden die Erkenntnisse der Feedback-Interventions-Theorie auf eine Testsituation mit Feedback angewandt, lassen sich folgende präzisierete Aussagen treffen: Während einer Intelligenztestung angebotenes Feedback sollte aufgrund des potentiell selbstwertbedrohlichen Inhalts stets eine Veränderung des Aufmerksamkeitsfokus des Testanden auslösen. Ob im Testverlauf angebotene Leistungsrückmeldungen zu einer Verbesserung der Leistung führen, ist durch die *Angemessenheit* der Aufmerksamkeitsausrichtung auf Seiten des Testanden bestimmt. Da dem Testanden nur begrenzt kognitive Ressourcen zur Verfügung stehen, wird nur eine Aufmerksamkeitslenkung auf die

Aufgabe potentiell in einer Leistungssteigerung, d.h. in höheren Testwerten, resultieren. Wird die Aufmerksamkeit auf selbstrelevante Ziele gerichtet (z.B. Leistungsziele, wie Intelligenz zu demonstrieren), sind Leistungsbeeinträchtigungen zu erwarten, die nicht durch mangelnde intellektuelle Fähigkeiten erklärt werden dürfen. Eigentliche Gründe für Leistungseinbußen können stattdessen in anderen Prozessen, wie z.B. in einer Anstrengungsreduktion (Ablehnung der Art der Aufgabe, des Tests), in affektiven Reaktionen (Stimmungsabfall, Überaktivierung), einer aufgabenunspezifischen Beanspruchung kognitiver Ressourcen (Worry-Kognitionen), vermutet werden. Der so zentrale Aufmerksamkeitsshift induziert durch Feedback ist durch Personmerkmale, wie Selbstwertgefühl, Ängstlichkeit, oder Zielstrukturen (z.B. Leistungs- versus Lernzielorientierung) moderiert.

Die verschiedenen in der Literatur beschriebenen *Funktionen* von Feedback, die sich auch auf Intelligenztests mit Feedback anwenden lassen, sind:

Förderung der Leistung durch

- Aufmerksamkeitsausrichtung auf eine Ist-Soll-Diskrepanz,
- Angebot zusätzlicher Information,
- Erhöhung der Informationsverarbeitungstiefe,
- Aufmerksamkeitsausrichtung auf bisher unbeachtete Aspekte eines Problems,
- Förderung der Aufgabenmotivation, Stimmung und (angemessener) Aktivierung,
- realistische Anpassung der Selbstevaluation,
- Erhöhung der Antwortsicherheit.

Beeinträchtigung der Leistung durch

- Beanspruchung von kognitiven Ressourcen (durch Förderung meta-kognitiver Prozesse, speziell Worry-Kognitionen, durch Aufmerksamkeitsausrichtung auf selbstrelevante Themen),
- Störung der Ausführung von automatischen Skripten,
- Beeinträchtigung der Motivation, Stimmung und Aktivierung (durch selbstwert-schädigendes Feedback).

2.3 Zusammenfassung zu Kapitel 2

Forschungsarbeiten zu Feedback-Phänomenen haben eine etwa 100jährige Tradition, die ihren Anfang im frühen Behaviourismus nahm. Das zunächst lernpsychologische Verständnis von Feedback wurde Mitte der 70er Jahre durch ein informationstheoretisches

Verständnis abgelöst. Modernere Ansätze, wie etwa konstruktivistische Instruktionsmodelle, stellen die Informationsnutzung dem Informationsangebot gegenüber. Die „plausible“ und weit vertretene Annahme, Feedback als „Verstärker“ oder später als „Information“ stelle eine effektive Interventionsstrategie zur Förderung von Lernprozessen dar, ließ sich jedoch empirisch nicht stringent nachweisen. Zahlreiche Gegenbefunde führten zu einer eher pessimistischen Bewertung der Effizienz von Feedback-Strategien Ende der 90er Jahre. Mit Hilfe eines bis heute vorherrschenden kognitiven Feedback-Modells war es nicht möglich, die innerhalb der Feedback-Forschung häufig beobachteten „enttäuschend“ geringen, oder sogar „erwartungswidrig“ negativen Effekte von Leistungsfeedback zu erklären. Nicht zuletzt deshalb wurde Ende der 90er Jahre die Forderung nach einem theoretischen Modell erhoben, das neben kognitiven auch nicht-kognitive Personmerkmale bei der Erklärung und Vorhersage von Verhalten in Lern- und Leistungssituationen mit Feedback berücksichtigt. Mit diesem Ziel publizierten Kluger und DeNisi (1996) die Feedback-Interventions-Theorie, die vor allem eine erste Integration ganz unterschiedlicher theoretischer Ansätze zur Vorhersage von Verhalten in Leistungssituationen, speziell unter Feedback-Bedingungen realisiert. Dazu wurden Erkenntnisse aus verschiedenen etablierten Forschungsrichtungen (z.B. Kontrolltheorie, Zielsetzungstheorie, Theorie zur Gelernten Hilflosigkeit) herangezogen. Bisherige empirische Arbeiten untermauern die Feedback-Interventions-Theorie in ihrem Erklärungswert für Leistungsverhalten unter Feedback-Bedingungen. Mit diesen Forschungsarbeiten ist die Hoffnung verbunden, auf dem Hintergrund der Kontroverse zur Wirksamkeit von Feedback-Interventionen effektivere Feedback-Strategien zu gestalten. Die Widerlegung einer zunächst angenommenen rein informativen Wirkung von Feedback lässt sich in der Literatur allerdings schon viel eher, nämlich bereits in den 80er Jahren nachvollziehen (z.B. Mandl, et al., 1985, Stapf et al., 1986, Abschnitt 2.1.3).

Ein Anwendungsgebiet für Feedback-Interventionen stellt die psychologische Diagnostik dar. Vertreter des dynamischen Testansatzes (Abschnitt 3.1) verwenden Feedback-Strategien im Testprozess, um die *Veränderung* eines Personmerkmals zu erfassen und so validere diagnostische Aussagen zu treffen, als mit Hilfe traditioneller, auf den Status eines Personmerkmals abzielender Testverfahren (z.B. klassische Intelligenztests). Speziell im Rahmen der Intelligenzmessung wird der Feedback-Profit, den Testanden in einem Testprozess mit Feedback zeigen, als zusätzlicher valider Indikator für ihre intellektuelle Leistungsfähigkeit interpretiert. Eine Beeinflussung der Feedback-

Testleistung durch nicht-kognitive außerintellektuelle Personmerkmale, wie im Rahmen der Feedback-Forschungstradition postuliert, ist im Sinne der Validität eines Verfahrens zur Erfassung der intellektuellen Leistungsfähigkeit natürlich nicht wünschenswert. Es muss daher geprüft werden, ob die Leistung in einer Testsituation mit Feedback tatsächlich vorrangig als ein Indikator für intellektuelle Fähigkeiten (Intelligenzpotenz, Abschnitt 3.1) des Testanden interpretiert werden darf, oder ob nicht andere außerintellektuelle Merkmale beachtet werden müssen, die den Erfolg bei der Bearbeitung von Leistungstests mit Feedback bestimmen.

Im folgenden Kapitel wird deshalb überblickshaft das Testkonzept „Dynamisches Testen“ und die hier vorgenommene spezifische Implementierung von Feedback-Interventionen am Beispiel des Lerntest-Konzepts vorgestellt. Außerdem werden bisherige Befunde zur Rolle außerintellektueller Personmerkmale bei der Bearbeitung dynamischer Intelligenztests in einer Zusammenschau dargelegt.

3 Dynamisches Testen als besondere Variante einer Leistungsmessung mit Feedback

Im folgenden Kapitel soll eine Forschungsrichtung innerhalb der Psychodiagnostik vorgestellt werden, die in Abgrenzung zur traditionellen Psychodiagnostik Feedback-Interventionen nutzt, um validere diagnostische Aussagen über Personmerkmale zu treffen. Dazu wird zunächst das Testkonzept vorgestellt sowie im Anschluss die spezifische Implementierung von Feedback-Strategien erläutert.

3.1 Zum Testkonzept

Das Konzept des „Dynamischen Testens“ ist in Auseinandersetzung mit der Kritik an der traditionellen Intelligenzdiagnostik als eine Ergänzung oder auch Alternative konzipiert und inzwischen weltweit mit über 20 Verfahren realisiert worden (Quelle: www.dynamicassessment.com). Das ursprünglich aus der Intelligenz- und Eignungsdiagnostik stammende Konzept des Dynamischen Testens wurde später auch auf andere Leistungsmerkmale (Konzentration, Merkfähigkeit) und vereinzelt auch auf den Bereich der Persönlichkeitsdiagnostik übertragen (z.B. Ettrich & Guthke, 1988; siehe auch Guthke & Wiedl, 1996). Im Folgenden werden einige Begriffsbestimmungen zum Dynamischen Testen (auch *Dynamic Testing*, *Dynamic Assessment*, *Testing-the-Limits*) aufgeführt, denen auf den ersten Blick das Interesse an der Messung von Veränderung eines psychischen Merkmals gemeinsam ist, die sich aber in der jeweiligen diagnostischen Umsetzung unterscheiden.

- „Dynamic assessment is a procedure that determines whether substantive *changes* occur on examinee behavior if *feedback* is provided across an array of increasingly complex or challenging tasks“ (Swanson & Lussier, 2001, S.321, Hervorhebung durch die Autorin).
- „In a standardized dynamic test, the *responsiveness* of an examinee to systematically and objectively changing testing conditions is measured“ (Embretson, 2000, S. 507, Hervorhebung durch die Autorin, siehe auch schon Bethge, Carlson & Wiedl, 1982; Guthke & Wiedl, 1996).
- „Broadly defined dynamic assessment is naturally linked with intervention. In essence, the goal of dynamic assessment is to evaluate, to intervene and to *change*“ (Grigorenko & Sternberg, 1998, S. 76, Hervorhebung durch die Autorin).

- „Dynamic assessment refers to an approach in which *processes* of thinking, perception, learning, and problem solving are assessed through an active teaching process aimed at *modifying* the individual's cognitive functioning“ (Tzuriel, Samuels & Feuerstein, 1988, S. 144, Hervorhebung durch die Autorin).
- „Dynamic measurement procedures are supposed to uncover the zone of proximal development and to increase predictive validity in comparison to conventional, static measurement procedures“ (Meijer & Elshout, 2001, S. 93).

Historische Wurzeln sieht Guthke (1992) bereits in Arbeiten von Stern, Thorndike (siehe Abschnitt 2.1) oder auch Meumann zu Beginn des 19. Jahrhunderts, also bereits in den Anfängen der modernen Psychodiagnostik. Insbesondere aber Wygotskis (1934, 1964) erkenntnistheoretischer und diagnostischer Zugang wird von den verschiedenen Autoren konsistent als Ausgangspunkt einer dynamischen Psychodiagnostik verstanden (z.B. Elliott, 2003; Embretson, 2000; Guthke, 1992; Guthke & Wiedl, 1996; Grigorenko & Sternberg, 1998; Meijer & Elshout, 2001).

Wygotski (1934, 1964) ging davon aus, dass die kognitive Entwicklung, einschließlich der Entwicklung der Intelligenz, als *Aneignungsprozess* verstanden werden kann, den ein Individuum in Auseinandersetzung mit seiner sozialen Umwelt leistet (Kulturhistorische Schule), z.B. ein Kind in Kooperation/Kommunikation mit einem Erwachsenen oder seinen Peers. Er forderte deshalb, diesen (Wissens-)Aneignungsprozess selbst näher zu untersuchen, und weniger nur sein Resultat, wie in der traditionellen Intelligenzmessung üblich. Für die Intelligenzdiagnostik hieß das: Neben der Erfassung der „Zone der gegenwärtigen Entwicklung“, d.h. der Messung bereits ausgereifter Funktionen durch traditionelle Intelligenztests, sollte eben auch die „Zone der nächsten Entwicklung“ betrachtet werden, also der Leistungsstand eines Kindes, den es in Kooperation mit einem Erwachsenen erreicht. Von der Erfassung der „heranreifenden Funktionen“ werden, im Vergleich zur traditionellen, statischen Intelligenztestung, angemessenere Indikatoren für zukünftiges Lernen erwartet (einen Überblick zu den theoretischen Grundlagen und Entwicklungslinien des Dynamischen Testens geben Guthke und Wiedl, 1996; siehe auch z.B. Guthke, 1992; Guthke, Beckmann & Wiedl, 2003).

Zunächst weitestgehend unabhängig von einander, beginnend in den 60er und 70er Jahren, haben Autoren in den USA (z.B. Budoff, 1970, 1975; Campione, Brown und Mitarbeiter, 1985; Carlson & Wiedl, 1976) und in Europa (z.B. Feuerstein, 1972; Guthke, 1972, 1977; Hamers & Ruijsenaars, 1984; Haywood & Switsky, 1974; Klein, 1975; Tzuriel & Klein, 1985; Tzuriel, Samuels & Feuerstein, 1988; Wiedl, 1981) in zum Teil ganz verschiedener Art und Weise versucht, die Idee der Erfassung einer „Zone der nächsten Entwicklung“ testdiagnostisch umzusetzen. *Gemeinsam* ist allen Ansätzen der Versuch, die Testbedingungen im Gegensatz zu traditionellen Verfahren so zu variieren, dass neben einem Status auch eine Veränderungskomponente erfassbar wird – die *intraindividuelle Variabilität* der zu messenden Eigenschaft (siehe Guthke & Wiedl, 1996). Gegen-

wärtig spiegelt sich in der umfangreichen Literatur (etwa 330 englischsprachige Veröffentlichungen, davon 186 Artikel und 38 Dissertationen, 76 Buchkapitel sowie 30 Bücher, Quelle: PsychINFO) zu Konzept und Applikation des Dynamischen Testens das internationale Interesse und die weite Verbreitung der Testidee wider, z.B. auch in aktuellen Arbeiten aus den Niederlanden (z.B. Hessels, 2000; Resing, 2000), England (z.B. Elliott, 2003; Lauchlan & Elliott, 2001) und den USA (z.B. Embretson, 2000; Grigorenko & Sternberg, 1998; Sternberg & Grigorenko, 2002).

Divergenzen zwischen den genannten Ansätzen bestehen z.B. in der Beantwortung der Frage, was mit den Verfahren des Dynamischen Testens eigentlich gemessen werden soll, also der Frage nach dem Diagnostizierungsgegenstand. So unterscheiden Guthke und Wiedl (1996) die angezielte validere Erfassung eines Merkmals über seine im Testprozess auslösbare Veränderbarkeit (z.B. *Testing-the-Limits-Ansatz*; Ansatz der Kurzzeit-Lerntests, Abschnitt 3.1.1), von der angezielten Erfassung einer „neuen“ Eigenschaft (z.B. Lernfähigkeit, Lernpotential, *Modifiability*, siehe Diskussion des Gültigkeitsbereichs des Dynamischen Testens in Guthke et al., 2003). Aus methodischer Perspektive lassen sich die Verfahren weiter nach der Einhaltung psychometrischer Standards bei der gewünschten Erfassung intraindividuelle Variabilität einordnen. So bevorzugen manche Vertreter des Dynamischen Testens auch als bewusste Abgrenzung zur traditionellen Intelligenzmessung ein wenig standardisiertes, klinisch-kasuistisches Vorgehen (Feuersteinsche Schule, z.B. Feuerstein, 1972; Haywood & Tzuriel, 1992, siehe Abschnitt 3.2.2). Andere unterstreichen gerade die Notwendigkeit der Psychometrisierung einer Diagnostik der intraindividuellen Variabilität (z.B. Lerntest-Konzept, schon Guthke, 1972, siehe Abschnitt 3.1.1, aber auch z.B. Embretson, 2000), bis hin zu einem mehr experimentellen Vorgehen (z.B. *Testing-the-Limits*, Carlson & Wiedl, 1976, 1980; Konzept der experimentellen Diagnostik, Berg & Schaarschmidt, 1984). Zusammenfassend definieren Guthke und Wiedl (1996) eine dynamische Testdiagnostik als:

„... Sammelbegriff für testdiagnostische Ansätze, die über die gezielte Evozierung und Erfassung der intraindividuellen Variabilität im Testprozess entweder auf eine validere Erfassung des aktuellen Standes eines psychischen Merkmals und/oder seiner Veränderbarkeit abzielen.“ (S.8).

Aus der deutschen Schule, vor allem durch Guthke in Leipzig begründet, stammen die *Lerntests*, die im Folgenden kurz vorgestellt werden. Wie noch gezeigt wird,

basiert die vorliegende Arbeit auf einer Gegenüberstellung von Kurzzeit-Lerntest-Prozeduren und (feedbackfreien) Intelligenztests.

3.1.1 Das Lerntest-Konzept

Lerntests werden als eine *leistungsdiagnostische*, standardisierte Variante der dynamischen Testdiagnostik verstanden. Genauer gesagt sind Lerntests nach Guthke (1992) durch den Einbau standardisierter Lernanregungen (z.B. Feedback) in den Testprozess gekennzeichnet mit dem Ziel, neben dem Intelligenzstatus die Intelligenzpotenz bzw. die intellektuelle Lernfähigkeit zu erfassen. Mit der intellektuellen Lernfähigkeit soll nach Guthke und Wiedl (1996) allerdings keine gänzlich neue Eigenschaft erfasst werden, sondern die im Lerntest interindividuell unterschiedlich induzierbare *Fähigkeitssteigerung*, die als „eine über den Intelligenzstatus hinausgehende zusätzliche diagnostische Information (Indikator) über 'Intelligenz' im Sinne von Intelligenzpotenz bzw. intellektueller Lernfähigkeit interpretiert“ werden kann (Guthke & Wiedl, 1996, S. 81). Es wird hier schon deutlich, dass eine klare Einordnung in einen der beiden genannten Gültigkeitsbereiche, also die Erfassung der Veränderbarkeit eines Merkmals oder die Erfassung einer „neuen“ Eigenschaft, nicht vorgenommen wird. Ein Grund mag darin liegen, dass sich innerhalb des Lerntest-Konzepts verschiedene Testtypen abgrenzen lassen (siehe unten), welche jeweils einem der Gültigkeitsbereiche zugeordnet werden können.

In Anlehnung an die Unterscheidung von Kliegl und Baltes (1987), die zwischen *Baseline Performance* (Ergebnis rein statusorientierter Intelligenztests), *Baseline Reserve Capacity* (die zum Zeitpunkt unter leistungsoptimierender Testdarbietung noch entdeckbare Intelligenzpotenz) und *Developmental Reserve Capacity* (nach längerer Trainingsphase noch zusätzlich erzeugbare neue Kapazitäten) unterscheiden, betonen Beckmann und Guthke (1999), dass in Langzeit-Lerntests oder nach länger dauernden kognitiven Trainingsprogrammen die Feststellung der Intelligenzpotenz im Sinne der *Developmental Reserve Capacity* möglich ist, dagegen in Kurzzeit-Lerntests die Intelligenzpotenz im Sinne der *Baseline Reserve Capacity* erfasst wird, die aber ebenso zu validen Prognosen zukünftigen Lernens führen soll.

Nach Guthke (1992, auch Guthke & Wiedl, 1996) lassen sich nach der Art und der zeitlichen Ausdehnung der eingebauten standardisierten Lernanregungen *Langzeit-Lerntests* und *Kurzzeit-Lerntests* unterscheiden. Im Gegensatz zum Langzeit-Lerntest

mit dem Aufbau „Prätest/Trainings- (Pädagogisierungs-)phase/Posttest“ finden Kurzzeit-Lerntests in nur einer Testsitzung statt, in welcher anstatt eines Trainingsprogramms zwischen Prätest und Posttest standardisiert *Rückmeldungen* und Denkhilfen bzw. elaborierte Rückmeldungen präsentiert werden. Weiterhin werden Kurzzeit-Lerntests vom Typ 1, die im Testprozess „nur“ Informationen über die Richtigkeit der Itemantwort (Richtig/Falsch-Feedback) anbieten, von Kurzzeit-Lerntests vom Typ 2 unterscheiden, die zusätzliche fehlerorientierte Hilfen, d.h. elaboriertes Feedback, vorgegeben. Es existieren hier aber auch „Mischformen“ nämlich Kurzzeit-Lerntests, die beide Arten an Feedback beinhalten (z.B. die ACIL, Guthke et al., 1995).

Die ACIL (vgl. Beckmann & Guthke, 1999; Beckmann, 2001) erfasst als „Kernfaktor der Intelligenz“ das schlussfolgernde Denken (Reasoning nach Thurstone) in drei Materialbereichen, dem numerischen Bereich mit Hilfe von Zahlenfolgen, dem figuralen mit Hilfe von Figurenfolgen und dem verbalen Bereich mit Hilfe von Analogien. Entsprechend dem Konzept des Dynamischen Testens wird nicht nur die momentane Kompetenz sondern auch die Potenz bezüglich der Fähigkeit „schlussfolgerndes Denken“ erfasst. Die ACIL zählt zu den *Diagnostischen Programmen*. Diagnostische Programme wurden durch Guthke und seine Mitarbeiter in den 80er Jahren als eine spezielle Variante von Kurzzeit-Lerntests entwickelt, um einige Kritikpunkte an den bis dahin veröffentlichten Lerntests zu überwinden. Nach Beckmann und Guthke (1999) werden mit den Diagnostischen Programmen Tests im Sinne eines programmierten Lehrprogramms eingesetzt. Folgende Kennzeichen eines Diagnostischen Programms werden postuliert (Beckmann & Guthke, 1999; siehe auch Beckmann, 2001): Sicherung der Kontenvalidität des Itempools bzw. Charakterisierung der objektiven kognitionspsychologisch analysierten Anforderungsstruktur des Itempools; sequentieller, möglichst hierarchischer Aufbau der Testanforderungen aufgrund einer Anforderungsstrukturanalyse; Einbau von systematischen Rückinformationen (Feedbacks) bzw. Denkhilfen in den Testprozess; Bearbeitung von nachfolgenden schwierigeren Aufgaben erst nach Lösung (selbständig oder mit Hilfestellung) aller vorhergehenden leichteren Aufgaben; Individualisierung und Ökonomisierung der Testung durch Verknüpfung des Lerntest-Konzepts mit dem Prinzip des adaptiven Testens; Verlaufsanalysen im Sinne einer Prozessdiagnostik, sowie qualitative Fehleranalysen.

3.1.2 Feedback in dynamischen Tests

Die Entwicklung der ersten dynamischen Tests fand in den 60er und 70er Jahren zeitgleich mit der Entwicklung des programmierten Unterrichts (Skinner, siehe Abschnitt 2.1.1), der Lehrmaschinen und etwas später der *Intelligent Tutoring Systems* (ITS) statt. Die damalige Vorrangstellung einer behaviouristischen Lernpsychologie, also eines Verständnisses von *Lernen durch Konsequenzen* sowie ab den 70er Jahren des *Regelungsprinzips*, also der Vorstellung von einer Verhaltensregulation durch Rückmeldungsschleifen, lassen sich sicherlich auch als wesentliche Kernideen einer dynamischen Diagnostik bewerten. Diese Aspekte werden von Vertretern des Dynamischen Testens jedoch kaum thematisiert (siehe aber eine Nennung bei Guthke, 1992; sowie Guthke & Wiedl, 1996). Trotz der augenscheinlich gemeinsamen Wurzeln und teilweise vergleichbaren Forschungsfragen fand meines Wissens kaum ein Austausch, im Sinne von publizierter gegenseitiger Kenntnisnahme der Arbeiten zwischen beiden Forschungslinien statt, z.B.

hinsichtlich der Erkenntnisse zur Gestaltung von effektiven Feedback-Prozeduren oder Moderatoren von Feedback-Effekten. In der Tat sind die Gestaltungsprinzipien, nämlich die Nutzung von Feedback-Prozeduren mit der damit verbundenen Hoffnung auf eine Förderung von Lernprozessen zum einen in Lerntest-Prozeduren zur Vermittlung von prozeduralem Wissen als auch zum anderen in pädagogisch-psychologischen Lehrprogrammen zur Vermittlung von vorrangig deklarativem aber auch prozeduralem Wissen, vergleichbar.

Die Art und Weise der Darbietung sowie der Inhalt des in dynamischen Tests implementierten Feedbacks variiert je nach diagnostischem Zugang der verschiedenen Arbeitskreise (vgl. auch Kapitel 2 zu Feedback in Lehrprogrammen). Grundsätzlich handelt es sich aber um „echtes“, d.h. kontingentes Feedback, also die erreichten Leistungen des Testanden tatsächlich widerspiegelndes, *verbales* Feedback, welches

- entweder nach jedem Lösungsversuch, oder nach einer Sequenz von Lösungsversuchen,
- mehr oder weniger *standardisiert* (hohe Standardisierung, z.B. das Lerntest-Konzept, versus hohe Individualisierung, z.B. im Rahmen des *Mediated Learning*-Ansatzes der Feuerstein Schule) sowie
- mehr oder weniger *elaboriert* (von Information über die Korrektheit des Lösungsversuchs zu spezifischen Hilfen bis hin zur Präsentation der richtigen Lösung),
- anfangs ausschließlich durch den Testleiter, heute zunehmend computerisiert dargeboten wird (Guthke et al., 1995; Meijer, 2001; siehe aber auch Tzuriel, 2002) und
- als im Testprozess selbst oder innerhalb eines Trainings zwischen zwei Testphasen eingebautes, zusätzliches *Informationsangebot* im Sinne einer Schaffung von Lernmöglichkeiten *während* der Merkmalerfassung (meist Intelligenz, aber auch andere Merkmale, wie z.B. Konzentration, Eignung) und somit als Optimierung der Testbedingungen angedacht ist.

Die Anzahl der benötigten Feedbacks (meist Hilfen) bis zur korrekten Lösung wird in den meisten Verfahren auch als Indikator für das zu messende Konstrukt, z.B. Intelligenzpotential, genutzt oder geht zumindest in seine Berechnung ein (z.B. Beckmann & Guthke, 1999; Guthke et al., 1995; Meijer, 2001).

Performanzbezogene Feedback-Effekte beim Dynamischen Testen

Experimentell angelegte Untersuchungen zur Wirksamkeit der vorrangig hinsichtlich ihrer Elaboriertheit variierenden Feedback-Arten beim Dynamischen Testen, die oft gemeinsam innerhalb der Testprozedur sozusagen als Paket angeboten werden, sind von nur wenigen Autoren vorgelegt worden. So konnten Carlson und Wiedl (1980) zeigen, dass impulsive Grundschulkinder, Kinder mit Lernstörungen, aber auch normalintelligente Zweit- und Viertklässler unter *elaborierten* Rückmeldungen höhere Leistungen erzielten (*Coloured Progressive Matrices*, CPM, Raven, 1976) als unter Rückmeldemodalitäten, die ausschließlich über die Korrektheit der Lösung informierten.

In einer Metaanalyse ($N = 5104$) von Swanson und Lussier (2001) erreichten Testanden unter dynamischen Testbedingungen im Allgemeinen höhere Testleistungen als unter Standardtest-Bedingungen (ungewichtetes $d = 1.12$). Die Autoren stellten jedoch fest, dass das Ausmaß der Leistungseffekte in Abhängigkeit von der Art der Pädagogisierung bedeutsam variierte. Teststrategien, die „Stützung“ (*Scaffolding*, z.B. Ergänzungsfragen, Hilfen; $d = .77$) und „Strategietraining“ (Vermittlung von Strategien und Feedback, $d = .66$) enthielten, erzielten moderate aber immer noch weitaus höhere Leistungs-Effekte als zwischen zwei Tests geschaltete Trainingsprogramme (*Coaching*, meist *Test-Training-Test-Design*; $d = -.08$). Sie fanden außerdem größere Effekte einer dynamischen Teststrategie auf das Leistungsverhalten für den Altersbereich bis 13 Jahre, sowie innerhalb der Gruppe der Underachiever ($d = 1.01$). Bei regressionsanalytischen Berechnungen stellten sich im gewählten Modell als beste Prädiktoren³ der durchschnittlichen Effektgröße von dynamischen Testprozeduren (gewichtet, zwischen $d = .80$ und $d = .90$) die verwendete Instruktion (die so genannte Pädagogisierung) und die Aufgabendomäne heraus. Hier konnte gezeigt werden, dass Studien, die Strategietraining, Feedback⁴ und *Modeling* verwendeten, die höchsten Leistungs-Effekte erzielten ($d = .65$), gefolgt von Studien mit *Testing-the-Limits-Design*, also z.B. *Scaffolding* ($d = .48$). Die geringsten Effekte ergaben auch hier Bedingungen, die Training (*Coaching* oder *Mediated Learning*, $d = .21$) beinhalteten. Außerdem erwiesen sich visuell-räumliche Aufgaben als geeigneter für eine dynamische Teststrategie als verbale Aufgaben. Der Effekt der Leis-

³ Andere Prädiktoren waren die Altersgruppe, die Leistungsgruppe, die Anzahl der Testsitzungen, das Versuchsdesign und die Stichprobengröße.

⁴ Swanson und Lussier (2001) geben nicht genauer an, um welche Feedback-Art es sich dabei handelt. Es ist aber zu vermuten, dass das hier betrachtete Feedback über einfaches Richtig/Falsch-Feedback hinausging.

tungsgruppe (z.B. höhere Leistungseffekte für Underachiever) verschwand im Übrigen bei Kontrolle anderer Variablen wie Alter, Design, Instruktion. Die Autoren interpretierten diesen Befund als Hinweis darauf, dass bei Verwendung traditioneller Verfahren immer wieder berichtete Leistungsdefizite bestimmter Untergruppen, z.B. hörgeschädigter Kinder, innerhalb von dynamischen Testprozeduren nivelliert werden können (eine weitere Zusammenfassung bisheriger Befunde zum Dynamischen Testen geben auch Guthke & Wiedl, 1996). Als ein Beispiel für die unzureichende gegenseitige Zurechnung lässt sich auch anführen, dass Swansons Metaanalyse keine Arbeiten von Guthke und Mitarbeitern berücksichtigt.

3.1.3 Schlussfolgerungen

In der internationalen Diskussion zur Beantwortung der Frage, *was* dynamische Teststrategien eigentlich erfassen (Gültigkeitsbereich), spiegelt sich wohl auch die Unklarheit darüber wider, *was* tatsächlich Feedback-Interventionen im Testprozess auf kognitiver Ebene auslösen (siehe aber Modell nach Carlson & Wiedl, 1980, 1992). Sind beobachtete Leistungsverbesserungen im Gegensatz zur traditionellen Intelligenzmessung auf die Ausprägung einer „neuen“ Eigenschaft (z.B. Lernfähigkeit, auch kognitive Modifiabilität) zurückzuführen, die es dem Testanden erlaubt, Feedback kognitiv zu verarbeiten und seine Leistung zu verbessern? Oder wird die Eigenschaft Intelligenz einfach „nur“ valider gemessen, handelt es sich also um ein messmethodisches Problem (z.B. durch Kompensation von ungünstigen Leistungsvoraussetzungen, wie Testängstlichkeit, siehe z.B. Meijer & Elshout, 2001 sowie Abschnitt 3.2.1)?

Die Frage nach den zugrunde liegenden Prozessen wird auch aufgrund ausbleibender leistungsbezogener Feedback-Effekte ebenso in der Forschungstradition zu Feedback-Phänomenen gestellt und ist auch hier noch nicht ausreichend beantwortet (siehe Kluger & DeNisi, 1996, Abschnitt 2.2.2). Erste Analysen zu Moderatoren von Feedback-Effekten zeigen, dass neben intellektuellen auch außerintellektuelle Variablen, z.B. die Motivation, die Verarbeitung von Feedback bestimmen (siehe Abschnitt 2.2.3). Hier ergeben sich viel versprechende Ansätze zur Aufklärung von unerwarteten Feedback-Effekten, die im Rahmen der Feedback-Forschung beobachtet wurden. Insgesamt sind die innerhalb der Forschungstradition zum Dynamischen Testen berichteten Befunde zur Effektivität von Feedback-Interventionen im Testprozess jedoch optimistischer zu werten als das in Kapitel 2 zusammengefasste Befundbild der traditionellen Feedback-Forschung. Zum Beispiel werden in Wirksamkeitsstudien zu dynamischen

Tests keine negativen performanzbezogenen Feedback-Effekte sowie insgesamt eine geringere Bandbreite der Feedback-Effekte berichtet.

3.2 Differentielle Effekte beim Dynamischen Testen

Ziel dieses Abschnitts ist es, einen Überblick über die bisherige Befundlage zur Rolle außerintellektueller Personmerkmale beim Einsatz von Testprozeduren mit implementiertem Feedback, also dynamischen Testverfahren, zu geben. Dabei wird ein Kompensationsmodell (standardisierte Verfahren) von einem Modifikationsmodell (nicht-standardisierte Verfahren) unterschieden.

3.2.1 Das Kompensationsmodell

In verschiedenen Untersuchungen konnte nachgewiesen werden, dass bestimmte Personengruppen (vor allem Unterprivilegierte und Leistungsschwache) von Lerntest-Prozeduren profitieren (für einen Überblick siehe Guthke & Wiedl, 1996). Im Folgenden sollen Befunde aus Studien zur Rolle einzelner Persönlichkeitsmerkmale bei der Bearbeitung von *Kurzzeit*-Lerntests zusammengefasst werden. Untersuchungen zum Einfluss außerintellektueller Faktoren wie Motivation, Ängstlichkeit, Kreativität, Stresssensibilität und Frustrationstoleranz auf die Leistung in *Langzeit*-Lerntests werden von Guthke und Mitarbeitern (Guthke, 1972; Guthke & Lehwald, 1984) referiert, sollen hier aber aufgrund des spezifischen Aufbaus der *Langzeit*-Lerntests nicht näher dargestellt werden. Im Paradigma der *Langzeit*-Lerntests werden im Gegensatz zu einer *Kurzzeit*-Lerntest-Prozedur weiterhin *feedbackfreie* Statustests appliziert. Leistungsrückmeldungen erfolgen sozusagen versetzt, nämlich erst innerhalb eines Trainingsprogramms nach der ersten Testsitzung. Da die vorliegende Arbeit aber auf Tests mit *implementierten*, itemspezifischen Feedback-Strategien fokussiert, konzentriert sich die Darstellung auf die Befundlage zu potentiellen außerintellektuellen Moderatoren der Leistung in *Kurzzeit*-Lerntest-Prozeduren.

3.2.1.1 Kognitive Impulsivität

Wiedl und Bethge (1981) stellen fest, dass Untersuchungstechniken, die an den Ansatz des Dynamischen Testens angelehnt sind, wie z.B. *Testing-the-Limits*-Strategien, eine validere Diagnose der intellektuellen Leistungsfähigkeiten kognitiv impulsiver Kinder ermöglichen als traditionelle Intelligenztests (auch Spillner & Spillner, 1979 bei Einsatz des Raven-Kurzzeit-Lerntests nach Frohriep, 1978). Unter Verwendung eines von Wiedl

(1980) eingeführten Leistungsmaßes⁵ konnten sie nachweisen, dass Problemverbalisation als auch elaboriertes Feedback im Testprozess zu einer Nivellierung der Leistungsunterschiede im Farbigen Matrizentest (*Coloured Progressive Matrices*, CPM, Raven, 1956) zwischen kognitiv impulsiven und reflexiven Kindern (*Matching Familiar Figures Test*, MFF, Kagan, 1966) führen und zwar aufgrund einer Angleichung der Leistungen der impulsiven Kinder an die der reflexiven Kinder (siehe auch Carlson & Wiedl, 1980; zur Verwendung von nicht-standardisierten dynamischen Testverfahren siehe auch Tzuriel, 2001). In einer Replikationsstudie zeigen Wiedl und Bethge (1981), dass sich kompensatorische Effekte weniger deutlich in den Testleistungen widerspiegeln, dagegen vielmehr in den Variablen des *Blickverhaltens* als Indikatoren des Lösungsprozesses. Durch Veränderung des diagnostischen Vorgehens unter Einbezug von Rückmeldungen und Problemverbalisation ließen sich das Blickverhalten kognitiv impulsiver Kinder und damit der Problemlöseprozess günstig beeinflussen. So beobachteten sie bei impulsiven Kindern unter dynamischen Testbedingungen höhere Fixationszeiten und -häufigkeiten bei der Betrachtung der Testaufgaben als unter Standard-Testbedingungen (vgl. auch Wiedl & Bethge, 1983).

3.2.1.2 Problemlöseverhalten

Nach Bethge, Carlson und Wiedl (1982) fördert elaboriertes Feedback auch bei „unauffälligen“ (hier nicht-impulsiven) Drittklässlern das visuelle Suchverhalten im Testprozess. Das zeigte sich in höheren Fixationszeiten und -häufigkeiten bei der Betrachtung der Testaufgaben sowie selteneren Auslassungen der Fixation von Antwortalternativen (Digital-Analog-System nach Zack) und ging mit signifikant höheren Testleistungen einher verglichen mit den Leistungen bei einer traditionellen Testvorgabe der CPM (*Coloured Progressive Matrices*, Raven, 1956). Eine Leistungsverbesserung bei Angebot von elaboriertem Feedback sowie unter der Bedingung Problemverbalisation wird mit der Förderung planvoller und systematischer Problemlösestrategien erklärt, also durch eine Modifikation des Problemlöseprozesses im Testprozess. So konnten Cormier,

⁵ Wiedl (1980) führt zwei Leistungsmaße für die Auswertung der Leistungen im Farbigen Matrizentest (CPM nach Raven, 1956) ein: die *aufgabenspezifische Fehlerlatenz*, d.h. die Anzahl richtig gelöster Items bis zum Auftreten des ersten Fehlers innerhalb der gegebenen Aufgabengruppierung sowie die *aufgabenspezifische Lösungsmenge*, d.h. die Anzahl richtig gelöster Items pro Aufgabengruppierung. Carlson und Wiedl (1979) berichten aber, dass bei (unauffälligen) Viert- und Sechstklässlern, zumindest in der Bedingung Problemverbalisation kompensatorische Effekte auch bezüglich des *Gesamt-Testwertes* der CPM (Anzahl korrekter Lösungen) auszumachen waren. Die Befundlage ist aber nicht ganz eindeutig.

Carlson und Das (1990) auch zeigen, dass Leistungsdefizite schlechter Planer (zehnjährige Kinder, *Trial-Making* sowie *Visual Search Planning Tests* nach Naglieri & Das, 1987) durch Problemverbalisation im Testprozess, als eine Form so genannter optimierender Testbedingungen, kompensiert werden konnten. Es wird vermutet, dass die frontallappenbestimmte Kontrolle über das Problemlöseverhalten vor allem durch die Verbalisation der Testaufgaben günstig beeinflusst werden kann.

3.2.1.3 Neurotizismus

Carlson und Wiedl (1976) weisen nach, dass Grundschulkinder mit hohen Neurotizismuswerten von der Durchführung einer Kurzzeit-Lerntest-Variante des Farbigen Matrizentests (vgl. Wiedl, 1980) profitieren und in ihren Leistungen im Gegensatz zur traditionellen Intelligenztestung nicht beeinträchtigt sind. Nach Carlson und Wiedl (1979) ließ sich Neurotizismus, der mit der Hamburger Neurotizismus- und Extraversionskala für Kinder und Jugendliche (HANES, Bugge & Baumgärtel, 1972) gemessen wurde, insbesondere unter einer Testbedingung kompensieren, die elaboriertes Feedback sowie Problemverbalisation beinhaltet. Dagegen war eine alleinige Problemverbalisation nicht ausreichend, um Kompensationseffekte hervorzurufen.

3.2.1.4 Leistungsmotivation

Burkert (1981) fand keine höheren Zusammenhänge von Ausdauer und motivationalen Eigenschaften von Schulanfängern und deren Leistungen im Kurzzeit-Lerntest im Gegensatz zum Statustest. Nagler (1995) zeigt, dass das Anstrengungsvermeidungsverhalten Leistungen in computergestützten Kurzzeit-Lerntests (ACIL, Abschnitt 3.1.1) mitbedingt.

3.2.1.5 Stresssensibilität

Lussier (2002) konnte bei ihrer Untersuchung des Zusammenhangs von Alltagsstress, Arbeitsgedächtnis und Reasoning innerhalb einer studentischen Stichprobe zwar nachweisen, dass bei Verwendung einer dynamischen Teststrategie zur Erfassung des Arbeitsgedächtnisses (Swanson *Cognitive Processing Tests*, S-CPT, Swanson, 1995) der Einfluss von berichtetem Alltagsstress bei der Vorhersage von Reasoning (*Standard Progressive Matrices*, Raven, 1983) reduziert wurde. Der dynamische Arbeitsgedächtnistest war aber diesbezüglich einem statischen Arbeitsgedächtnistest nicht überlegen.

3.2.1.6 Leistungsangst

Mit der Untersuchung einer kompensatorischen Wirkung dynamischer Testverfahren bezüglich des Merkmals *Testangst* hat sich vor allem Meijer (1993, 1996, 2001) beschäftigt. Schüler der dritten Klasse der universitätsvorbereitenden Sekundarstufe zwei in den Niederlanden, die hoch leistungsmotiviert waren und gleichzeitig hohe Leistungsangst zeigten, erreichten die niedrigsten Testwerte in einem traditionellen Intelligenztest, d.h. in einem Statustest (Prüfsystem für Schul- und Bildungsberatung, PSB, Horn, 1969), unterschieden sich dagegen im dynamischen Testverfahren nicht (*Transfer Test for Mathematics*, entspricht dem Kurzzeit-Lerntest-Ansatz nach Guthke, 1992). Die Ursache der beobachteten Leistungsdefizite hochängstlicher und gleichzeitig hoch motivierter Personen wird in einer Komponente der Leistungsangst, der *Worry*-Komponente (Morris et al., 1981), vermutet. Durch grübelnde Gedanken (Worry-Kognitionen) wird die Aufmerksamkeit und Konzentration auf die Aufgabe gestört. Des Weiteren ließ sich interessanterweise zeigen, dass die prädiktive Validität dynamischer Testverfahren mit der Persönlichkeitseigenschaft „Ängstlichkeit“ der Testanden variierte. So genügte die Kenntnis der Statustest-Werte wenig ängstlicher Schüler (*Mathematics Attitude Test*, Gaudry & Spielberger, 1971) bereits, um zukünftige Leistungen hinreichend zu präzisieren. Der Einbezug des Lerntest-Wertes konnte keine zusätzliche Varianz aufklären. Dagegen ließen sich zukünftige Leistungen *leistungsängstlicher* Personen valider mit Hilfe von Lerntest-Werten vorhersagen. Hier konnten zum Statustest zusätzlich bis zu 10% der Kriteriumsvarianz aufgeklärt werden.

Auf der Ebene von latenten Konstrukten belegte Meijer (2001), dass sowohl bei statischen als auch bei dynamischen Testleistungen die Testangst eine wesentliche Rolle spielt. Testergebnisse aus dynamischen Testverfahren sind demnach nicht unbedingt als reine Indikatoren der jeweils zu messenden Eigenschaft anzusehen. Allerdings ergaben sich hinsichtlich der Verfälschbarkeit der Testwerte durch den Faktor „Testangst“ für das eingesetzte dynamische Testverfahren bei wiederholter Testanwendung günstigere Ergebnisse im Vergleich zum Statustest-Pendant. Meijer und Elshout (2001) belegen, dass Leistungsunterschiede in dynamischen und statischen Testverfahren nicht ausschließlich mit der unterschiedlich starken Konfundierung der Testwerte durch den Faktor „Testangst“ erklärbar sind, sondern dass dynamische Verfahren einen „ängstlichkeitsunabhängigen“ Beitrag zur Prädiktion leisten. Darin äußert sich unter anderem die inkrementelle Validität dynamischer Testprozeduren im Vergleich zu tra-

ditionellen, statischen Tests. Im Sinne der Konstruktvalidierung wird dies als ein Hinweis für die Existenz des Konstrukts der Zone der nächsten Entwicklung (Wygotski, 1964, 1934) interpretiert (siehe auch Beckmann, 2001).

Embretson (2000) untersuchte im Rahmen eines dynamischen Testansatzes die Veränderung der Reasoning-Leistung (*Abstract Reasoning Test*, ART, Embretson, 1995⁶) Erwachsener unter Stressbedingungen, wobei sie Stress als Zeitdruck operationalisierte. Nach Embretson (2000) sind dynamische Tests durch im Testprozess eingebaute leistungsförderliche Bedingungen, z.B. Rückmeldungen, aber auch durch leistungshinderliche Bedingungen, z.B. Zeitdruck oder Störgeräusche, gekennzeichnet. Dabei soll die interindividuelle Variabilität in der Reaktionsfähigkeit des Testanden auf die sich systematisch ändernden, fördernden oder störenden Testbedingungen erfasst werden (*Modifiability*). Ein hohes Ausmaß an *Worry*-Gedanken (hier als eine Form der selektiven Aufmerksamkeit, so genannte aufdringliche Gedanken) war allerdings generell leistungshinderlich, unabhängig von der Testprozedur, was sich in niedrigeren Testscores unter stressinduzierendem Zeitdruck als dynamisches Performanzmaß als auch in statischen Performanzmaßen niederschlug. Zeitdruck reduzierte zudem die Bedeutung der *Worry*-Kognitionen am Zustandekommen der Leistungsvarianz. In einer weiteren, vergleichbaren Studie zeigte Embretson, dass der Faktor *Gewissenhaftigkeit* (*Conscientiousness*, *Big Five*) erwartungswidrig negativ gewichtet die Leistung im dynamischen Testverfahren (Reasoning unter Stress) vorhersagte.

Hinweise bezüglich der Wirkung von Kurzzeit-Lerntest-Prozeduren auf die *Leistungsangst* und die *Bewertung der Testsituation* ergeben Befunde nach Wiedl, Bethge und Bethge (1982). Danach bewirken neben der Problemverbalisierung elaborierte Rückmeldungen sowohl eine Verringerung der Leistungsangst als auch eine positivere Situationsbewertung (siehe auch Bethge, Carlson & Wiedl, 1982). Die Reduktion der Leistungsangst wird mit einer stärkeren Strukturierung der Testsituation durch die Rückmeldungen und Problemverbalisationen erklärt. Als Konsequenz mag sich die dadurch bessere Abschätzung der Erfolgs- bzw. Misserfolgchancen als förderlich erwiesen haben. Weiter wird vermutet, dass eine stärker testandenzentrierte Aktivität des Testlei-

⁶ Die hier verwendeten Matrixaufgaben sind angelehnt an die sich auf kognitive Prozesse (Arbeitsgedächtniskapazität, Abstraktionskapazität) beim Lösen von progressiven Matrizen (z.B. Ravens Progressive Matrizen) beziehende Theorie nach Carpenter, Just und Shell (1990). Der Test (ART, Embretson, 1995) wurde computergestützt vorgegeben. Da nur eine Testsitzung stattfand, kann dieser Test als eine Form der Kurzzeit-Lerntests aufgefasst werden.

ters unter optimierenden Testbedingungen, wie sie in Lerntest-Verfahren angestrebt werden (vgl. Wiedl et al., 1982), dazu führt, dass die Testsituation als angenehmer und hilfreicher wahrgenommen wird und somit angstreduzierend wirkt.

3.2.1.7 Schlussfolgerungen

Zusammenfassend machen die Ausführungen deutlich, dass die Befundlage zur Wirkung von außerintellektuellen Merkmalen in standardisierten dynamischen Verfahren, in welchen unter anderem Feedback angeboten wird, wenig umfangreich und keineswegs eindeutig ist.

Stichprobe: Die publizierten Untersuchungen variieren hinsichtlich der Zielgruppe der eingesetzten Verfahren. Die Mehrzahl der Untersuchungen bezieht sich auf Verfahren für den Kinderbereich, wenige Befunde liegen dabei für den Jugend- und Erwachsenenbereich vor (siehe Meijer, 2001, Embretson, 2000).

Lerntest-Typ: Die Untersuchungen unterscheiden sich hinsichtlich der Test-Prozedur (Langzeit- versus Kurzzeit-Lerntest). Der Prätestwert im Langzeit-Lerntest stellt im Prinzip einen klassischen Statuswert dar, im Posttestwert soll hingegen das intellektuelle Veränderungspotential ausgelöst durch das nach der ersten Testung erfolgte Trainingsprogramm erfasst werden. Langzeit-Lerntests unterscheiden sich somit in der singulären Vorgabe des Prä- bzw. Posttests nicht wesentlich von einem traditionellen Intelligenztest. Dagegen erfolgt im Paradigma des Kurzzeit-Lerntests eine einmalige Testung mit im Testprozess integriertem Feedback-System. Es ist also zu erwarten, dass diese beiden Testsituationen vom Testanden unterschiedlich erlebt und bewertet werden, was zum Teil die heterogene Befundlage zur Wirkung außerintellektueller Faktoren auf die Testleistung erklären könnte (zum Langzeit-Lerntest-Paradigma siehe Guthke, 1972; Guthke & Lehwald, 1984; Günther & Günther, 1981, 1982). Nach Guthke und Lehwald (1984) lassen sich z.B. keine entsprechenden kompensatorischen Effekte eines impulsiven Arbeitsstils bei der Durchführung von *Langzeit-Lerntests* (im Gegensatz zu *Kurzzeit-Lerntests*) nachweisen. Sie fanden sogar Defizite kognitiv impulsiver Kinder verglichen mit kognitiv reflexiven Kindern in den *Langzeit-Lerntest-Leistungen*, die in den *Status-Test-Leistungen* nicht nachweisbar waren (vgl. auch Hentrich & Reich, 1979). Auch ergaben Untersuchungen zu *Langzeit-Lerntests* keine kompensatorischen Effekte der Testdurchführung bezüglich der Persönlichkeitsmerkmale Neurotizismus (Guthke, 1972; siehe auch Blutke, 1979; Heyme & Militzer, 1977) sowie Ängstlichkeit, im

Gegensatz zu den berichteten Ergebnissen beim Einsatz von Kurzzeit-Lerntests (Wiedl & Bethge, 1981).

Die hier interessierenden Tests mit implementiertem Feedback-Angebot entsprechen am ehesten einer *Kurzzeit-Lerntest-Prozedur*. Die Befunde zu Kurzzeit-Lerntests zeigen, dass differentielle (kompensatorische) Effekte bei *leistungsängstlichen* Jugendlichen, *anstrengungsvermeidenden* Schülern, *impulsiven* Kindern oder Kindern mit *Planungsdefiziten* sowie auch Kindern mit *neurotischen* Tendenzen zu erwarten sind. Testangst scheint generell die Testleistungen zu beeinträchtigen. Die Befunde zu dynamischen Testverfahren sind hier nicht ganz eindeutig, weisen aber auf eine geringere Bedeutung des Faktors „Testangst“ beim Einsatz von Kurzzeit-Lerntests, insbesondere bei Wiederholungsmessung hin. Dabei ist die *Worry*-Komponente der Angst bedeutsam. Andererseits zeigen Befunde zur Leistungsangst nach Meijer (2001) auch, dass die Validität von Kurzzeit-Lerntests differentiell variiert. Für bestimmte Personengruppen – beispielsweise Leistungsängstliche – besitzen Kurzzeit-Lerntests eine höhere prädiktive Validität als für andere Personengruppen – beispielsweise nicht leistungsängstliche Personen. Auch sollen Kurzzeit-Lerntest-Prozeduren angstreduzierend wirken (Wiedl et al., 1982).

Analysestrategie: Es handelt sich bei der Mehrzahl der Untersuchungen um bivariate Korrelationsanalysen die bestenfalls quasiexperimentelle Ansätze verfolgen. Ausnahmen hierbei bilden Studien von Wiedl und Carlson (*Testing-the-Limits*, 1976, 1979, 1980, 1989, vor allem zu Impulsivität/Reflexivität, aber auch zu Neurotizismus). Meijer (1993, 2001) bildet hier insofern eine doppelte Ausnahme. Seine Analysen zu moderierenden Effekten von Leistungsangst im Dynamischen Testen bedienen sich darüber hinaus der Methode der Strukturgleichungsmodellierung und beziehen sich somit auf die Ebene latenter Konstrukte, d.h. sie verharren nicht auf der Ebene manifester Variablen.

Feedback-Interventionen: Nur Wiedl und Carlson (1979, 1980) haben sich ganz gezielt mit der Betrachtung einzelner Aspekte der Intervention in dynamischen Verfahren (hier dynamische Variante der CPM) beschäftigt und diese experimentell kontrastiert. Hier wurde deutlich, dass differentielle Effekte von der *Feedback-Art* abhängen. Dabei erwies sich *elaboriertes Feedback* im Gegensatz zu einfachem Richtig/Falsch-Feedback bei Vorgabe progressiver Matrizen im Grundschulalter als kompensatorisch im Hinblick auf impulsives Arbeitsverhalten und deshalb als leistungsförderlich.

Validitätsdiskussion: Von den verschiedenen Autoren werden unterschiedliche Argumentationslinien zur Rolle außerintellektueller Personmerkmale im Testprozess vertreten (differentielle Validität). Guthke (1972, Guthke & Lehwald, 1984) war bestrebt nachzuweisen, dass dynamische Tests (Lerntests) nicht im stärkeren Maße als traditionelle Verfahren Persönlichkeitsaspekte erfassen (siehe auch Meijer, 1993), wie von einigen Kritikern des Lerntest-Konzepts vermutet. Carlson und Wiedl (1979) dagegen wollen vor allem klären, *warum* eine dynamische Veränderung der Testadministration (hier der Raven-Matrizen) zur Verbesserung der Testleistungen führt (siehe auch Carlson, 1989; Guthke & Wiedl, 1996). Dabei wird eine *Kompensation* ungünstiger Leistungsvoraussetzungen, wie *Testangst*, *Motivationsdefizite*, *Impulsivität* oder mangelndes *Planungsverhalten* angenommen. Genauer soll mit Hilfe einer dynamischen Testprozedur der Problemlöseprozess systematischer und planvoller ablaufen (z.B. leistungsförderliche Veränderung des visuellen Suchverhaltens beim Lösen von progressiven Matrizen) und dadurch bestimmte außerintellektuelle Personmerkmale, wie z.B. Testangst, kompensiert werden. Carlson (1989) stellt dazu fest:

„... that dynamic assessment is effective because it can help certain children compensate for specific personal characteristics and/or dispositions that tend to reduce performance on mental ability measures. These include anxiety, lack of motivation, impulsivity and ineffective use of planning strategies.“ (Carlson, 1989, S. 19).

Meijer nimmt eine andere Argumentationslinie auf. Er versucht vor allem einen Validitätsgewinn dynamischer Verfahren gegenüber statischen Testverfahren nachzuweisen (siehe auch Beckmann, 2001), der nicht auf eine Reduktion der Leistungsangst oder anderer leistungshemmender nicht-intellektueller Merkmale zurückzuführen ist. Unterschiedliche Leistungen in beiden Verfahrenstypen dürfen sich demnach nicht mit einem unterschiedlichen Grad der Erfassung von Testangst erklären lassen. Dynamische Testverfahren sollten einen davon unabhängigen eigenen Beitrag zur Varianzaufklärung des angezielten Fähigkeitskonstrukts liefern. Eine Reduktion der gemessenen Testangst in dynamischen Testprozeduren, wie bei Wiedl und Carlson nachgewiesen, im Sinne einer Kompensation ungünstiger Personmerkmale des Testanden, ist für diesen Nachweis nicht ausreichend, sondern wird eher als Validitätsproblem angesehen (siehe vor allem Meijer & Elshout, 2001; aber auch Beckmann, 2001).

Eine entscheidende Erkenntnis der Analysen zur Kompensation, z.B. von kognitiver Impulsivität unter optimierenden Testbedingungen, ist aber, dass die Vorgabe einer bestimmten Testprozedur offensichtlich mit kognitiven und nicht-kognitiven (außerintellektuellen) Voraussetzungen des Testanden *interagiert*, somit also differentielle Effekte beim Einsatz von entsprechenden Testverfahren zu erwarten sind.

3.2.2 Das Modifikationsmodell

Bereits 1988 haben Tzuriel, Samuels und Feuerstein die Bedeutung außerintellektueller Merkmale beim Einsatz nicht-standardisierter dynamischer Testverfahren (hier das *Learning Potential Assessment Device*, LPAD, Feuerstein, 1979⁷) herausgestellt und diskutiert. Die diesen Testansatz motivierende Hauptkritik am Statustest-Ansatz bezieht sich in erster Linie auf die Vernachlässigung der Interaktion intellektueller und außerintellektueller Personmerkmale bei der Manifestation intellektueller Leistungsfähigkeit im Testscore (siehe auch Tzuriel, 2001). Im Rahmen dieses Ansatzes wird davon ausgegangen, dass auch im Gegensatz zu anderen bis dahin publizierten dynamischen Testansätzen im LPAD ungünstige Effekte außerintellektueller Merkmale, wie *intrinsische Motivation, Frustrationstoleranz, Kontrollüberzeugungen, Angst vor Misserfolg* und *Defensivität, Selbstvertrauen, Vitalität* und *Wachheit* auf die Testleistung nicht nur kontrolliert werden können, sondern auch über den Prozess der *Mediation* so verändert werden, dass im Testprozess ein Effekt außerintellektueller Merkmale von dem kognitiver Merkmale unterschieden werden kann.

Entscheidend hierbei sei die vom Testleiter gestaltete Beziehung zum Testanden (oft Kinder mit Lernschwierigkeiten), die als zentrales Merkmal des LPAD charakterisiert wird. Eine besondere Rolle kommt dem vom Testleiter während der Testung vermittelten *Feedback* zu, das sich nicht nur auf den kognitiven Prozess beziehen soll, sondern auch auf motivationale und temperamentbezogene Verhaltenscharakteristika des Testanden, die seine Denkprozesse und damit seine Leistung beeinflussen. Eine Standardisierung der vom Testleiter erwarteten Interventionen besteht dabei einzig in 10 aufgestellten Kriterien⁸ (vgl. z.B. Tzuriel, 1988), die als Richtlinien für den Testleiter für

⁷ Das LPAD nach Feuerstein stellt eines der ersten Verfahren der Verfahrensgruppe dar, die eine nicht-standardisierte dynamische Testprozedur (*Dynamic Assessment*) vertreten. Erste Veröffentlichungen stammen, ebenfalls wie solche zum Lerntest-Konzept (Guthke, 1972), bereits aus den 70er Jahren und bilden eine weitere Schule des Dynamischen Testens.

⁸ Als drei *Hauptkriterien*, die medierte Lerninteraktionen ausmachen, werden genannt: *Intentionality* (wechselseitige Determiniertheit von Verhalten), *Transcendence* (Transferfunktion der Lernsituation) und

die gewünschte Beeinflussung der genannten außerintellektuellen Variablen dienen sollen. Diese Kriterien sind allerdings eher allgemein gehalten, so dass dem Testleiter, ähnlich einem Therapeuten, Gestaltungsspielraum zur Verfügung steht, solange er eine entsprechende Grundhaltung umsetzt, d.h. die zentrale Testleiter-Testanden-Beziehung herstellt und aufrechterhält. Als Folge dieser Testprozedur wird ein Umschwenken von einer *extrinsischen* Motivationslage seitens des Testanden, mit dem vordergründigen Ziel „dem Testleiter zu gefallen“, also sozial erwünschtem Verhalten, zu einer *intrinsischen* Motivationslage erwartet, also einer Motivation um der Aufgabenlösung selbst willen. Eine solche Motivierung soll sich in erhöhter Aufgabenfreude und Zufriedenheit des Testanden äußern (siehe auch Gamlin, 1996; Ryder, 1995). Gamlin (1996) argumentiert, dass innerhalb des Feuersteinschen Ansatzes das Kind in der Gesamtheit einschließlich seiner ganz individuellen Bedürfnisse betrachtet werde, und eben nicht nur sein Denken. Zur Zielstellung einer nicht-standardisierten dynamischen Testprozedur schreibt Gamlin (1996): „DAI [Dynamic Assessment Instruction, d.V.] should motivate the youngsters to want to learn how to learn. ... We want our youngsters to become excited by the learning process itself.“ (S. 85). Die Not einer „unstandardisierten“ Testsituation wird hier sozusagen zur Tugend stilisiert. Problematisch ist, dass diese Ausführungen als auch Tzuriels (1988) Argumentationen nach meiner Kenntnis empirisch nicht überprüft sind.

Die Erfassung der Veränderbarkeit (*Modifiability*) nichtintellektueller Persönlichkeitsmerkmale im Testprozess wird als ein Hauptziel des LPAD aufgefasst (siehe auch PLAD⁹, Lidz, 1990). Das Ausmaß, indem eine solche Veränderung der außerintellektuellen Merkmale des Testanden möglich ist, soll im Anschluss an die Diagnostik als Anhaltspunkt für die Auswahl geeigneter Interventionsmethoden herangezogen werden. Das innerhalb des LPAD-Ansatzes vertretene Modell der Rolle außerintellektueller Variablen im Testprozess kann deshalb als *Modifikationsmodell* verstanden werden.

Meaning (affektive und motivationale Bedeutung des Lerninhalts für den Lernenden, siehe z.B. Tzuriel, 1988).

⁹ PLAD (*Preschool Learning Assessment Device*, Lidz, 1990): Das PLAD ist ein auf dem Ansatz von Feuerstein et al. (1979) basierendes dynamisches Testverfahren zur Erfassung der kognitiven Veränderbarkeit bei Vorschulkindern. Außerintellektuelle Variablen, die erfasst und modifiziert werden sollen, sind beispielsweise: Impulskontrolle/*Selbstregulation* (z.B. der Testand toleriert Frustration, wartet Instruktionen ab, denkt bevor er handelt, folgt Regeln), *Selbstbewusstheit* (z.B. belohnt sich bei Erfolg, kann eigene Fähigkeiten beschreiben), *Motivation* (z.B. Enthusiasmus, Arbeitsbereitschaft, Ausdauer), Interaktionsverhalten (z.B. Affektausdruck, zeigt Freude an der Interaktion mit dem Testleiter, reagiert auf Lob und Tadel).

Schlussfolgerungen

Zusammenfassend muss kritisch angemerkt werden, dass im vorgestellten Ansatz das dynamische Testkonzept einem *standardisierten*, psychometrischen Testkonzept gegenübergestellt wird (siehe auch Tzuriel, 2001; siehe aber Tzuriel & Shamir, 2002 zum computergestützten *Mediated-Learning*-Ansatz). Vorzüge einer standardisierten Testprozedur werden allgemein vor allem in der Durchführungs-, Auswertungs- und Interpretationskonkordanz gesehen und sind als allgemeine Richtlinien zur Konstruktion von psychodiagnostischen Instrumenten anerkannt. In *psychometrischen* dynamischen Testansätzen werden Rückmeldungen standardisiert vorgegeben. Es ist also für den Testleiter genau festgelegt bzw. im Computerprogramm definiert, nach welcher Falsch-Antwort des Testanden bezogen auf ein bestimmtes Item welche Rückmeldung angeboten wird. So wird zum Beispiel in der ACIL eine auf den Fehler des Testanden abgestimmte, d.h. fehlerspezifische helfende Lehreinheit präsentiert (Guthke et al., 1995; Beckmann & Guthke, 1999). Somit ist gesichert, dass jedem Testanden in Abhängigkeit der gezeigten Leistung individualisierte Interventionen zuteil werden, dies jedoch im Sinne einer interindividuellen Vergleichbarkeit standardisiert erfolgt.

Ein weiteres Problem besteht in der empirischen Untermauerung der Ergebnisse zur Veränderbarkeit außerintellektueller Merkmale im nicht-standardisierten, dynamischen Testprozess. Mit Ausnahme von Untersuchungen zum Merkmal Impulsivität/Reflexivität (Tzuriel, 2001) liegen meines Wissens keine Befunde zu einer systematischen Erfassung und experimentellen Variation außerintellektueller Merkmale beim Einsatz von dem Feuersteinschen Ansatz verpflichteten dynamischen Verfahren vor. Stattdessen wird der Leser zurückgelassen mit der Vermutung, dass die Ausführungen zur Veränderbarkeit außerintellektueller Merkmale auf subjektiven Beobachtungen einzelner Testleiter beruhen, und in keiner Weise systematisch erfasst und analysiert wurden.

Ein weiterer Unterschied des vorgestellten Ansatzes im Vergleich zu psychometrischen dynamischen Testansätzen besteht darin, dass sich hier das vom Testleiter verabreichte Feedback eben nicht nur auf Aspekte des kognitiven Prozesses beziehen soll, wie beispielsweise bei Reasoning-Aufgaben der Hinweis auf eine bestimmte zu beachtende Regel, sondern auch auf *motivationale* und *emotionale* Verhaltensaspekte des Testanden. Die hierbei postulierte therapeutische Grundhaltung des Testleiters kennzeichnet und unterscheidet diesen Testansatz von anderen in der Tradition dynamischer

Verfahren stehender Testansätze, der es dem Testleiter in besonderer Weise ermöglicht (via *Mediated Learning*), auf bestimmte außerintellektuelle Verhaltensmerkmale des Testanden therapeutisch (Tzuriel et al., 1988) einzugehen. Zahlreiche Untersuchungen zum *Mediated Learning* vor allem auch bezogen auf die Mutter-Kind-Dyade, als Prototyp einer natürlichen Intervention (vgl. Tzuriel et al., 1999), nicht allerdings bezogen auf die Testleiter-Testanden-Beziehung, haben gezeigt, dass außerintellektuelle Personmerkmale des Kindes den *Mediated Learning* -Prozess determinieren. Emotionale und motivationale Faktoren werden somit im Rahmen dieses Testansatzes als entscheidende Einflussgrößen verstanden, die über die kognitive Veränderbarkeit des Testanden bestimmen und also über den Prozess des *Mediated Learning* vermittelt, weniger allerdings direkt, Leistungsvarianz aufklären.

3.3 Zusammenfassung zu Kapitel 3

In Kapitel 3 wurde eine Forschungsrichtung – das Dynamische Testen – vorgestellt, die die Nutzung von Feedback-Interventionen im Rahmen der Psychodiagnostik untersucht mit dem Ziel, validere Diagnosen über intellektuelle Fähigkeiten (aber auch außerintellektuelle Personmerkmale, z.B. Konzentration, Eignung) zu stellen als die vielfältig kritisierten traditionellen Intelligenztests. Unter den verschiedenen international publizierten Verfahren lassen sich standardisierte Testprozeduren, z.B. das Lerntest-Konzept nach Guthke (1972), von nicht-standardisierten Testprozeduren, z.B. das *Dynamic Assessment* nach Feuerstein (1972), unterscheiden. Innerhalb standardisierter dynamischer Testprozeduren wird *Feedback* als *leistungskontingentes, verbales, standardisiertes*, mehr oder weniger *elaboriertes* Informationsangebot durch einen Testleiter oder computergestützt vorgegeben, mit dem Ziel, Lernmöglichkeiten im Testprozess zu schaffen bzw. die Testbedingungen zu optimieren. Testverfahren, die Feedback und Strategietraining (z.B. Kurzzeit-Lerntests) beinhalten, sind von höherer Effektivität als zwischen zwei Testphasen geschaltete Trainingprogramme (*Coaching*, z.B. Langzeit-Lerntests; *Mediating*). Insgesamt ist das Befundbild zur Wirksamkeit von Feedback-Interventionen beim Dynamischen Testen weniger kontrovers als das in der traditionellen Feedback-Forschung diskutierte Befundmuster. Vertreter beider Forschungstraditionen – Dynamisches Testen sowie traditionelle Feedback-Forschung – betonen die Notwendigkeit, nicht-kognitive außerintellektuelle Personmerkmale stärker in die Analysen zur Effektivität von Feedback-Interventionen einzubeziehen.

Innerhalb der Forschungstradition zum Dynamischen Testen können bei der Betrachtung der Wirkung außerintellektueller Merkmale beim Zustandekommen der Testleistung zwei Modellvorstellungen unterschieden werden. Autoren, die sich der Entwicklung standardisierter Testverfahren verpflichtet sehen (z.B. Carlson & Wiedl, 1979; Guthke, 1972) gehen davon aus, dass dynamische Verfahren unter anderem deshalb validere Diagnosen liefern, da sie *kompensatorisch* auf bestimmte leistungshinderliche nicht-intellektuelle Merkmale wirken und somit Subgruppen von Testanden, z.B. impulsive Kinder, aber auch beispielsweise Angehörige ethnischer Minoritäten, zu einer Leistungsverbesserung verhelfen (*Kompensationsmodell*). Dagegen betonen Vertreter eines nicht-standardisierten dynamischen Testansatzes die Beziehungsgestaltung zwischen Testleiter und Testanden (z.B. Feuerstein, 1972; Tzuriel, 1988), die im Sinne einer therapeutischen Intervention schon im Testprozess zu einer deutlichen positiven Veränderung ungünstiger außerintellektueller Leistungsvoraussetzungen auf Seiten des Testanden führen soll (*Modifikationsmodell*).

Guthke und Wiedl (1996) fassen zusammen, dass Leistungen in Lerntests als ein Ansatz des standardisierten Dynamischen Testens nicht so sehr durch leistungshemmende Persönlichkeitseigenschaften (Ängstlichkeit, Stressempfindlichkeit, Impulsivität) und dafür stärker durch leistungsfördernde Persönlichkeitseigenschaften (Erkenntnisstreben, reflexiver Arbeitsstil) beeinflusst werden. Das gilt vor allem für die in den 70er und 80er Jahren eingesetzten Langzeit-Lerntests und die als Papier und Bleistift-Test applizierten Kurzzeit-Lerntests. Insgesamt existieren aber nur wenige und oft widersprüchliche Erkenntnisse zur Frage einer unterschiedlichen Verarbeitung und Nutzung von Feedback in standardisierten dynamischen Testprozeduren in Abhängigkeit von verschiedenen Persönlichkeitseigenschaften. In dazu vorliegenden Studien dominiert die Analyse bivariater Korrelationsmuster.

Carlson und Wiedl fordern bereits 1980 weitere Forschungen zur Interaktion von Testprozedur und Persönlichkeit. Dieser Forderung wurde zwar zum Teil in Bezug auf unterschiedliche Personengruppen mit spezifischen Defiziten nachgekommen, z.B. Kinder aus ethnischen Minoritäten. Eine isolierte Betrachtung einzelner Personmerkmale unter experimentell kontrollierten Bedingungen fand jedoch kaum statt. Ebenso wurde in den entsprechenden Studien nicht nach verschiedenen Feedback-Arten unterschieden, die beim Dynamischen Testen angeboten werden. Die Forschungslage zu Feedback-Phänomenen im Leistungsbereich belegt jedoch die unterschiedliche Effektivität

von Feedback-Interventionen in Abhängigkeit ihres Elaborationsgrades. Eine Ausnahme bilden hier Arbeiten nach Carlson und Wiedl (z.B. 1976) zu einer Kurzzeit-Lerntest-Variante der *Raven Coloured Progressive Matrices* (CPM, Raven, 1956) in den 70er und 80er Jahren. Unter Kenntnis der Befundlage der Feedback-Forschung kann daher nicht ausgeschlossen werden, dass Personen mit ausgewählten auch nicht-kognitiven Persönlichkeitsmerkmalen, z.B. bestimmten Zielstrukturen, bei Anwendung unterschiedlicher Feedback-Interventionen in Lerntest-Prozeduren (vor allem Kurzzeit-Lerntests mit computergestützter Darbietung) Leistungsdefizite zeigen, die nicht auf mangelnde kognitive, speziell intellektuelle Fähigkeiten zurückgeführt werden dürfen.

4 Problemstellung

Die Zusammenschau der zahlreichen Forschungsarbeiten zur Effektivität von Feedback-Interventionen im Leistungsbereich stellt die große Variabilität der Leistungseffekte nach Feedback heraus. Insbesondere die aus der Perspektive eines rein kognitiven Feedback-Modells erwartungswidrigen Befunde (geringe bis fehlende oder sogar negative Leistungseffekte nach Feedback) führten konsequenterweise zur Forderung, auch nicht-kognitive Variablen in die Analysen einzubeziehen. Die Feedback-Interventions-Theorie (Kluger & DeNisi, 1996) stellt das bisher einzige Feedback-Modell dar, das eine Integration von kognitiven und nicht-kognitiven Erklärungsansätzen leistet, um Leistungsverhalten unter Feedback-Bedingungen aufzuklären und vorherzusagen. Die Feedback-Interventions-Theorie bietet sich daher als ein theoretischer Orientierungsrahmen für die vorliegende Studie an.

Ein Anwendungsgebiet für Feedback-Interventionen findet sich in der psychologischen Diagnostik. So verwenden Vertreter des Dynamischen Testens (Kapitel 3) Feedback-Strategien im Testprozess, um die interindividuelle Veränderbarkeit eines Personmerkmals zu erfassen und schließlich validere diagnostische Aussagen zu treffen. Traditionelle, auf den Status eines Personmerkmals abzielende Testverfahren, z.B. klassische Intelligenztests, vernachlässigen diesen Aspekt. Eine Moderation der Testleistung unter Feedback-Bedingungen durch nicht-kognitive außerintellektuelle Personmerkmale, wie im Rahmen der Feedback-Forschungstradition postuliert, muss dabei im Sinne der oft angestrebten Eindimensionalität der Validität eines Verfahrens ausgeschlossen werden.

Speziell in *Kurzzeit-Lerntests* werden typischerweise und im Gegensatz zu traditionellen Intelligenztests nach jeder Itembearbeitung Rückmeldungen über die Lösungsqualität gegeben. Darüber hinaus sollen spezifische Denkhilfen, d.h. elaboriertes Feedback, und eine performanzabhängige Schwierigkeitssteigerung in der Itemabfolge (adaptives Testen) dem Erwerb von Lösungsheuristiken und somit einer potentiellen Leistungssteigerung im Test dienen (siehe Abschnitt 3.1). Erwartungsgemäß unterscheiden sich Testanden in ihrer Fähigkeit von Feedback zu profitieren, welche als zusätzliche, diagnostisch relevante Information über ihre intellektuelle Leistungsfähigkeit interpretiert wird. Die vorliegende Arbeit soll vor diesem Hintergrund die Frage klären, ob die Nutzung von Feedback in Leistungstests ausschließlich von intellektuellen Merkmalen einer Person bestimmt wird, oder ob der Erfolg in Lerntests eventuell sogar im

stärkeren Maße von nicht-intellektuellen Persönlichkeitsmerkmalen abhängt. Mit anderen Worten soll der Frage nachgegangen werden, ob die Fähigkeit eines Testanden, im Testverlauf angebotenes Feedback zur Lösung nachfolgender Aufgaben zu nutzen, als Indikator für seine intellektuelle Leistungsfähigkeit interpretiert werden darf.

Werden die im Rahmen der Feedback-Interventions-Theorie dargelegten Wirkungszusammenhänge von Persönlichkeit und Leistungsverhalten auf eine Testsituation mit (Richtig/Falsch-)Feedback angewandt, lässt sich folgendes Erwartungsbild entwickeln: Während einer Intelligenztestung angebotenes Richtig/Falsch-Feedback löst aufgrund des gegebenenfalls selbstwertbedrohlichen Inhalts stets eine Veränderung des Aufmerksamkeitsfokus' des Testanden aus. Nur eine Aufmerksamkeitsausrichtung auf Aspekte der Aufgabe führt jedoch potentiell zu einer Leistungssteigerung nach Richtig/Falsch-Feedback. Wird die Aufmerksamkeit auf selbstrelevante Ziele (z.B. Leistungsziele, wie die Demonstration von Intelligenz im Test) gelenkt und eine Ist-Soll-Diskrepanz festgestellt, dann sind Leistungsbeeinträchtigungen zu erwarten, die nicht durch mangelnde intellektuelle Fähigkeiten erklärt werden dürfen. Gründe für Leistungsdefizite können stattdessen z.B. in einer Anstrengungsreduktion, in affektiven Reaktionen (Überaktivierung, Stimmungsabfall), oder in einer aufgabenunspezifischen Beanspruchung kognitiver Ressourcen (z.B. Worry-Kognitionen) vermutet werden. Der beschriebene Aufmerksamkeitsshift nach Feedback sollte durch außerintellektuelle Personmerkmale wie Selbstwertgefühl, Ängstlichkeit und motivationale Zielorientierungen moderiert sein. Zum Beispiel tendieren insbesondere leistungsängstliche Personen zu sorgenvollem Grübeln (Worry-Kognitionen) nach Misserfolg und zeigen deshalb eventuell Leistungsdefizite unter Feedback-Bedingungen.

Da Leistungsfeedback potentiell als eine Bedrohung des Selbstwertes wahrgenommen wird (Kluger & DeNisi, 1996), sollte besonderes Augenmerk auf die Analyse des Selbstkonzepts bzw. des Vertrauens in die eigenen Fähigkeiten in Leistungssituationen mit Feedback-Angebot gelegt werden. Bewährte persönlichkeits-theoretische Konzepte, die Aufklärung bei der Untersuchung von Leistungsverhalten unter Feedback-Bedingungen versprechen, sind z.B. das Konzept der *motivationalen Orientierung* in Lern- und Leistungssituationen (Dweck & Leggett, 1988; Dweck, 2000; Nicholls, 1984) und das Konzept der *Testangst* (Morris et al, 1981; Meijer, 1996, 2001). Beide Konzeptionen beinhalten Annahmen zur Rolle des Selbstkonzepts der eigenen Fähigkeiten (Begabungskonzept) in Leistungssituationen mit Erfolgs- und Misserfolgspotential. Das Konzept

der motivationalen Orientierung lässt zudem Aussagen über Zielstrukturen von Personen in Lern- und Leistungssituationen zu, die nach der Feedback-Interventions-Theorie wesentlich die Effektivität von Feedback bestimmen. Das Konzept der Testangst erlaubt Aussagen über die Tendenz einer Person in Leistungssituationen Worry-Kognitionen zu entwickeln, die als eine häufige Ursache für Leistungsdefizite nach Feedback-Angebot bewertet werden.

Metaanalytische Auswertungen bisheriger Befunde zur Effektivität verschiedener Feedback-Strategien unterstützten die von Kluger und DeNisi (1996, 1998) formulierten Modell-Annahmen (Kapitel 3). Der Wirkungszusammenhang von Persönlichkeit und Leistung unter Feedback-Bedingungen gilt jedoch bisher als nicht genügend aufgeklärt (Kluger & DeNisi, 1998; Bangert-Drowns, et al., 1991). Vor allem fehlt es an systematischen Untersuchungen auf der Basis eines geeigneten Feedback-Modells.

Mögliche leistungsbeeinträchtigende Konsequenzen einer Testprozedur mit Feedback, wie beispielsweise eine Anstrengungsreduktion nach Misserfolg bei wenig selbstkonzeptstarken Testanden, sind in der Forschungstradition zum Dynamischen Testen bisher ebenfalls nur unzureichend untersucht worden. Bisherige Studien haben aber gezeigt, dass dynamische Teststrategien geeignet sind, Leistungsdefizite von Personen mit so genannten ungünstigen, außerintellektuellen Leistungsvoraussetzungen (Ängstlichkeit, Stressempfindlichkeit, Impulsivität, Motivationsdefizite) zu kompensieren (Kompensationsmodell). Daher sind differentielle Effekte bei Applikation dynamischer Tests zu erwarten. Eine isolierte Betrachtung einzelner Personmerkmale unter experimentell kontrollierten Bedingungen, d.h. auch unter Beachtung der verschiedenen beim Dynamischen Testen eingesetzten Feedback-Strategien, fand mit Ausnahme weniger Untersuchungen zur kognitiven Impulsivität und Leistungsangst bisher nicht statt.

Ließen sich außerintellektuelle differentielle Effekte einer dynamischen Teststrategie tatsächlich nachweisen, kann die in diesen Testprozeduren typischerweise erfasste Fähigkeit des Testanden, von leistungsbezogenem Feedback zu profitieren, nicht mehr ausschließlich, wie von den Testautoren angenommen, als eine Abbildung des intellektuellen Leistungsniveaus des Testanden verstanden werden. Für eine geeignete Interpretation von Leistungen in Testverfahren mit implementiertem Feedback wäre eine solche Erkenntnis von entscheidender Bedeutung. Unterdurchschnittliche Leistungen nach Feedback dürften dann nicht mehr direkt auf eine mangelnde intellektuelle Begabung des Testanden zurückgeführt werden, sondern ließen sich unter Umständen auch

z.B. mit einem geringen Begabungskonzept erklären, welches eine geeignete kognitive Verarbeitung des Feedbacks beeinträchtigt und damit einen Feedback-Gewinn verhindert haben könnte.

Ziel der vorliegenden Studie ist es deshalb, einen Beitrag zur Aufklärung des Zusammenhangs von Persönlichkeit und Leistung unter Feedback-Bedingungen zu leisten. Speziell soll geprüft werden, ob die Leistung in einem Intelligenztest mit Feedback tatsächlich als ein Indikator für die intellektuelle Fähigkeit (Intelligenzpotenz, Abschnitt 3.1) des Testanden interpretiert werden darf, oder ob nicht im Widerspruch zum Kompensationsmodell andere, außerintellektuelle Merkmale berücksichtigt werden müssen. In diesem Rahmen stellt die vorliegende Arbeit gleichzeitig einen weiteren Beitrag zur differentiellen Validierung eines computergestützten dynamischen Tests, der Adaptiven Computergestützten Intelligenz-Lerntest-Batterie (ACIL, Guthke et al., 1996), dar.

Eine Prüfung von möglichen Moderatoreffekten wird im Sinne der Testfairness bei jeder Testentwicklung gefordert. Jäger (1978) spricht in diesem Kontext auch von der differentiellen Diagnostizierbarkeit, auch als *Differential Prediction* bei Aguinis (2004) referiert. Der von Michel und Iseler (1968) vorgeschlagene Begriff beinhaltet, dass unterschiedliche Testgütekriterien für auf der Basis von Moderatoren, z.B. Geschlecht, gebildete Teilgruppen zu erwarten sind. Routinemäßig werden deshalb bei Testentwicklungen potentielle Moderatoren wie das Geschlecht oder Alter berücksichtigt. Im Rahmen der Entwicklung von Tests mit Feedback liegt es nahe, auch andere Personmerkmale, z.B. das Begabungskonzept, auf eine potentielle Moderatorfunktion hin zu untersuchen. Die vorliegende Arbeit dient deshalb der weiteren Klärung der Frage nach potentiellen Moderatoren in Leistungstests mit Feedback.

Dynamische Teststrategien integrieren üblicherweise verschiedene Arten von Feedback. In Lerntests (z.B. der ACIL) implementiertes Feedback lässt sich vor allem nach seinem Informationsgehalt unterscheiden. So kann Feedback den Testanden „nur“ über die Richtigkeit seiner Aufgabenlösung informieren, Hinweise zu Lösungswegen enthalten (Denkhilfen, auch elaboriertes Feedback) oder auch die Aufgabenlösung präsentieren (siehe Abschnitt 3.1). Im Rahmen der Feedback-Forschung stellte sich jedoch Richtig/Falsch-Feedback als wenig effektiv heraus (siehe auch die berichteten Metaanalysen, z.B. Bangert-Drowns, et al., 1991; sowie Delgado & Prieto, 2003; Stankov & Crawford, 1997). Ein vergleichbarer (singulärer) Befund wird auch von Vertretern des Dynamischen Testens berichtet (Carlson & Wiedl, 1980). Dem scheinbar ungeachtet

stellt Richtig/Falsch-Feedback eine routinemäßig verwendete Feedback-Strategie in dynamischen Tests dar, speziell in Kurzzeit-Lerntests (auch Kurzzeit-Lerntests vom Typ I). In der vorliegenden Arbeit soll deshalb die Effektivität von Richtig/Falsch-Feedback in einem computergestützten Testverfahren überprüft werden. Die Arbeit fokussiert somit auf eine der in Lerntests enthaltenden Feedback-Arten. Ziel ist dabei eine höhere Generalisierbarkeit der Befunde zu erreichen. Zum einen sollen Schlussfolgerungen für den Aufbau von Feedback-Strategien in computergestützten, dynamischen Testprozeduren ermöglicht und nicht „nur“ Aussagen zu einem ganz konkreten dynamischen Testverfahren abgeleitet werden. Zum anderen gestattet dieses Vorgehen auch über die dynamische Testsituation hinaus Aussagen zum Zusammenhang von Persönlichkeit und Leistung unter Richtig/Falsch-Feedback-Bedingungen.

Zusammenfassend ist Folgendes festzuhalten: In vorliegender Arbeit soll zunächst geprüft werden, ob das in dynamischen Testverfahren häufig verwendete Richtig/Falsch-Feedback eine effektive Feedback-Strategie darstellt. Dabei wird vor dem Hintergrund der Forschung zur Effektivität von Feedback-Interventionen im Rahmen des Dynamischen Testens als auch der Feedback-Forschungstradition ein moderater, positiver performanzbezogener Feedback-Effekt erwartet. Dies wird im Folgenden als *generelle Feedback-Hypothese* bezeichnet. Des Weiteren soll geprüft werden, ob die Leistung in einem Intelligenztest mit Feedback tatsächlich als ein Indikator für die intellektuelle Fähigkeit (auch Intelligenzpotenz, Abschnitt 3.1) des Testanden interpretiert werden darf, oder ob nicht andere außerintellektuelle Merkmale, die den Erfolg bei der Bearbeitung von Leistungstests mit Feedback bestimmen, beachtet werden müssen. Diese sind die Testangst, die motivationale Orientierung sowie das Selbstvertrauen in die eigenen Fähigkeiten. Darauf wird im Folgenden mit dem Begriff *differentielle Feedback-Hypothese* Bezug genommen. Aus der Zusammenschau der Befunde sollen Erkenntnisse zur differentiellen Validität von Leistungstests mit Feedback abgeleitet sowie Vorschläge für die Konstruktion von geeigneten Feedback-Strategien formuliert werden.

Zur Prüfung der Hypothesen zum Zusammenhang von Persönlichkeit und Leistung (*differentielle Feedback-Hypothese*) empfiehlt sich der *Moderatoransatz* (z.B. Aguinis, 2004; Bartussek, 1970; Baron & Kenny, 1986; Jäger, 1978) im Allgemeinen und die multiple moderierte Regression (Saunders, 1954, 1956) im Speziellen. Die Methode der moderierten Regression wurde erstmals von Saunders (1954) in seinem Kongressbeitrag „The moderator variable as a useful tool in prediction“ zur Bestimmung von

Moderatoreffekten vorgeschlagen, den er 1954 auf der *Invitational Conference on Testing Problems* des *Educational Testing Service* vortrug (herausgegeben 1966 von Anne Anastasi).

Die multiple moderierte Regression (MMR) ist anderen Ansätzen zur Bestimmung von Moderatoreffekten, z.B. der Berechnung von Gruppen-Korrelationskoeffizienten (siehe Jäger, 1978) vorzuziehen. Ein Vorteil wird zum Beispiel in der Möglichkeit gesehen, kontinuierliche potentielle Moderatorvariablen ohne Stichprobenfraktionierung analysieren zu können (Stichprobenfehler, Willkürlichkeit der Stichprobenteilung; Bartussek, 1970). Die multiple moderierte Regression erlaubt demnach ein multiplikatives Modell entsprechend eines varianzanalytischen Designs. Das bedeutet, dass interaktionale Effekte auch bei Verwendung von kontinuierlichen Variablen geprüft werden können.

Fragestellung

Vor den fragestellungsbezogenen Analysen wird, der psychometrischen Tradition folgend, die Qualität der verwendeten Messinstrumente als Indikatoren für die jeweils angezielten Konstrukte zu prüfen sein. Dazu sind Validitäts- und Reliabilitätskennwerte zu bestimmen. Erst der Nachweis einer befriedigenden psychometrischen Qualität der eingesetzten Verfahren stellt die Voraussetzung für die sich anschließenden Analysen folgender Fragen dar.

1. Besteht ein genereller Treatmenteffekt?

- Es wird erwartet, dass das Angebot von Feedback im Testprozess im Allgemeinen zu einer Leistungsverbesserung – operationalisiert als Anzahl der gelösten Items – der Testanden führt verglichen mit feedbackfreien Testbedingungen. Die Prüfung dieser Hypothese soll zum einen über eine Kontrastierung zweier Testvarianten (mit versus ohne Feedback) zu zwei Messzeitpunkten innerhalb einer Experimentalgruppe erfolgen. Diese Hypothese wird im Folgenden als *intraindividuelle Feedback-Hypothese* bezeichnet. Zum anderen sollen Leistungsunterschiede zwischen verschiedenen Experimentalgruppen (Feedback/kein Feedback) zu einem Messzeitpunkt geprüft werden. Diese Hypothese wird im Folgenden *interindividuelle Feedback-Hypothese* genannt.
- Die Hypothesenprüfung soll für zwei Materialbereiche (figurales Reasoning, numerisches Reasoning) getrennt erfolgen, um eine bessere Generalisierbarkeit der Befunde zu ermöglichen. Es wird erwartet, dass die feedbackbedingte Leistungsverbesserung

serung im figuralen Materialbereich größer ist als im numerischen Materialbereich, da eine fehlerhafte Bearbeitung der numerischen Reasoning-Aufgaben (Zahlenfolgen) neben Defiziten im schlussfolgernden Denken auch durch mangelnde Rechenfertigkeiten bedingt sein kann. Da der Erwerb von Rechenfertigkeiten aber einen relativ langen Wissenserwerbs- und Automatisierungsprozess voraussetzt, können keine Verbesserungen in den Rechenfertigkeiten durch Feedback im Testprozess erwartet werden.

Darauf aufbauend gilt es die Frage zu beantworten:

2. Bestehen differentielle Treatmenteffekte?

- Zunächst ist im Sinne der Kontrolle des Wilderschen Ausgangswertgesetzes (Wilder, 1931) zu prüfen, ob eine Leistungsverbesserung, die auf die erfolgreiche Nutzung von Feedback-Interventionen zurückzuführen ist, in Abhängigkeit des Ausgangswertes variiert. Auf diese Hypothese wird mit dem Begriff *Ausgangswerthypothese* Bezug genommen.
- Weiterhin wird angenommen, dass ein Teil der Varianz zwischen den experimentellen Gruppen (Feedback/kein Feedback) neben Unterschieden in der Fähigkeit, von Feedback zu profitieren, als Indikator für die intellektuelle Leistungsfähigkeit des Testanden, auch auf Unterschiede in *außerintellektuellen Personmerkmalen* zurückzuführen ist bzw. durch diese moderiert wird. In diesem Sinne werden folgende Annahmen zu prüfen sein:
 - Personen mit niedrigem Begabungskonzept verbunden mit einer *Leistungszielorientierung* sollten nicht von Feedback-Interventionen profitieren oder sogar Leistungseinbußen unter Feedback-Bedingungen zeigen, da sie Feedback vor allem als Evaluation ihrer Kompetenz wahrnehmen und auf Misserfolge oder Hindernisse bei der Aufgabenlösung mit hilflosem Verhalten (z.B. Anstrengungsreduktion) reagieren. Dagegen sollten sowohl Personen mit hohem Begabungskonzept als auch Personen mit Lernzielorientierung einen Feedback-Gewinn erzielen, da diese Personen Feedback vor allem als zusätzliches Informationsangebot (z.B. über ihr erreichtes Leistungsniveau) wahrnehmen und auch bei Misserfolgen oder Hindernissen bei der Aufgabenlösung meisterndes Verhalten zeigen (z.B. Aufrechterhaltung der Anstrengung). Das Begabungskonzept stellt dabei einen Moderator für den Zusammenhang von Zielorientierung und Leistung unter

Feedback-Bedingungen dar. Diese Hypothese wird als *Moderatorhypothese Begabungskonzept* bezeichnet.

- Um Aussagen zur *Verursachung* der angenommenen Leistungsbeeinträchtigung durch Feedback zu treffen, ist weiterhin zu prüfen, ob Personen mit niedrigem Begabungskonzept und Leistungszielorientierung tatsächlich Besonderheiten in ihren Attributionstendenzen und ihrer Anstrengungsbereitschaft aufweisen. Hier wird angenommen, dass Personen mit niedrigem Begabungskonzept verbunden mit einer Leistungszielorientierung eher ungünstige Ursachenerklärungen für Erfolg/Misserfolg nach Feedback-Erhalt vornehmen (beispielsweise Misserfolge stabil und internal attribuieren) und bei Misserfolg mit einer Verringerung der Anstrengungsbereitschaft bzw. mit Anstrengungsvermeidung reagieren.
- Ebenso wird erwartet, dass testängstliche Personen nicht von Feedback profitieren oder sogar durch ein Feedback-Angebot bezüglich ihrer Performanz beeinträchtigt werden. Dabei wird angenommen, dass Testangst mit einem geringen Begabungskonzept einhergeht, oder durch ein geringes Begabungskonzept *verursacht* wird. Diese Hypothese wird im Folgenden als *Mediatorhypothese Testangst* bezeichnet.

5 Methodik

5.1 Versuchsplan

Wie in den vorangegangenen Kapiteln herausgearbeitet, stellt Feedback ein wesentliches Merkmal dynamischer Tests im Allgemeinen und Lerntests im Besonderen dar. Folgende Darstellung soll nochmals das Verhältnis von Lerntests, Statustests und Feedback aufzeigen, um somit die Wahl der in vorliegender Studie verwendeten Leistungstests zu begründen.

Lerntest-Prozeduren (speziell Kurzzeit-Lerntests) unterscheiden sich von traditionellen Statustest-Prozeduren ohne Feedback in mindestens drei Merkmalen:

- Adaptivität der Itemabfolge,
- Angebot von Richtig/Falsch-Feedback,
- Angebot fehlerspezifischer Hilfen (elaboriertes Feedback).

In Tabelle 5.1 sind die möglichen Kombinationen dieser Merkmale und ihre Testpendants aufgeführt. Die vorgestellte Untersuchung fokussiert auf die Gegenüberstellung von Verfahren mit Information über die Qualität der Lösung als ein typisches Merkmal von dynamischen Testprozeduren auf der einen Seite und feedbackfreien Tests, d.h. traditionellen Intelligenztests als Statustests auf der anderen. Die hier interessierenden Tests sind in Tabelle 5.1. hervorgehoben.

Tabelle 5.1

Mögliche Kombinationen der Merkmale von Intelligenzstatustest und Lerntest	
<i>Merkmalskombination</i>	<i>Anforderungscharakterisierung</i>
A+ R+ H+	<i>fehlerorientiert-adaptiver Kurzzeit-Lerntest</i>
A+ R+ H-	adaptiver Test mit expliziter Information über die Richtigkeit der Antwort
A+ R- H-	adaptiver Test
A+ R- H+	adaptiver Test mit Hilfeangebot, ohne explizite Information über die Richtigkeit der Antwort
A- R+ H+	Test mit expliziter Information über die Richtigkeit der Antwort und Hilfeangebot
A- R+ H-	Test mit expliziter Information über die Richtigkeit der Antwort
A- R- H-	<i>traditioneller Statustest</i>
A- R- H+	Test mit Hilfeangebot, ohne explizite Information über die Richtigkeit der Antwort
<i>Anmerkung:</i>	A = Adaptivität der Testprozedur R = Information über die Richtigkeit der Antwort (Richtig/Falsch-Rückmeldung) H = Hilfeangebot (fehlerspezifische Denkhilfen im Sinne einer elaborierten Rückmeldung) + = Merkmal vorhanden - = Merkmal nicht vorhanden

Zur Prüfung der Effekte von Richtig/Falsch-Feedback auf Testleistungen in Abhängigkeit von auch außerintellektuellen Personmerkmalen wurde ein experimentelles Vorgehen gewählt. Ziel dabei war, Leistungen in Tests mit Feedback Leistungen in feedbackfreien jedoch sonst identischen Itempools gegenüberzustellen. Um neben interindividuellen auch intraindividuelle Leistungsveränderungen unter der experimentellen Bedingungsvariation (Feedback/kein Feedback) analysieren zu können (*intraindividuelle* Feedback-Hypothese), wurde ein messwiederholtes Versuchsdesign realisiert (Tabelle 5.2). Eine Balancierung der Testreihenfolge diente der Kontrolle von Sequenzeffekten. Im Sinne einer Generalisierbarkeit der Befunde sollten die Effekte von Feedback auf die Testleistung in zwei Reasoning-Domänen untersucht werden (figural, numerisch). Tabelle 5.2 gibt eine Übersicht über das angestrebte Untersuchungsdesign.

Tabelle 5.2

Versuchsplan (Leistungstestung, siehe auch Tabelle 5.6)

Materialbereich		Sitzung 1	Sitzung 3	N in Prozent (geplant)
figural	Versuchsgruppe 1	Kein Feedback	Feedback	25
	Versuchsgruppe 2	Feedback	Kein Feedback	12.5
	Kontrollgruppe	Kein Feedback	Kein Feedback	12.5
numerisch	Versuchsgruppe 1	Kein Feedback	Feedback	25
	Versuchsgruppe 2	Feedback	Kein Feedback	12.5
	Kontrollgruppe	Kein Feedback	Kein Feedback	12.5

Anmerkung: abhängige Variable: Antwortqualität (Anzahl Richtige)

Somit kamen acht Testvarianten zum Einsatz. Basis hierfür bildeten die *Figurenfolgen* aus dem Itempool des Adaptiven Figurenfolgen-Lerntest sowie die *Zahlenfolgen* aus dem Itempool des Adaptiven Zahlenfolgen-Lerntest (beide Bestandteil der ACIL, Guthke, et al., 1995). Die Testprozedur wurde entsprechend der Anforderungen an die Untersuchung modifiziert (nicht adaptiv, je nach experimenteller Bedingung mit oder ohne Richtig/Falsch-Feedback). Die Itempools wurden in jeweils zwei Parallelversionen mittels Odd/Even-Technik aufgeteilt. Zur Prüfung von Reihenfolgeeffekten der so entstandenen Paralleltests wurde ihre Vorgabe ebenfalls balanciert (Tabelle 5.6). Eine weitere zwischen beiden Leistungstests durchgeführte Sitzung diente der Erfassung eines Teils der persönlichkeitsdiagnostischen Daten. Der Untersuchungsablauf beinhaltete somit insgesamt drei Testsitzungen je Versuchsperson.

Für die Untersuchung sollte eine Stichprobe im Altersbereich zwischen 13 bis 16 Jahren rekrutiert werden, die ein möglichst breites und somit repräsentatives Spektrum der intellektuellen Leistungsfähigkeit abdeckt.

5.2 Stichprobe

Die Datenerhebung erfolgte im Rahmen eines Forschungsaufenthalts an der Universität Sunderland (Kooperationspartner in Sunderland: Prof. Julian G. Elliott) und fand am Macmillan College in Middlesbrough (England) im Juni und Juli 2002 statt. Das Macmillan College wurde ausgewählt, da die Schulaufnahmerichtlinien der Maßgabe der Normalverteilung hinsichtlich der Leistungsfähigkeit der Schüler folgen. Somit war mit einer hinreichenden Leistungsvarianz innerhalb der Stichprobe zu rechnen. Der angestrebte Altersbereich war vorrangig durch den Geltungsbereich der einzusetzenden Verfahren determiniert. Hauptgrund dafür ist die Tatsache, dass die leistungsdiagnostischen Verfahren hinsichtlich ihrer psychometrischen Qualität im Geltungsbereich für Schüler von 11 bis 16 Jahren als bewährt gelten können (ACIL, siehe Abschnitt 3.1.1).

Nachdem von Seiten der pädagogischen Entscheidungsträger des Macmillan College Interesse an der Untersuchung bekundet wurde, erfolgte die Rekrutierung der Versuchspersonen mit Hilfe des Schulleiters. Dazu wurden alle Eltern der Schüler der achten und neunten Klassen¹⁰ des Jahrgangs 2002 schriftlich über die Zielstellung und den Ablauf der geplanten Untersuchung informiert. Die Teilnahme der Schüler an der Untersuchung basierte auf Freiwilligkeit und war unentgeltlich. Die vertrauliche Behandlung der Daten wurde zugesichert. Voraussetzung für die Teilnahme war die schriftliche Einwilligung der Eltern.

Fünf von 422 angesprochenen Schülern willigten nach Auskunft des Schulleiters nicht ein, an der Untersuchung teilzunehmen. Die Untersuchungsstichprobe bestand somit zunächst aus 417 Schülern im Alter von 13-15 Jahren. Von 346 Schülern sind aus mindestens einer der beiden Leistungstest-Sitzungen Daten vorhanden. Insgesamt liegen von 249 Schülern (143 in der Bedingung Figurenfolgen; 106 in der Bedingung Zahlenfolgen) vollständige Datensätze aus allen drei Testsitzungen vor. Der Drop-Out von 40% (168 von 417) erklärt sich im Wesentlichen durch Gründe, wie sie in Untersuchungen im Schulsetting zu erwarten sind (organisatorische Gegebenheiten wie Abwesenheit

¹⁰ Die achte und neunte Klassenstufe entspricht hinsichtlich der Altersgrenzen der siebten und achten Klassenstufe in deutschen Schulen.

der Schüler zu einem der drei Testtage, einschließlich Drop-In zum zweiten Testzeitpunkt), sowie durch technische Probleme (Datenverlust durch ein Serverproblem). Die zufällige Zuweisung der Versuchspersonen zu den verschiedenen experimentellen Bedingungen verhinderte, dass Drop-Outs einen systematischen Einfluss auf potentielle Effekte hatten.

Zur Analyse der Stichprobeneigenschaften wurde auf zusätzlich durch die Schule bereitgestellte Daten, nämlich routinemäßig an der Schule durchgeführte Schulleistungstests¹¹ sowie Intelligenztests¹², zurück gegriffen. So lässt sich nachweisen, dass die Stichprobe von 249 Schülern, für die vollständige Datensätze vorhanden sind, aus derselben Grundgesamtheit stammt, wie die Stichprobe von 168 Schülern mit unvollständigen Datensätzen (Tabelle 5.3 und Tabelle 5.4). Zur Stichprobe der Schüler mit unvollständigen Datensätzen werden auch diejenigen Schüler gezählt, für die ausschliesslich persönlichkeitsdiagnostische Daten aus der hier vorgestellten Studie vorliegen. Da für beide Klassenstufen unterschiedliche zusätzliche Daten vorhanden waren, wird die Betrachtung der Stichprobeneigenschaften für beide Klassenstufen getrennt vorgenommen. Die Tabellen 5.3 und 5.4 zeigen die Gegenüberstellung der beiden Stichproben mit vollständigen ($n_1 = 249$) bzw. unvollständigen Datensätzen ($n_2 = 168$) in den zusätzlich von der Schulleitung bereit gestellten Daten zur Schulleistung (KS2, 8. Klasse: $U_{KS2, English} = 3909.0$, $p = .167$, $N = 190$; $U_{KS2, Math} = 4020.0$, $p = .350$, $N = 189$; $U_{KS2, Science} = 3808.0$, $p = .092$, $N = 190$; KS3, 9. Klasse: $U_{KS3, English} = 3973.0$, $p = .057$, $N = 201$; $U_{KS3, Math} = 4597.5$, $p = .795$, $N = 201$; $U_{KS3, Science} = 4558.0$, $p = .715$, $N = 201$) und zur intellektuellen Leistungsfähigkeit (MidYIS, 8. Klasse: $t = .637$, $p = .525$, $N = 210$; CAT, 9. Klasse: $t = -.321$, $p = .749$,

¹¹ *Schulleistungstests*: Als Indikatoren für die schulische Leistungsfähigkeit in drei Fächern (Englisch, Mathematik, Science) liegen Daten des *National Curriculum Key Stage 2 Tests* (KS2 SATs, 10/11 Jährige bzw. 6. Klassenstufe) für die achte Klassenstufe und des *National Curriculum Key Stage 3 Tests* (KS3 SATs, 13/14 Jährige bzw. 8. Klassenstufe) für die neunte Klassenstufe vor. Beide Schulleistungstests sind als lehrzielorientierte Tests konstruiert, wobei das staatlich festgelegte Kriterium (Klassenziel) für die achte Klasse „4“ und für die neunte Klasse „5“ beträgt. Beide Verfahren sind national normiert (England, Wales, Northern Ireland) und werden routinemäßig jeweils im Mai des entsprechenden Schuljahres per Gesetz in allen Schulen durchgeführt (Quelle: <http://www.bbc.co.uk/schools>).

¹² *Intelligenztests*: Als Indikatoren für die intellektuelle Leistungsfähigkeit liegen für die achte Klasse Daten des *Middle Years Information System* (MidYIS, Durham University) und für die neunte Klasse Daten des *Cognitive Abilities Tests* (CAT, Thorndike & Hagen, 1993) vor. Das *MidYIS* besteht aus vier Untertests (*Vocabulary*, *Math*, *Non-verbal*, *Skills*) mit Aufgaben aus den Bereichen Wortschatz, Mathematik, Korrekturlesen, Informationsverarbeitungsgeschwindigkeit und Wahrnehmungsgenauigkeit (einschließlich des räumlichen Vorstellungsvermögens und des logischen Denkens, Quelle: www.midysisproject.org). Der CAT dient der Erfassung von Reasoning in drei Bereichen (verbal, non-verbal, numerisch) und besteht aus drei entsprechenden Untertestbatterien, einer verbalen (4 Tests, z.B. Wortschatz), einer quantitativen (3 Tests, z.B. Fortsetzen von Zahlenfolgen) und einer non-verbalen (3 Tests, z.B. figurale Analogien). Beide Verfahren sind national normiert (UK). Die Ergebnisse werden in IQ-Punkten (100 +/- 15) angegeben und repräsentieren das Leistungsniveau einer Person verglichen mit ihrer englischen Altersstichprobe.

$N = 201$). Die unterschiedlichen Stichprobenzahlen sind durch fehlende Daten für einige Schüler zu erklären, z.B. bedingt durch einen vorausgegangen Schulwechsel. Für insgesamt 26 Schüler lagen keine Ergebnisse in den Schulleistungstests vor, für 6 Schüler lagen keine Ergebnisse in den Intelligenztests vor.

Tabelle 5.3

	Kennwerte zur intellektuellen Leistungsfähigkeit der Stichprobe					
	Gesamtstichprobe ($N = 411/417$)		Teilstichprobe 1 ($n_1 = 246/249$)		Teilstichprobe 2 ($n_2 = 165/168$)	
	\bar{x}	SD	\bar{x}	SD	\bar{x}	SD
MidYIS (8. Klasse), Gesamtscore (IQ, $N = 210$)	107.13	11.92	106.69	13.07	107.71	10.27
CAT (9. Klasse), Gesamtscore (IQ, $N = 201$)	105.14	9.25	105.3	9.47	104.86	8.9

Anmerkung: CAT = Cognitive Abilities Tests (Thorndike & Hagen, 1993); MidYIS = Middle Years Information System (Durham University); IQ = 100+/-15; Teilstichprobe 1 = Stichprobe mit vollständigen Datensätzen; Teilstichprobe 2 = Stichprobe mit unvollständigen Datensätzen

Tabelle 5.4

		Kennwerte zur schulischen Leistungsfähigkeit der Stichprobe					
		Gesamtstichprobe ($N = 391/417$)		Teilstichprobe ($n_1 = 239/249$)		Teilstichprobe ($n_2 = 152/168$)	
		Median	Interquartilbereich	Median	Interquartilbereich	Median	Interquartilbereich
KS2 SAT (8. Klasse, $N = 190$)	Englisch	4.0	4.0-5.0	4.0	4.0-5.0	4.0	4.0-5.0
	Mathematik	4.0	4.0-5.0	4.0	4.0-5.0	4.0	4.0-5.0
	Science	4.0	4.0-5.0	4.0	4.0-5.0	5.0	4.0-5.0
KS3 SAT (9. Klasse, $N = 201$)	Englisch	5.5	5.0-6.0	5.5	5.0-6.0	5.5	5.0-5.5
	Mathematik	6.0	5.5-6.5	6.0	5.5-6.5	6.0	5.5-6.5
	Science	5.5	5.0-6.0	5.5	5.0-6.0	5.5	5.0-6.0

Anmerkung: KS2 = National Curriculum Key Stage 2 Test; KS3 = National Curriculum Key Stage 3 Test; kriteriumsorientierte Schulleistungstests mit den Kriterien 4 (Klassenstufe 8); bzw. 5 (Klassenstufe 9)

Die Gruppenunterschiede in den Untertests KS2 Science der achten Klassenstufe sowie KS3 English der neunten Klassenstufe sind statistisch nicht signifikant ($\alpha = 5\%$; $U_{KS2, Science} = 3808.0$, $p = .092$, $N = 190$; $U_{KS3, English} = 3973.0$, $p = .06$, $N = 201$). Die hier betrachteten Schulleistungstests stellen lehrzielorientierte Tests dar. In allen Substichproben wurde von mindestens 75% der Schüler (siehe Interquartilbereich) das Kriterium bzw. Klassenziel erreicht (achte Klasse: 4; neunte Klasse: 5; Tabelle 5.4).

Die endgültige Stichprobe mit vollständigen Datensätzen setzt sich somit aus 249 Schülern der siebten (48.2%) und achten Klassen im Alter von 13 bis 15 Jahren zusam-

men (Tabelle 5.5). Insgesamt waren 47.8% der Versuchspersonen weiblich. Die gezogene Stichprobe kann hinsichtlich der Kennwerte zur intellektuellen Leistungsfähigkeit als repräsentativ angesehen werden (Tabelle 5.3). Auch die Schulleistungen der Stichprobe in den Fächern Englisch, Mathematik und Science entsprechen dem Erwartungswert des national festgelegten Klassenziels der jeweiligen Klassenstufe (Tabelle 5.4). Die gewählte Stichprobe ist also hinsichtlich ihrer schulischen Leistungsfähigkeit sowie ihrer intellektuellen Leistungsfähigkeit repräsentativ für die Klassenstufen acht und neun, bzw. für den Altersbereich von 13 bis 15 Jahren der englischen Grundgesamtheit.

Tabelle 5.5

Klassenstufen- und Geschlechtszugehörigkeit der Untersuchungsstichprobe

		<i>Klassenstufe</i>		<i>Summe</i>
		<i>8. Klasse</i>	<i>9. Klasse</i>	
Geschlecht	männlich	64	66	130
	weiblich	56	63	119
Summe		120	129	249

Tabelle 5.6 zeigt die Zuweisung der Schüler auf die experimentellen Gruppen. Es werden nur vollständige Datensätze angegeben ($N = 249$). Zur Prüfung von Feedback-Effekten sollten die Daten der Versuchsgruppen herangezogen werden, die zum zweiten Messzeitpunkt die Aufgaben unter Feedback zu bearbeiten hatten. Vor dem Hintergrund limitierter Ressourcen wurden im Sinne der Maximierung der statistischen Power der anzustellenden Analysen die Kontrollgruppen prozentual niedriger besetzt. Auf die in Tabelle 5.6 aufgeführte Kennzeichnung der einzelnen Versuchsgruppen wird später zurückgegriffen.

Tabelle 5.6

Versuchsplan (Leistungstestung)

Materialbereich	Versuchsgruppe	Sitzung 1	Experimentelle Bedingung	Sitzung 3	Experimentelle Bedingung	Prozent (geplant)	N = 249 (realisiert)
figural (4 Testvarianten)	VG1	Odd-	101	Even+	201	12.5	37
		Even-	102	Odd+	202	12.5	38
	VG2	Odd+	103	Even-	203	6.25	17
		Even+	104	Odd-	204	6.25	19
	KG	Odd-	105	Even-	205	6.25	16
		Even-	106	Odd-	206	6.25	16
numerisch (4 Testvarianten)	VG1	Odd-	107	Even+	207	12.5	28
		Even-	108	Odd+	208	12.5	23
	VG2	Odd+	109	Even-	209	6.25	15
		Even+	110	Odd-	210	6.25	13
	KG	Odd-	111	Even-	211	6.25	13
		Even-	112	Odd-	212	6.25	14

Anmerkung: abhängige Variablen: Antwortqualität (Anzahl Richtige)
 Odd = ungerade Items
 Even = gerade Items
 - = kein Feedback
 + = Feedback nach jeder Antwort des Testanden
 VG1 = Versuchsgruppe 1
 VG2 = Versuchsgruppe 1
 KG = Kontrollgruppe

5.3 Prozedur

Die Datenerhebung erfolgte in jeweils drei Gruppensitzungen innerhalb einer Woche. Alle Testsitzungen konnten während der regulären Unterrichtszeit durchgeführt werden. Die Schulleitung stellte zu diesem Zweck die schuleigenen Computerlabore zur Verfügung, in denen jeweils die gesamte Klasse getestet wurde. Neben dem Versuchsleiter war während den Sitzungen ein Fachlehrer anwesend.

Die Versuchspersonen wurden wie im Versuchsplan vorgesehen insgesamt 12 experimentellen Gruppen nach folgender Prozedur zugeteilt. Der Versuchsleiter legte vor Testbeginn die experimentelle Bedingung je Sitzplatz im Computerlabor fest. Die Schüler wählten zu Beginn der Testsitzung ihre Arbeitsplätze selbst aus. Auf diese Weise konnte unter anderem gesichert werden, dass Sitznachbarn verschiedenen experimentellen Gruppen zugehörten und Reasoning-Aufgaben aus unterschiedlichen Materialbereichen bearbeiteten.

Die Reasoning-Testung erfolgte jeweils in der ersten und dritten Sitzung (je 60 Minuten). Jede Sitzung wurde durch den Versuchsleiter standardisiert eingeleitet. Unmittelbar vor und nach der eigentlichen Reasoning-Testung hatten die Versuchsperso-

nen acht Fragen zur aktuellen Testangst, zur Zielorientierung, zur Anstrengungsvermeidung und zum aktuellen Begabungskonzept zu beantworten. Dazu wurde ihnen je ein Beispielitem von geringer sowie hoher Komplexität aus der Parallelvariante des nachfolgenden Leistungstests am Bildschirm präsentiert. Danach bearbeiteten alle Versuchspersonen zwei Übungsaufgaben, um den Umgang mit dem Programm zu trainieren. Unter der Feedback-Bedingung wurden den Versuchspersonen nach der Leistungstestung zusätzlich drei Fragen zur Einschätzung der Nützlichkeit des erhaltenen Feedbacks vorgegeben. Die erste und dritte Sitzung unterschieden sich im Ablauf nicht. Jedoch wurde in der dritten Sitzung jeweils die Parallelversion des in der ersten Sitzung vorgegebenen Leistungstests administriert.

Die zweite Sitzung (45 Minuten) diente der Erfassung der Testangst, der Zielorientierung, der Anstrengungsvermeidung, des Begabungskonzepts und der Attribution von Erfolg/Misserfolg unter einer generellen Perspektive bezogen auf das Erleben im schulischen Kontext. Der dazu vorgelegte Fragebogen umfasste 52 Items.

Die Lehrer ($N = 18$) beantworteten im Juli 2002 eine fünfstufige Skala zur Einschätzung der Leistungsmotivation jedes ihrer an der Untersuchung teilnehmenden Schülers.

Zusätzliche Daten

Es lagen weiterhin aus vorherigen am Macmillan College routinemäßig durchgeführten Untersuchungen Daten zu

- Intelligenzleistungen (*Cognitive Abilities Test [CAT]*, Thorndike & Hagen, 1993; *Middle Years Information System [MidYIS]*, University of Durham, 2000),
- Schulnoten (Examensleistungen in allen Fächern sowie Schuljahresendnoten),
- Teilleistungen (*Non Verbal Reasoning, NFER Entrance Test, National Foundation for Educational Research*, 1998),
- schulnotenunabhängige Einschätzungen der Schulleistungen in Mathematik und Englisch (*Key Stage 2, National Foundation for Educational Research*, 2001, Schulleistungstest),
- Lese-Rechtschreibfähigkeiten unabhängig von Schulnoten (*Neal Accuracy and Comprehension Test and Spelling Age*, Lese- und Rechtschreibtest)

der Schüler vor. Allerdings waren die zusätzlichen Daten nicht immer vollständig für alle Schüler vorhanden.

5.4 Messinstrumente

5.4.1 Leistungsdiagnostik

Zur Messung der intellektuellen Leistungsfähigkeit wurden Aufgaben zum *Reasoning* als Kernfaktor der Intelligenz aus zwei Merkmalsbereichen (figural, numerisch) vorgegeben. Die aus dem Itempool der ACIL stammenden Aufgaben lassen sich hinsichtlich ihrer objektiven Komplexität beschreiben. Der Itempool ist demnach in drei (Figurenfolgen, ADAFI) bzw. vier (Zahlenfolgen, AZAFO) aufeinander aufbauende Schwierigkeitsgruppen, so genannte Komplexitätsbereiche, aufgeteilt (Beckmann, 2001). Untersuchungen zur psychometrischen Qualität der Tests ergaben gute bis sehr gute Itemkennwerte und Güteabschätzungen (Beckmann & Guthke, 1999). Da die Itemdarbietung aber von der ursprünglichen Teststrategie abweicht, z.B. nicht adaptiv erfolgte, wird im nächsten Abschnitt nochmals eine in diesem Rahmen mögliche Abschätzung ihrer psychometrischen Eigenschaften vorgenommen.

Alle Reasoning-Items wurden in ansteigender Komplexität computergestützt dargeboten. Abweichend von der ursprünglich in Lerntests realisierten Bestimmung der Testleistung (siehe Abschnitt 3.1.1) wurde hier die Testleistung als Anzahl richtiger Lösungen operationalisiert.

5.4.1.1 Figurenfolgen

Der Figurenfolgen-Test besteht aus 12 Items mit fünfstufigem Multiple-Choice-Antwortformat. Abbildung 5.1 zeigt ein Beispielitem. Die im Original realisierte Farbigkeit der Elemente wurde hier durch verschiedene Muster ersetzt. Im ersten der drei Komplexitätsbereiche des Itempools variiert nur eine Dimension (Farbe oder Form), im zweiten Komplexitätsbereich variieren beide Dimensionen (Farbe und Form), und im letzten und dritten Komplexitätsbereich ist die Komplexität der darin enthaltenen Aufgaben durch die systematische Variation der drei Dimensionen Farbe, Form und Gestalt determiniert. Das in Abbildung 5.1 illustrierte Item stellt folglich ein Item aus dem zweiten Komplexitätsbereich dar. Der Testand muss den Wechsel der Farbe und der Form (Orientierung von Kreuz und Rechteck) beachten, um das Item korrekt (Antwortalternative 3) zu lösen.

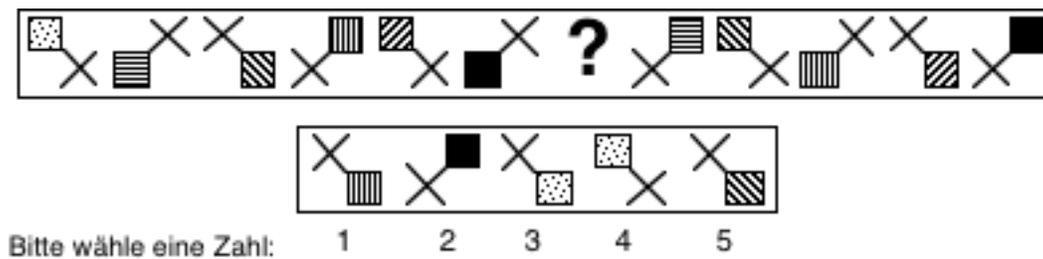


Abbildung 5.1 Beispielitem aus dem Figurenfolgen-Test (nach Beckmann, 2001, S. 82).

Zur Überprüfung der Verteilungseigenschaften, Aufgabenkennwerte und Reliabilität der so zusammengestellten Tests werden zunächst die Testrohwerte der zum ersten Testzeitpunkt eingesetzten feedbackfreien Testversionen betrachtet (Versuchsgruppen 101, 102, 105, 106; $N_{\text{gesamt}} = 146$; siehe Tabelle 5.6).

Die Testrohwerte sind normal verteilt (ungerader Itempool: $\bar{x} = 7.6$, $s = 2.1$, $Z = .985$, $p = .286$, $N = 75$; gerader Itempool: $\bar{x} = 8.0$, $s = 1.9$, $Z = 1.153$, $p = .140$, $N = 71$). Insgesamt sind die Items jedoch für die vorliegende Stichprobe etwas zu leicht (gerader Itempool: $\bar{p}_i = .67$ bzw. hinsichtlich möglicher Rateeffekte bei einem fünfstufigen Antwortformat korrigiert $\bar{p}_i = .59$; ungerader Itempool: $\bar{p}_i = .64$, bzw. hinsichtlich möglicher Rateeffekte bei einem fünfstufigen Antwortformat korrigiert $\bar{p}_i = .55$). Die zum Teil extremen Itemschwierigkeiten bedingen mäßige *Trennschärfen*¹³ (gerader Itempool: $.05 \leq r_{it} \leq .42$; ungerader Itempool: $.07 \leq r_{it} \leq .44$). Die weite Streuung der Itemschwierigkeiten, die für schwierigkeitsgestaffelte Tests erwünscht ist, führt außerdem zu geringen Iteminterkorrelationen. Anhand der Iteminterkorrelationen müssen daher beide Tests als wenig homogen bewertet werden (gerader Itempool: $\bar{r}_{ii} = .10$; ungeraden Itempool: $\bar{r}_{ii} = .10$). Weitere Analysen ergeben aber parallele Verläufe der Itemschwierigkeiten innerhalb von nach Geschlecht oder nach Klassenstufen gebildeten Teilgruppen (Teilgruppenkonstanz), so dass trotz geringer Iteminterkorrelationen Eindimensionalität¹⁴ der figuralen Tests angenommen werden kann. Infolge der geringen Iteminterkorrela-

¹³ Die punkt-biseriale Korrelation entspricht der Produkt-Moment-Korrelation, wenn in die Gleichung der Produkt-Moment-Korrelation für das dichotome Merkmal (hier Itemlösung) die Werte 0 und 1 eingesetzt werden. Der Korrelationskoeffizient wird dann noch um den doppelten Eingang des Items in die Korrelation zwischen Item und Testscore korrigiert (Teil-Ganzes-Korrektur).

¹⁴ Die Zusammenfassung der Itemlösungen zu einem Gesamtscore ist bei geringer Homogenität des Verfahrens problematisch. Aufgrund der gefundenen Teilgruppenkonstanz scheint es dennoch gerechtfertigt, einen Gesamtscore zu bilden. Interpretationen dieser Gesamtscores sind jedoch mit angemessener Vorsicht vorzunehmen.

tionen lässt sich auch eine nur mäßige interne *Konsistenz* der Verfahren ausmachen (gerader Itempool: Cronbachs $\alpha = .56$, bzw. ungerader Itempool: Cronbachs $\alpha = .58$). Die Paralleltest-Reliabilitäten, die in der Kontrollgruppe erfasst wurden, welche weder in der ersten noch in der Paralleltestung Feedback erhielt, liegen etwas höher (gerader Itempool: $r_{tAtB} = .62$, ungerader Itempool: $r_{tAtB} = .67$). Demnach lassen sich durch die Kenntnis der Leistung im ersten Test 38% bzw. 45% der Leistungsvarianz im Paralleltest vorhersagen. Die durchschnittliche Leistung (Anzahl Richtige) in beiden Paralleltests unterscheidet sich nicht signifikant ($t_{144} = -1.08$, $p = .282$, $N_{ungerade} = 75$, $N_{gerade} = 71$). Da die Teilung der Tests in zwei anforderungsparallele Testvarianten (Odd/Even-Technik) demnach gelungen ist, können die Versuchsgruppen 101 und 102 (VG1), 103 und 104 (VG2) sowie 105 und 106 (KG), wie geplant, zu jeweils einer Substichprobe zusammengefasst werden (siehe Tabelle 5.6).

5.4.1.2 Zahlenfolgen

Der Zahlenfolgen-Test besteht aus 12 Items mit freiem Antwortformat. Abbildung 5.2 zeigt ein Beispielitem.

1 2 4 8 16 32 64 ?

Abbildung 5.2 Beispielitem aus dem Zahlenfolgen-Test.

Die Aufgaben beinhalten arithmetische sowie geometrische Zahlenfolgen, die vier Komplexitätsbereichen zugeordnet werden. Diese sind durch die Art der Operationen (arithmetische oder geometrische) sowie die Anzahl der Lösungsebenen definiert. Die Aufbauregel der Zahlenfolge kann also entweder bereits auf der ersten Ebene der Folge bestimmt werden, oder ist erst unter Einbezug einer weiteren Ebene möglich (siehe Abbildung 5.3).

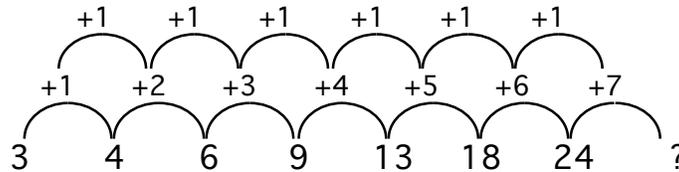


Abbildung 5.3 Beispiel für eine Zahlenfolge mit zwei Lösungsebenen aus dem zweiten Komplexitätsbereich (nach Beckmann, 2001, S. 90).

Zur Überprüfung der Verteilungseigenschaften, Aufgabenkennwerte und Reliabilität wurden wieder die Testrohwerte der zum Testzeitpunkt 1 eingesetzten feedback-freien Testversionen herangezogen (Versuchsgruppe 107, 108, 111, 112; $N_{gesamt} = 114$; siehe Tabelle 5.6).

Die Testrohwerte sind normal verteilt (gerader Itempool: $\bar{x} = 7.2$, $s = 2.9$, $Z = .787$, $p = .565$, $N = 54$; ungerader Itempool: $\bar{x} = 7.9$, $s = 2.6$, $Z = 1.064$, $p = .208$, $N = 60$). Die Items sind auch hier tendenziell etwas zu leicht (gerader Itempool $\bar{p}_i = .60$; ungerader Itempool, $\bar{p}_i = .62$). Aufgrund der teilweise extremen Itemschwierigkeiten sind keine sehr hohen *Trennschärfen* zu erwarten. Die berechneten Trennschärfen liegen überwiegend im akzeptablen Bereich (gerader Itempool: $.23 \leq r_{it} \leq .59$; ungerader Itempool: $.22 \leq r_{it} \leq .61$). Die Streuung der Itemschwierigkeiten innerhalb der Tests (gerader Itempool: $.26 \leq p_i \leq .93$; ungerader Itempool: $.18 \leq p_i \leq .90$) bedingt auch hier geringe Iteminterkorrelationen, die aber noch ausreichend sind, so dass beide Itemgruppen als *homogen* betrachtet werden können (gerader Itempool: $\bar{r}_{ii} = .26$; ungerader Itempool: $\bar{r}_{ii} = .25$). Die interne *Konsistenz* ist auch unter Beachtung der geringen Anzahl an Items akzeptabel (gerader Itempool: Cronbachs $\alpha = .80$; ungerader Itempool: Cronbachs $\alpha = .78$). Gleiches gilt für die *Paralleltest-Reliabilität*, geschätzt in der Kontrollgruppe (ungerader Itempool $r_{tAtB} = .89$; gerader Itempool $r_{tAtB} = .81$). Demnach lassen sich durch die Kenntnis der Leistung im ersten Test 79 % bzw. 66 % der Leistungsvarianz im Paralleltest vorhersagen. Die durchschnittliche Leistung – operationalisiert als Anzahl Richtige – in beiden Paralleltests unterscheidet sich nicht signifikant ($t_{144} = 1.254$, $p = .212$, $N_{ungerade} = 60$, $N_{gerade} = 54$). Da die Teilung der Tests in zwei anforderungsspezifische Testvarianten (Odd/Even-Technik) demnach gelungen ist, können die Versuchsgruppen 107 und 108 (VG1), 109 und 110 (VG2) sowie 111 und 112 (KG), wie geplant, zu jeweils einer Substichprobe zusammengefasst werden (siehe Tabelle 5.6).

Zusammenfassung: Die psychometrische Qualität der für die Itemauswahl zugrunde gelegten zwei Untertests der ACIL, AZAFO und ADAFI (Abschnitt 3.1.1), stellte sich bisher in anderen Untersuchungen anhand anderer Stichproben ($N_{ADAFI} = 885$; $N_{AZAFO} = 692$) als gut bis sehr gut heraus (Beckmann & Guthke, 1999). Deshalb schien es sinnvoll, diese bereits bewährten Items auch in der vorliegenden Studie zu verwenden. Abweichend von der ursprünglichen Teststrategie wurden die Items jetzt aber zum einen nicht-adaptiv und zum anderen ohne Hilfeangebot vorgegeben. Da im ADAFI sowie im AZAFO je zwei anforderungsgleiche Items nacheinander dargeboten werden, schien es weiterhin nahe liegend, eine Odd/Even-Teilung der Gesamttests vorzunehmen und somit zeitlich ökonomische Parallelverfahren zu bilden, die die Kontrolle eines Retesteffekts bei Wiederholungsmessung, wie im Design geplant, erlauben.

Die an der vorliegenden relativ kleinen Stichprobe vorgenommene Itemanalyse ergab allerdings teilweise schlechtere Itemkennwerte sowie Reliabilitätsindizes als in anderen Untersuchungen berichtet. Die Veränderung der Teststrategie resultierte offensichtlich in einer Einschränkung der psychometrischen Qualität vor allem der figuralen Items. Die Verteilungscharakteristika und Itemkennwerte legen nahe, dass die Items für die hier untersuchte englische Stichprobe etwas zu leicht waren. Die sich daraus ergebene Reduktion von Varianz innerhalb der Stichprobe kann somit auch die Reliabilitätsschätzungen sowie Itemkennwerte negativ beeinflussen. Auch die vorgenommene Testverkürzung schränkt potentiell die Reliabilität (Konsistenz) ein. Aufgrund der nur mäßigen Reliabilität der Verfahren zur Erfassung von *figuralem* Reasoning sind die in späteren Analysen zu prüfenden Feedback-Effekte jedoch schwieriger aufzudecken. So kann mit wenig reliablen Messinstrumenten nur der Nachweis der Existenz starker Effekte gelingen. Die psychometrische Qualität der beiden Tests zur Erfassung von Reasoning im *numerischen* Bereich ist dagegen zufrieden stellend.

Neben einer generell ausgerichteten Perspektive wurde eine differentielle Analyse der Leistungsdaten geplant. Die dazu eingesetzten Fragebogenverfahren werden im Folgenden vorgestellt und ihre psychometrische Qualität bestimmt.

5.4.2 Persönlichkeitsdiagnostik

Zur Erfassung der in dieser Studie einbezogenen außerintellektuellen Personmerkmale war die Zusammenstellung eines persönlichkeitsdiagnostischen Verfahrens erforderlich. Dazu wurden zum einen verschiedene publizierte Verfahren im englischen Sprachbereich zu den Merkmalen *Testangst* (Meijer, 2001), *Zielorientierung* (Duda & Nicholls, 1992

sowie Dweck, 2000), *Begabungskonzept* (Dweck, 2000), *Anstrengungsvermeidung* (Duda & Nicholls, 1992) sowie *Ursachenerklärungen* für Erfolg bzw. Misserfolg (Duda & Nicholls, 1992) herangezogen und teilweise modifiziert (siehe Anlage). Unter anderem galt es das Antwortformat der insgesamt 52 Items zu vereinheitlichen. Zur Abhängigkeit des Ausmaßes der Zustimmung oder Ablehnung der im jeweiligen Item getroffenen Aussage sollte vom Testanden eine Markierung auf einer sieben Zentimeter langen Analogskala vorgenommen werden. Abbildung 5.4 zeigt ein Beispielitem zur Veranschaulichung des im Fragebogen verwendeten Antwortformats.

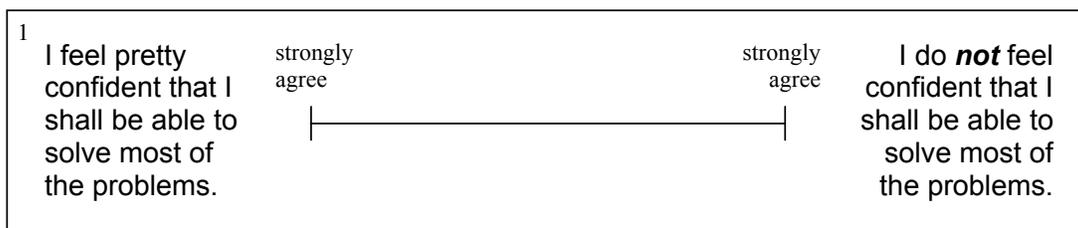


Abbildung 5.4 Beispielitem aus dem situationsspezifischen Fragebogen.

Um situations- bzw. anforderungsspezifische Informationen zu den entsprechenden Personmerkmalen zu gewinnen, wurde zum anderen ein Fragebogen entwickelt, der vor und nach der computergestützten Leistungstestung einzusetzen war. So sollte geprüft werden, ob Konstrukte wie beispielsweise das Begabungskonzept in der konkreten Testsituation tatsächlich verhaltenswirksam werden. Dazu wurden mit jeweils zwei Items situationsspezifisch *Testangst*, *Zielorientierung*, *Begabungskonzept*, *Anstrengungsvermeidung* vor als auch nach der Leistungstestung erfasst. Des Weiteren sollten die Testanden in der Feedback-Bedingung Auskunft geben, inwiefern sie das Feedback als störend, wichtig und hilfreich für die Lösung weiterer Aufgaben einschätzten (3 Items). Als Antwortformat wurde auch hier die oben beschriebene Analogskala verwendet. Die Formulierung der Items basierte auf den von Mummenday (1999) vorgeschlagenen Kriterien.

Bei der Konstruktion beider Fragebögen wurde angestrebt, möglichst maximale Vergleichbarkeit zu erreichen. So existiert für jedes Item im situationsspezifischen Fragebogen (State-Ebene) ein bis auf den Situationsbezug identisches Referenzitem im situationsunspezifischen Fragebogen (Trait-Ebene). Damit ist potentiell gewährleistet, dass Aussagen zur Konsistenz im Antwortverhalten getroffen werden können. Somit lässt sich also auch prüfen, ob auffällige Ausprägungen in den erfassten außerintellektuellen

Merkmale auf Trait-Ebene auch auf State-Ebene, d.h. in der Testsituation, zum Ausdruck kommen.

Es war notwendig, die für den amerikanischen Sprachraum entwickelten Verfahren (Duda & Nicholls, 1992; Dweck, 2000) von einem Experten (Muttersprachler mit Expertise im pädagogischen Bereich) an den britischen Sprachgebrauch anzupassen. Dabei wurden einige Items sprachlich verändert oder Items aufgrund inhaltlicher Überlegungen hinsichtlich kultureller Besonderheiten in nordenglischen Gesamtschulen eliminiert (siehe dazu die Darstellung der einzelnen Verfahren). Aus pragmatischen Gründen durfte die Bearbeitungszeit für das schulbezogene Fragebogenverfahren zudem 40 Minuten nicht überschreiten. Mit anderen Worten musste ein Kompromiss zwischen zuverlässigkeitsdienlicher Testlänge und pragmatischer Machbarkeit gefunden werden. Für die Itemselektion wurden – soweit vorhanden – die in entsprechenden Publikationen referierten Itemkennwerte einbezogen. Zur Gewöhnung an das hier genutzte, eher ungewöhnliche Antwortformat (Analogskala) wurde allen Fragebögen ein erklärendes Beispielitem vorangestellt.

Da die zu verwendenden persönlichkeitsdiagnostischen Instrumente also mehr oder weniger stark für den Zweck der hier vorliegenden Untersuchung modifiziert wurden, wird im Folgenden die Prüfung der psychometrischen Qualität der eingesetzten Verfahren dargestellt. Als Basis für die Schätzung der Itemkennwerte der situationspezifischen Items diente das Antwortverhalten der gesamten Stichprobe ($N = 410$) vor der ersten Reasoning-Testung. Alle Items wurden entsprechend ihrer Polung so verrechnet, dass hohe Skalenwerte einer hohen Merkmalsausprägung entsprechen.

5.4.2.1 Zielorientierung

Ein Ziel bestand darin, teilweise gegensätzliche in der Literatur beschriebene theoretische Vorstellung zur Dimensionalität des Konstrukts *Zielorientierung* (auch motivationale Orientierung) als auch seine psychodiagnostische Umsetzung zu prüfen und das hinsichtlich seiner Kennwerte angemessenere Verfahren für die Analyse der Fragestellung heranzuziehen. Aus diesem Grund fanden sowohl die Arbeiten nach Nicholls und Mitarbeiter (beispielsweise Duda & Nicholls, 1992), die von einem mehrdimensionalen Konstrukt ausgehen, als auch nach Dweck (2000), die ein eindimensionales Konstrukt präferiert, in beiden persönlichkeitsdiagnostischen Instrumenten Berücksichtigung (siehe Abschnitt 2.2.5).

5.4.2.1.1 Motivationale Orientierung

Die MOS (*Motivational Orientation Scales*, Duda & Nicholls, 1992) dienen der Erfassung der im schulischen Kontext verfolgten Zielorientierungen nach Nicholls (schon 1984), nämlich der Dimensionen Ich-Orientierung und Aufgaben-Orientierung. Daneben wird mit drei zusätzlichen Items Anstrengungsvermeidung erfasst, da es sich hierbei um ein wesentliches Kernelement des Konstrukts Zielorientierung handelt (siehe Abschnitt 2.2.5; Duda & Nicholls, 1992; Nicholls et al., 1990). Die hier verwendete Version der MOS (Duda & Nicholls, 1992) wurde für den Geltungsbereich amerikanische *High-School*-Schüler (von 14 bis 18) konstruiert. Der dreidimensionale Fragebogen besteht aus 19 Items mit 5 stufigem Antwortformat vom Typ einer Likert-Skala („strongly agree“ bis „strongly disagree“), die den Skalen „Aufgaben-Orientierung“ (*Task-Orientation*), „Ich-Orientierung“ (*Ego-Orientation*) und „Anstrengungsvermeidung“ (*Effort-Avoidance*) zugeordnet sind. Der Itemstamm jedes Items zielt auf persönliche Erfolgs-Kriterien ab. Der Schüler ist angehalten darüber nachzudenken, unter welchen Bedingungen er hinsichtlich seiner schulischen Leistungen Erfolg verspürt („I feel really successful when ...“).

Im Rahmen der Anpassung des Verfahrens an den britischen Sprachgebrauch wurden drei Items sprachlich verändert und auf ein Item aus inhaltlichen Gründen ganz verzichtet („I feel really successful when I can goof off.“). Somit gingen 18 Items aus den MOS in das hier verwendete Instrument ein. Das Antwortformat wurde an die oben dargestellte Analogskala angepasst. Die 18 Items wurden nach der Zugehörigkeit zu den Skalen alternierend dargeboten (siehe Anlage).

Die Prüfung der *Verteilungseigenschaften* erlaubt die Annahme einer Normalverteilung der Summenrohwerte der Subskalen „Ich-Orientierung“ ($\bar{x} = 2.83$; $s = 1.50$; $Z = 1.009$, $p = .26$, $N = 413$) und „Anstrengungsvermeidung“ ($\bar{x} = 3.97$; $s = 1.59$, $Z = .807$, $p = .532$, $N = 412$). Die Summenrohwerte der Subskala „Aufgaben-Orientierung“ sind nicht normal verteilt ($\bar{x} = 1.57$; $s = 1.01$, $Z = 1.498$, $p = .022$, $N = 413$), sondern zeigen eine rechtssteile Verteilung ($sch = -.996$, $SE_{sch} = .12$). Die Items der Subskala „Aufgaben-Orientierung“ wurden demnach von der Mehrheit der Untersuchungsstichprobe in Richtung hoher Merkmalsausprägung beantwortet, sind also für die hier untersuchte Stichprobe zu leicht. Da die Antworthäufigkeiten jedes der acht Items dieser Subskala rechtssteil verteilt sind, ist auch eine Eliminierung einzelner Items nicht sinnvoll. Damit kann eine Interpretation der Skala „Aufgaben-Orientierung“ (Skalenmittelwerte) nur mit Vorsicht vorgenommen werden. Es lässt sich allerdings diskutieren, ob eine Normal-

verteilung des Merkmals Aufgaben-Orientierung überhaupt erwartet werden kann (siehe Abschnitt 2.2.5).

Im Sinne einer groben Abschätzung der psychometrischen Eigenschaften des Verfahrens sollen im Folgenden die Ergebnisse der *Dimensionalitätsprüfung* mit Hilfe der Faktorenanalyse dargestellt werden. Die Hauptkomponenten-Analyse mit Extraktion und anschließender Varimax-Rotation der Faktoren mit einem Eigenwert > 1 ergab, wie erwartet, eine Drei-Faktoren-Lösung (Kriterium: Scree-Test, 60.01% aufgeklärte Varianz, $N = 408$, siehe Tabelle 5.7). Die Betrachtung der Faktorladungen zeigt, dass die Items einer Subskala der MOS auch in der hier untersuchten Stichprobe erwartungsgemäß interkorrelieren und einen Faktor bilden. Dabei liegen die entsprechenden Ladungen ($.63 \leq a_{ij} \leq .81$) über dem Kriterium zur Abschätzung der Bedeutsamkeit von Faktorladungen ($a_{ij} = .60$). Entsprechend der Einfachstruktur finden sich niedrigere Faktorladungen aller Items mit den jeweils beiden nicht zugehörigen Faktoren (Nebenladungen, $.11 \leq a_{ij} \leq .34$). Eine Interpretation der varimax-rotierten Faktorstruktur ist bei Betrachtung der Ladungsmatrix gerechtfertigt (siehe Bortz, 1999). Es lassen sich demnach zwei (unabhängige) Faktoren extrahieren, die jeweils eine Zielorientierung repräsentieren (zweidimensionales Modell). Eine dritte Komponente repräsentiert die beiden Items zur „Anstrengungsvermeidung“, sollte aber aufgrund der wenigen Items und einer hohen Nebenladung ($a_{ij} > .30$) zurückhaltend interpretiert werden.

Tabelle 5.7

Kennwerte der faktoranalytischen Dimensionalitätsprüfung der MOS (Duda & Nicholls, 1992)

Itemnummer, originale Skalenzugehörigkeit	Kommunalität	Faktor 1	Faktor 2	Faktor 3
Item 18, Ich-Orientierung	.63	.81		
Item 03, Ich-Orientierung	.47	.81		
Item 14, Ich-Orientierung	.66	.80		
Item 16, Ich-Orientierung	.64	.80		
Item 01, Ich-Orientierung	.51	.79		
Item 12, Ich-Orientierung	.66	.75		.20
Item 10, Ich-Orientierung	.53	.75		.21
Item 07, Ich-Orientierung	.52	.68		.26
Item 17, Aufgaben-Orientierung	.58		.81	
Item 15, Aufgaben-Orientierung	.61		.78	
Item 13, Aufgaben-Orientierung	.43		.78	
Item 04, Aufgaben-Orientierung	.61		.75	-.28
Item 05, Aufgaben-Orientierung	.65		.71	
Item 08, Aufgaben-Orientierung	.66		.70	
Item 02, Aufgaben-Orientierung	.64	.21	.64	
Item 11, Aufgaben-Orientierung	.67		.63	
Item 06, Anstrengungsvermeidung	.66			.81
Item 09, Anstrengungsvermeidung	.67	.34		.68
Eigenwerte der rotierten Faktoren		4.99	4.28	1.53
Durch Faktoren aufgeklärter Varianzanteil		27.74	23.79	8.48

Anmerkung: Ladungen < |.20| werden nicht dargestellt.

Nebenladungen der Items des „Faktors“ „Anstrengungsvermeidung“ auf den Faktoren der Zielorientierungen sind theoretisch nachvollziehbar (siehe Abschnitt 2.2.5)¹⁵. Die Polung ist dabei theoriekonform (siehe Item 9, Tabelle 5.7).

Die Überprüfung der *Itemschwierigkeiten* ergab angemessene Kennwerte ($.42 \leq p_i \leq .84$). Insgesamt zeigt sich aber auch hier, dass insbesondere die Items der Subskala „Aufgaben-Orientierung“ ($\bar{p}_i = .77$) von eher geringer Schwierigkeit sind, also von einer Mehrheit der Testanden in Richtung hoher Merkmalsausprägung beantwortet wurden (siehe Tabelle 5.8 und Tabelle 5.9). Die durchschnittliche Itemschwierigkeit der Subskala „Ich-Orientierung“ ist akzeptabel ($\bar{p}_i = .60$).

In die Analyse der *Trennschärfe* der Items wurden die Skalen „Ich-Orientierung“ und „Aufgaben-Orientierung“ einbezogen. Auf die Bestimmung der Trennschärfe für

¹⁵ *Rotation der Faktorenstruktur*: Duda und Nicholls (1992) haben eine schiefwinklige (Oblimin) Rotation der Faktoren vorgeschlagen. Aufgrund der Interpretationseindeutigkeit unabhängiger Faktoren wurde hier aber einer rechtwinkligen Rotation (Varimax) der Faktoren der Vorzug gegeben.

die Skala „Anstrengungsvermeidung“ wurde verzichtet, da sie aus nur zwei Items besteht, und stattdessen die Item-Interkorrelation bestimmt. Bei simultaner Betrachtung der Trennschärfen als auch der Fremdtrennschärfen für beide Subskalen lässt sich erkennen, dass alle Items jeweils höher mit dem eigenen Skalensummenwert (Trennschärfe) als mit den Skalensummenwerten der beiden anderen Skalen (Fremdtrennschärfe) korreliert sind. So sind alle Trennschärfekoeffizienten als hoch zu bewerten ($r_{ii} > .50$, Bortz & Döring, 1995), während ein Großteil der Fremdtrennschärfen unter dem vorgeschlagenen Kriterium ($r_{ii} \leq .30$) bleibt. Auffallend sind fünf Items der Skala „Ich-Orientierung“, die positive Fremdtrennschärfen im Bereich von $.29 \leq r_{ii} \leq .30$ zu den beiden Items der Skala „Anstrengungsvermeidung“ aufweisen. Ein moderater Zusammenhang zwischen den Items der Skalen „Ich-Orientierung“ und „Anstrengungsvermeidung“ ist inhaltlich nachvollziehbar und zeigt sich auch bereits in der Dimensionalitätsprüfung (siehe oben).

Die *Homogenität* der Skalen „Aufgaben-Orientierung“ und „Ich-Orientierung“ liegt mit $\bar{r}_{ii} = .55$ (Ich-Orientierung) bzw. $\bar{r}_{ii} = .46$ (Aufgaben-Orientierung) etwas über dem von Briggs und Check (1986) vorgeschlagenen Bereich ($.20 \leq \bar{r}_{ii} \leq .40$; Tabelle 5.8).

Als ein Maß für die Bestimmung der *Reliabilität* der Skalen „Aufgaben-Orientierung“ und „Ich-Orientierung“ wurde der Alpha-Koeffizient nach Cronbach (1951) herangezogen. Beide Skalen weisen eine mäßige bis hohe interne Konsistenz auf (Skala „Aufgaben-Orientierung“: $\alpha = .87$; Skala „Ich-Orientierung“: $\alpha = .91$, siehe Konventionen nach Lienert & Ratz, 1994). Da der Alpha-Koeffizient den auf eine Merkmalsdimension zurückgehenden Varianzanteil aller Items einer Skala schätzt, lässt sich an beiden Werten ebenso auf das Ausmaß der *Homogenität* der entsprechenden Subskalen schließen (Homogenitätsindex, Bortz & Döring, 1995). Der Zusammenhang zwischen beiden Items der Skala „Anstrengungsvermeidung“ ist signifikant (Spearman's Rangkorrelation $r_s = .41$, $p < .001$, $N = 412$), so dass angenommen werden kann, dass beide Items neben einer Fehlerkomponente systematische Varianz eines gemeinsamen Merkmals erfassen.

Tabelle 5.8

Item- und Skalenanalyse der Skalen „Ich-Orientierung“ und „Aufgaben-Orientierung“ der MOS (Duda & Nicholls, 1992)

Itemnummer, Skalenzugehörigkeit	Schwierig- keit	Trenn- schärfe	Fremdtrennschärfe: Zielorientierung	Fremdtrennschärfe: Anstrengungsver- meidung	Homogenität (Spearman's r_s)
01, Ich-Orientierung	.64	.68	.12	.19	.53
03, Ich-Orientierung	.70	.70	.05	.23	.54
07, Ich-Orientierung	.54	.64	.10	.31	.50
10, Ich-Orientierung	.64	.70	.13	.32	.54
12, Ich-Orientierung	.51	.70	-.01	.30	.54
14, Ich-Orientierung	.59	.75	.05	.29	.57
16, Ich-Orientierung	.55	.75	.12	.29	.57
18, Ich-Orientierung	.59	.74	.16	.22	.57
02, Aufgaben-Orientierung	.82	.56	.20	-.04	.41
04, Aufgaben-Orientierung	.80	.67	.08	-.20	.51
05, Aufgaben-Orientierung	.69	.61	.05	-.06	.44
08, Aufgaben-Orientierung	.84	.61	.05	-.11	.50
11, Aufgaben-Orientierung	.76	.52	.12	-.01	.45
13, Aufgaben-Orientierung	.82	.71	.11	-.16	.55
15, Aufgaben-Orientierung	.75	.68	.03	-.04	.54
17, Aufgaben-Orientierung	.73	.70	.06	-.06	.54

Anmerkung: Die Fremdtrennschärfe der Items einer Skala der Zielorientierung bezieht sich jeweils auf die andere Skala der Zielorientierung

Tabelle 5.9

Itemschwierigkeit und Fremdtrennschärfe der Skala „Anstrengungsvermeidung“ der MOS (Duda & Nicholls, 1992)

Itemnummer, Skalenzugehörigkeit	Schwierigkeit	Fremdtrennschärfe: Leistungszielorientierung	Fremdtrennschärfe: Anstrengungsvermeidung
06, Anstrengungsvermeidung	.45	.21	-.10
09, Anstrengungsvermeidung	.42	.38	-.09

Die im Rahmen der Item- und Skalenanalyse bestimmten Kennwerte entsprechen im Wesentlichen den bisher publizierten Befunden (Duda & Nicholls, 1992). Die im Interesse der hier vorgelegten Studie vorgenommene Modifizierung hatte somit keinen negativen Einfluss auf die psychometrische Qualität des Instruments. Zusammenfassend ergeben sich hinsichtlich der Reliabilität und faktoriellen Validität zufrieden stellende bis gute Kennwerte der MOS, die eine Interpretation der Testwerte zulassen. Lediglich die Subskala „Aufgaben-Orientierung“ stellte sich als problematisch dar.

5.4.2.1.2 Situationsspezifische Aufgaben- versus Ich-Orientierung

In Anlehnung an das dreidimensionale Fragebogenverfahren nach Duda und Nicholls (MOS, 1992) wurde je ein Item formuliert, welches jeweils einer Markiertvariablen der drei Faktoren inhaltlich entsprechen sollte. Die so konstruierten drei Items wurden zu beiden Testsitzungen vor der Reasoning-Testung vorgegeben, um die situationsspezifische Zielorientierung und Anstrengungsvermeidung zu erfassen. Da es aber wenig sinnvoll ist, nach der Testung, also rückblickend, Nicholls' Erfolgskriterien im Sinne von Zielen zu erfragen, ist stattdessen nach der Testung konkret eine interindividuelle (Ich-Orientierung) sowie eine intraindividuelle (Aufgaben-Orientierung) Leistungseinschätzung vom Testanden gefordert und somit die *Zielerreichung* und nicht die *Zielsetzung* erfragt worden. In gleicher Weise erfolgte die Erfassung der im Testverlauf tatsächlich investierten Anstrengung.

Die Prüfung der *Verteilungseigenschaften* ergab für die Items 3 und 4 eine rechtssteile Verteilung der Antworthäufigkeiten (Tabelle 5.10; siehe auch Anlage). Als Basis diente hier das Antwortverhalten der gesamten Stichprobe vor Beginn des ersten Reasoning-Tests und während der Konfrontation mit zwei Beispielitems, um somit eine Schätzung der Itemkennwerte unabhängig von der experimentellen Bedingungsvariation zu erhalten.

Tabelle 5.10

Verteilungskennwerte und Itemschwierigkeiten der Items zur Erfassung der situationsspezifischen motivationalen Orientierung, sowie der situationsspezifischen Anstrengungsvermeidung (in Anlehnung an Duda & Nicholls, 1992)

	Median	Interquartilbereich	Z-Wert	Schiefte	SE_{sch}	p_i	N
Item 3 (Ich-Orientierung)	4.4	3.5 - 6.0	1.639**	-.521	.121	.63	410
Item 4 (Aufgaben-Orientierung)	5.1	3.6 - 6.2	2.135**	-.809	.121	.68	410
Item 6 (Anstrengungsvermeidung)	3.6	1.6 - 5.8	2.037**	.017	.121	.51	408

Anmerkung: ** = $p < .01$

Auch die Bestimmung der *Itemschwierigkeit* zeigt, dass beide Items zur Erfassung der situationsspezifischen Zielorientierungen überwiegend in Richtung hoher Merkmalsausprägung beantwortet wurden (siehe auch die rechtssteile Verteilung). Die Mehrzahl der Testanden gab also vor dem ersten Reasoning-Test an, beide Ziele zu verfolgen, also sowohl aufgaben- als auch ich-orientiert zu sein. Dieses Ergebnis widerspricht der von Dweck (2000) vertretenen Auffassung zur Eindimensionalität des Konstruktes Zielorientierung. Die Zielorientierung stellt also offenbar kein bipolares Kon-

strukt dar. Des Weiteren zeigt sich eine neutrale Antworttendenz hinsichtlich der Anstrengungsabsicht. Die Schwierigkeitsindizes der Items 3, 4 und 6 sind aber akzeptabel (Tabelle 5.10).

Die Analyse der Zusammenhänge der Items zur Erfassung der test- also situationsspezifischen Zielorientierungen mit den Skalensummenwerten der Subskalen der MOS (Duda & Nicholls, 1992) ergab zum einen erwartungsgemäß substantielle Zusammenhänge mit der Skalensumme der jeweils entsprechenden Subskala der MOS, was als Beleg für die interne Validität der situationsspezifischen Items gelten kann. Zum anderen bestehen negative oder nicht-substantielle Zusammenhänge mit der Skalensumme der jeweils nicht zugehörigen Subskala der MOS (Tabelle 5.11). Es kann demnach im Sinne der *Validität* davon ausgegangen werden, dass die situationsspezifischen Items und die jeweils entsprechenden Subskalen der MOS gemeinsame Merkmalsvarianz erfassen. Da die Zusammenhänge aber moderat sind, kann weiter angenommen werden, dass mittels der situationsspezifischen Items auch situationsspezifische Merkmalsvarianz erhoben wird. Ebenso findet sich eine substantielle Korrelation des Items zur Erfassung der situationsspezifischen Anstrengungsvermeidung mit der entsprechenden allerdings aus nur zwei Items bestehenden „Subskala“ „Anstrengungsvermeidung“ der MOS (Spearman's $r_s = .23$, $p < .001$, $N = 401$), so dass auch hier angenommen werden kann, dass gemeinsame systematische Merkmalsvarianz erfasst wird.

Tabelle 5.11

Zusammenhänge zwischen den Items zur Erfassung der situationsspezifischen motivationalen Orientierung und den Subskalen der MOS (in Anlehnung an Duda & Nicholls, 1992)

situationsspezifische Zielorientierung	schulbezogene Zielorientierung (MOS)	
	Ich-Orientierung, Skalensumme	Aufgaben-Orientierung, Skalensumme
Item 3 (Ich-Orientierung)	.28**	.07
Item 4 (Aufgaben-Orientierung)	-.13**	.36**

Anmerkung: Spearman's r_s , $N = 401$, ** = $p < .01$

Sowohl die itemanalytischen Kennwerte als auch die validitätsbezogenen Zusammenhangsparameter lassen eine Interpretation der Itemwerte als Indikatoren der test- bzw. situationsspezifischen motivationalen Orientierung als gerechtfertigt erscheinen.

5.4.2.1.3 Lern- versus Leistungszielorientierung

Die Erfassung der schulbezogenen Zielorientierung als eindimensionales, bipolares Konstrukt orientierte sich an einem Verfahren nach Dweck (*Goal Choice Questionnaire*, Dweck, 2000). Das aus vier Items mit sechsstufiger Likert-Skala („strongly agree“ bis „strongly disagree“) bestehende Fragebogenverfahren ist für den Geltungsbereich ab 12 Jahre konzipiert. Die Besonderheit der Items zur Erfassung der im schulischen Kontext verfolgten Ziele, auch im Gegensatz zu den oben beschriebenen Items zur motivationalen Orientierung (Nicholls, 1984), besteht in der Gegenüberstellung von Lern- und Leistungszielen. Dabei wird vom Rezipienten eine Entscheidung für eine der beiden Zielkategorien gefordert (Eindimensionalität, Dweck 2000)¹⁶. Im Sinne des hier angestrebten explorativen Herangehens wurden zwei für die vorliegende Untersuchung geeignete Items, je eines zur Leistungszielorientierung und eines zur Lernzielorientierung, ausgewählt und an das hier gewünschte Antwortformat in Form einer Analogskala angepasst (siehe Anlage). Diese Items sind ebenfalls hinsichtlich ihrer Angemessenheit für den britischen Sprachgebrauch überprüft und leicht angepasst worden.

Für beide Items kann keine Normalverteilung angenommen werden (Tabelle 5.12). Es fällt weiter auf, dass das Item zur Erfassung der Lernzielorientierung (Item 50) geringe *Itemschwierigkeit* ($p_i = .76$) aufweist. Dies zeigt sich ebenfalls in den Verteilungseigenschaften. Die Häufigkeitsverteilung der Itemantworten ist deutlich rechtssteil. Die Mehrzahl der Testanden charakterisierte sich demnach als lernzielorientiert. Die Analyse der Itemschwierigkeiten als auch der Verteilungseigenschaften des Items zur Erfassung der Leistungszielorientierung (Item 49) zeigt dagegen, dass dieses Item von der überwiegenden Zahl der Testanden ablehnend beantwortet wurde (leicht linkssteile Verteilung, Tabelle 5.12). Die Mehrheit der Schüler beschrieb sich in Hinblick auf ihr schulisches Leistungsverhalten eher lernzielorientiert und weniger leistungszielorientiert. Die beiden Items zur Zielorientierung sind theoriekonform signifikant negativ korreliert ($r_s = -.18, p < .001, N = 413$).

¹⁶ *Eindimensionalität der Zielorientierung*: Bei Betrachtung der Items nach Dweck (2000) fällt auf, dass die Lern- sowie Leistungszielorientierung dennoch getrennt, d.h. mit verschiedenen Items erfasst wird. Eine hohe Ausprägung auf einem Item zur Erfassung der Lernzielorientierung wird jedoch nicht nur als Indikator für Lernzielorientierung aufgefasst, sondern auch als Hinweis auf geringe Leistungszielorientierung interpretiert. Auch Dweck (2000) stellte fest, dass Schüler häufig beide Ziele angeben, wird ihnen die Möglichkeit dazu eingeräumt. Sie bevorzugt deshalb ein diagnostisches Vorgehen, das Schüler zwingt, sich für eine Zielkategorie zu entscheiden. Nur auf diese Weise ließen sich in verschiedenen Studien mit dem Dweckschen Motivations-Modell konforme Befunde aufzeigen und replizieren (siehe Dweck, 2000).

Tabelle 5.12

Verteilungskennwerte und Itemschwierigkeiten der Items zur Erfassung der Lern- bzw. Leistungszielorientierung nach Dweck (2000)

	<i>Median</i>	<i>Interquartilbereich</i>	<i>Z-Wert</i>	<i>Schiefe</i>	<i>SE_{sch}</i>	<i>Schwierigkeit</i>	<i>N</i>
Item 49 (Leistungsziel)	2.3	.80 - .48	2.561**	.394	.12	.40	413
Item 50 (Lernziel)	5.7	4.50 - 6.60	2.857**	-1.178	.12	.76	413

Anmerkung: ** = $p < .01$

Vor dem Hintergrund der hier vorangestellten Analysen, sollten die Items 49 und 50 als Indikatoren für die Lern- bzw. Leistungszielorientierung nach Dweck (2000) akzeptiert werden.

5.4.2.1.4 Situationsspezifische Lern- versus Leistungszielorientierung

In Anlehnung an die beschriebenen Items nach Dweck (2000) wurde ebenfalls ein Item konstruiert, das die in der spezifischen Testsituation ausgelöste Lern- bzw. Leistungszielorientierung im Dweckschen Sinne erfassen sollte. Im Itemstamm wurden dazu beide Ziele gegenübergestellt und im Antwortfeld eine gestufte Zustimmung zu einem der beiden Ziele vom Testanden gefordert.

Die *Verteilungseigenschaften* zeigen, dass die Antworthäufigkeiten dieses Items nicht normal, sondern linkssteil verteilt sind (Tabelle 5.13). Als Basis diente wieder das Antwortverhalten der gesamten Stichprobe vor Beginn des ersten Reasoning-Tests und nach der Konfrontation mit zwei Beispielitems. Auch der *Schwierigkeitsindex* legt nahe, dass die Mehrheit der Schüler das Item im Sinne der Lernzielorientierung¹⁷ beantwortete (Tabelle 5.13).

Tabelle 5.13

Kennwerte der Verteilung und Itemschwierigkeit des Items zur Erfassung der situationsspezifischen Lern- bzw. Leistungszielorientierung (in Anlehnung an Dweck, 2000)

	<i>Median</i>	<i>Interquartilbereich</i>	<i>Z-Wert</i>	<i>Schiefe</i>	<i>SE_{sch}</i>	<i>Schwierigkeit</i>	<i>N</i>
Item 5	1.7	0.5 – 3.6	2.708*	.746	.121	.33	408

Anmerkung: * = $p < .05$

Die validitätsbezogene Zusammenhanganalyse, d.h. die substantiellen Korrelationen der Items zur Erfassung der situationsspezifischen und schulbezogenen Zielori-

¹⁷ Das Item 5 wurde konsistent mit allen situativen Items umgepolt. Eine niedrige Ausprägung weist auf Lernzielorientierung, eine hohe Ausprägung auf Leistungszielorientierung vor bzw. nach der Testsitzung.

entierung (Tabelle 5.14), belegen, dass neben der gemeinsamen Merkmalsvarianz auch ein eigenständiger Anteil an systematischer, situationsspezifischer Merkmalsvarianz durch das situationsspezifische Item abgebildet wird.

Tabelle 5.14

Zusammenhang der Items zur Erfassung der situationsspezifischen und schulbezogenen Lern- bzw. Leistungszielorientierung (in Anlehnung an Dweck, 2000)

situationsspezifische Zielorientierung	schulbezogene Zielorientierung	
	Item 49 (Leistungsziel)	Item 50 (Lernziel)
Item 5 (Lern- versus Leistungsziel)	.16**	-.30**

Anmerkung: $N = 402$, ** = $p < .01$

Das situationsspezifische Item 5 kann als Messgelegenheit zur Erfassung der Lern- bzw. Leistungszielorientierung im Dweckschen Sinne akzeptiert werden.

5.4.2.2 Attribution von Erfolg/Misserfolg

Duda und Nicholls (1992) interpretieren die motivationale Orientierung einer Person auch als persönliche Erfolgskriterien oder Ziele, die eine Person in einer Leistungssituation aufstellt. Folglich werden von Personen angenommene Ursachen für (schulischen) Erfolg als wesentliche Bestimmungsstücke ihrer motivationalen Orientierung angesehen (Abschnitt 2.2.5). Ein entsprechendes Verfahren nach Duda und Nicholls (1992) stellt 19 mögliche Ursachen für schulischen Erfolg auf, die hier auf einer fünfstufigen Skala („strongly agree“ bis „strongly disagree“) bewertet werden sollen. Aus diesem Verfahren wurden nach Anpassung an den britischen Sprachgebrauch je fünf für die vorliegende Untersuchung angemessene Items aus den Kategorien „Betrug/Täuschung“, „Motivation/Anstrengung“ und „Fähigkeit“ ausgewählt und an das in der vorliegenden Studie verwendete Antwortformat angepasst. Diese 15 Items wurden nach ihrer Skalenzugehörigkeit alternierend vorgegeben.

Die Überprüfung der *Dimensionalität* der eingesetzten Skalen mittels Hauptkomponenten-Analyse mit Extraktion und anschließender Varimax-Rotation der Faktoren mit einem Eigenwert > 1 ergab, erwartungsgemäß, eine Drei-Faktoren-Lösung (Kriterium: Scree-Test, 55.91 % aufgeklärte Varianz, $N = 410$) mit allerdings sehr unbefriedigendem Ladungsmuster. Es war deshalb notwendig, drei Items (Item 23, Item 30, Item 31) mit hohen Nebenladungen ($a_{ij} > .30$) bei vergleichsweise geringen Hauptladungen ($a_{ij} < .60$) zu eliminieren. Ein weiteres Item (Item 19) wurde als Ergebnis der nachfolgen-

den Skalenanalyse (extrem geringe Trennschärfe $r_{it} = .09$, bei geringer aber mit anderen Items vergleichbarer Schwierigkeit, $p_i = .78$) entfernt¹⁸. Tabelle 5.15 zeigt die Drei-Faktoren-Lösung der verbleibenden 11 Items. Auch hier zeigt sich für ein Item (Item 32) eine noch hohe Nebenladung ($a_{ij} > .30$) bei vergleichsweise geringer Hauptladung ($a_{ij} < .50$). Bei Betrachtung der Ladungsmuster der übrigen Items scheint diese Drei-Faktoren-Lösung im Sinne der Einfachstruktur jedoch akzeptabel¹⁹ (Kriterium: Scree-Test, 56.59% aufgeklärte Gesamtvarianz, $N = 410$) und somit interpretierbar. Es lassen sich demnach drei Faktoren extrahieren, die jeweils eine Ursachenerklärung für schulischen Erfolg/Misserfolg repräsentieren, diese sind Anstrengung, Betrug bzw. Täuschung sowie Fähigkeit.

Tabelle 5.15

Kennwerte der faktoranalytischen Dimensionalitätsprüfung des Verfahrens zur Erfassung von wahrgenommenen Erfolgsursachen (in Anlehnung an Duda & Nicholls, 1992)

Itemnummer, originale Skalenzugehörigkeit	Kommunalität	Faktor 1	Faktor 2	Faktor 3
Item 26, Anstrengung	.67	.80		
Item 20, Anstrengung	.56	.75		
Item 28, Anstrengung	.53	.70		
Item 33, Anstrengung	.50	.66		-.23
Item 22, Betrug/Täuschung	.58		.76	
Item 27, Betrug/Täuschung	.59		.75	
Item 25, Betrug/Täuschung	.62		.74	.27
Item 32, Betrug/Täuschung	.43	-.36	.47	.28
Item 21, Fähigkeit	.69			.81
Item 24, Fähigkeit	.70		.28	.78
Item 29, Fähigkeit	.37			.59
Eigenwerte		2.30	2.07	1.86
Aufgeklärte Varianz		20.88	18.79	16.93

Anmerkung: Ladungen $< |.20|$ werden nicht dargestellt.

¹⁸ *Eliminierte Items*: Item 19: „People succeed if they enjoy thinking about their school subjects“; Item 23: „People succeed if they help each other learn.“; Item 30: „ People succeed if they are just lucky.“; Item 31: „People succeed if they try to understand instead of just memorizing things.“

¹⁹ *Rotation der Faktorenstruktur*: Hier ist anzumerken, dass Duda und Nicholls (1992) eine schiefwinklige Rotation (Oblimin) der Faktorenstruktur vorschlugen. In der hier untersuchten Stichprobe unterschied sich das Ladungsmuster der oblimin-rotierten Faktorenstruktur jedoch nicht wesentlich von der varimax-rotierten Faktorenstruktur. Aufgrund der höheren Eindeutigkeit in der Interpretation von unabhängigen Faktoren wurde sich deshalb für eine rechtwinklige Faktorenstruktur (Varimax) für die weiteren Analysen entschieden.

Die Skalensummen der Subskala „Betrug/Täuschung“ sowie der Subskala „Fähigkeit“ sind normal verteilt (Subskala „Betrug/Täuschung“: $Z = 1.295$, $p = .07$, $N = 413$; Subskala „Fähigkeit“: $Z = .831$, $p = .495$, $N = 413$), die Skalensumme der Subskala „Anstrengung“ ist dagegen rechtssteil verteilt ($Z = 2.654$, $p < .001$, $N = 410$, $sch = -1.379$, $SE_{sch} = .121$). Die Analyse der *Schwierigkeiten* der Subskalen zeigt, dass die Mehrzahl der Schüler unter anderem Anstrengungsbereitschaft als Ursache für schulischen Erfolg/Misserfolg anführt ($\bar{p}_i = .85$, $N = 410$). Die Items dieser Subskala „Anstrengung“ sind für die hier untersuchte Stichprobe demnach zu leicht. Dagegen wird „Täuschung/Betrug“ eher selten als ursächlich für schulischen Erfolg/Misserfolg angesehen ($\bar{p}_i = .28$, $N = 413$). Dieser Befund kann auch als Hinweis auf sozial erwünschtes Antwortverhalten in der Untersuchungsstichprobe interpretiert werden. Hinsichtlich der Einschätzung von vorhandenen Fähigkeiten als Ursache für schulischen Erfolg/Misserfolg ist keine Bevorzugung einer Antwortrichtung in der hier untersuchten Stichprobe erkennbar ($\bar{p}_i = .52$, $N = 413$). Die durchschnittlichen Schwierigkeiten der Subskala „Betrug/Täuschung“ sowie der Subskala „Fähigkeit“ sind demnach akzeptabel.

Die simultane Betrachtung der *Trennschärfen* und Fremdtrennschärfen der einzelnen Items (Tabelle 5.16) zeigt, dass alle Items jeweils höher mit der Skalensumme der eigenen Skala korrelieren (Trennschärfe) als mit den Skalensummen der beiden anderen Skalen (Fremdtrennschärfe). Die Trennschärfen liegen insgesamt im bevorzugten Bereich ($r_{it} > .30$). Jedoch finden sich auch recht hohe Fremdtrennschärfen. Das ungünstigste Verhältnis von Trennschärfe und Fremdtrennschärfen zeigt Item 32, welches sich bereits in der Prüfung der Dimensionalität als nur mäßig geeignet erwiesen hat. Hohe Fremdtrennschärfen einiger Items zeigen einen Verstoß gegen das Postulat der Orthogonalität der Varimax-Rotation an. Die Tatsache, dass ein Ausschluss dieser Items die Konsistenz der eigenen Skala sogar reduziert, spricht aber dafür, diese Items beizubehalten. Insgesamt scheinen derartige Items sowohl Messgelegenheiten für das „skaleneigene“ Konstrukt darzustellen, als eben auch für Konstrukte „fremder“ Skalen.

Tabelle 5.16

Item- und Skalenanalyse des Verfahrens zur Erfassung von wahrgenommenen schulbezogenen Erfolgsursachen (in Anlehnung an Duda & Nicholls, 1992)

Itemnummer, Skalenzugehörigkeit	Trennschärfe bzw. Fremdtrennschärfe			Homogenität (innerhalb der Subskala)	Schwierigkeit	N
	Anstrengung	Betrug/Täuschung	Fähigkeit			
26, Anstrengung	.61	-.22	-.02	.54	.85	412
20, Anstrengung	.52	-.11	.02	.49	.86	413
28, Anstrengung	.48	-.15	.02	.48	.81	413
33, Anstrengung	.43	-.13	-.09	.48	.86	411
27, Betrug/Täuschung	-.13	.55	.31	.42	.19	413
22, Betrug/Täuschung	-.11	.46	.28	.35	.39	413
25, Betrug/Täuschung	-.07	.53	.40	.41	.36	413
32, Betrug/Täuschung	-.30	.39	.30	.35	.18	413
21, Fähigkeit	.02	.35	.53	.44	.54	413
24, Fähigkeit	.02	.42	.54	.44	.53	413
29, Fähigkeit	-.10	.26	.32	.29	.49	413

Die internen Konsistenzen der drei Subskalen („Anstrengung“ Cronbachs $\alpha = .72$, $N = 410$; „Betrug/Täuschung“ Cronbachs $\alpha = .70$, $N = 413$, „Fähigkeit“ Cronbachs $\alpha = .65$, $N = 413$) liegen im für Fragebogen-Skalen tolerierbaren Bereich. Die durchschnittlichen Homogenitäten der Subskalen sind akzeptabel („Anstrengung“ $\bar{r}_{ii} = .50$, $N = 410$; „Betrug/Täuschung“ $\bar{r}_{ii} = .38$, $N = 413$, „Fähigkeit“ $\bar{r}_{ii} = .39$, $N = 413$).

Insgesamt kann vor dem Hintergrund der vorangestellten Analysen geschlossen werden, dass die zum Zweck der Anpassung an die hier vorgestellte Studie vorgenommenen Modifikationen die psychometrische Qualität der Skalen zu Erfolgsursachen nicht maßgeblich gefährden.

5.4.2.3 Begabungskonzept

Das in der hier vorgestellten Studie einzusetzende Verfahren zur Erfassung des Begabungskonzepts wurde von Dweck (2000) im Rahmen ihrer Untersuchungen zum Konstrukt Lern- versus Leistungszielorientierung für das Schulalter entwickelt. Nach Dweck (2000) ist das Begabungskonzept einer Person eine wesentliche Komponente der Verhaltensregulation durch Zielsetzung (siehe Abschnitt 2.2.5). Von den drei zu diesem Zweck von Dweck (2000) vorgeschlagenen Items wurden zwei für die vorliegende Stu-

die ausgewählt²⁰ und das ursprünglich sechsstufige Antwortformat entsprechend modifiziert. Das erste Item (Item 51) fordert vom Testanden eine Einschätzung der eigenen intellektuellen Fähigkeiten, das zweite Item (Item 52) bezieht sich auf die eigene schulische Lernfähigkeit.

Das Antwortverhalten beider Items zur Erfassung des schulbezogenen Begabungskonzepts (Item 51, Item 52) ist nicht normal (Item 51: $Z = 1.286$, $p < .073$, $N = 413$; Item 52: $Z = 1.722$, $p < .005$, $N = 412$), sondern rechtssteil verteilt (Item 51: $sch = -.542$, $SE_{sch} = .12$; Item 52: $sch = -.816$, $SE_{sch} = .12$). Insbesondere Item 52 zur Erfassung der wahrgenommenen schulischen Lernfähigkeit wird von der Mehrheit der Schüler in Richtung hoher Merkmalsausprägung beantwortet. Die *Itemschwierigkeiten* beider Items weisen darauf hin, dass die Fähigkeitseinschätzungen insgesamt eher optimistisch ausfallen (Item 51: $\bar{p}_i = .61$, $N = 413$; Item 52: $\bar{p}_i = .67$, $N = 412$). Aufgrund des substantiellen Zusammenhangs beider Items ($r_s = .49$, $p < .001$, $N = 412$) lässt sich weiter schlussfolgern, dass beide Items gemeinsame systematische Merkmalsvarianz abbilden, also beide Items jeweils als Messgelegenheit für das Begabungskonzept angesehen werden können.

Situationsspezifisches Begabungskonzept

Für die Erfassung des situationsspezifischen Begabungskonzepts wurden in Anlehnung an das Verfahren von Dweck (2000) zwei Items konstruiert. Ein Item sollte die Einschätzung der eigenen Fähigkeit erfassen (Item 1), die im Test präsentierten Reasoning-Aufgaben zu lösen. Ein weiteres Item (Item 2) bezog sich auf die Einschätzung der eigenen Fähigkeit, im Testverlauf zu lernen, wie entsprechende Reasoning-Aufgaben zu lösen sind.

Die *Verteilungskennwerte* zeigen, dass das Antwortverhalten der situationsspezifischen Items nicht normal, sondern auch hier rechtssteil verteilt ist (Tabelle 5.17). Auch unter situationsspezifischer Perspektive, also der Einschätzung der eigenen Fähigkeiten in Hinblick auf die zu lösenden Reasoning-Aufgaben, fielen die Urteile eher optimistisch aus. Die Iteminterkorrelationen ($r_s = .26$, $p < .001$, $N = 410$) deuten darauf hin, dass beide Items auf das gleiche Konstrukt abzielen.

²⁰ Das dritte von Dweck (2000) vorgeschlagene Item unterscheidet sich nicht von dem ersten genannten (hier eingesetzten) Item und muss auch hinsichtlich seiner sprachlichen Formulierung hinterfragt werden („*I usually think I am intelligent.*“ versus das hier verwendete Item „*I feel pretty confident about my intellectual ability.*“), so dass darauf verzichtet wurde.

Tabelle 5.17

Verteilungskennwerte und Itemschwierigkeiten der Items zur Erfassung des situationsspezifischen Begabungskonzepts (in Anlehnung an Dweck, 2000)

	<i>Median</i>	<i>Interquartilbereich</i>	<i>Z-Wert</i>	<i>Schiefe</i>	<i>SE_{sch}</i>	<i>Schwierigkeit</i>	<i>N</i>
Item 1 (Fähigkeit)	5.0	3.7-5.9	1.812*	-.674	.121	.68	410
Item 2 (Lernen)	4.9	3.5-6.0	1.867*	-.652	.121	.66	410

Anmerkung: * = $p < .05$

Die Zusammenhänge der Items zum schulbezogenen und situationsspezifischen Begabungskonzept (siehe Tabelle 5.18) belegen, dass die Modifikation der Items in Hinblick auf die situationsspezifische Erfassung des Begabungskonzepts angemessen gelungen ist.

Tabelle 5.18

Zusammenhang der Items zur Erfassung des situationsspezifischen und schulbezogenen Begabungskonzepts (in Anlehnung an Dweck, 2000)

situationsspezifisches Begabungskonzept	<i>schulbezogenes Begabungskonzept</i>	
	<i>Item 51 (Fähigkeit)</i>	<i>Item 52 (Lernen)</i>
Item 1 (Fähigkeit)	.42**	.30**
Item 2 (Lernen)	.20**	.19**

Anmerkung: $N = 404$, ** = $p < .01$

5.4.2.4 Testangst

Zur Erfassung von Testangst wurde ein Verfahren nach Meijer (2001) eingesetzt. Das faktoranalytisch konstruierte Persönlichkeitsverfahren basiert auf älteren Arbeiten von Morris (1981) zum Konstrukt Testangst. Meijer (2001) schlägt neben der allgemein anerkannten Unterteilung der Testangst in eine Worry-Komponente und eine Emotionality-Komponente noch eine dritte Komponente von Testangst vor: der Mangel an Selbstsicherheit bzw. ein geringes Begabungskonzept. Das folglich dreidimensionale Verfahren besteht aus 34 Items mit jeweils fünfstufigem Antwortformat und ist für Schüler der Sekundarstufe I und II konzipiert. Auf der Basis der älteren Arbeiten von Morris (1981) zu einer Kurzform, sowie unter Beachtung der faktoriellen Struktur der Skalen (Meijer, 2001) wurden jeweils fünf Items pro Faktor für die vorliegende Untersuchung ausgewählt und an das hier verwendete Antwortformat (Analogskala) angepasst. Diese 15 Items wurden bezüglich ihrer theoretischen Faktorenzugehörigkeit alternierend vorgegeben.

Die Überprüfung der *Dimensionalität* der Items zur Erfassung von Testangst nach Meijer (2001) mittels Hauptkomponenten-Analyse mit Extraktion und anschließender Varimax-Rotation der Faktoren mit einem Eigenwert > 1 ergab zwar tatsächlich eine Drei-Faktoren-Lösung (Kriterium: Scree-Test, 53.9 % aufgeklärte Varianz, $N = 410$) mit jedoch unbefriedigendem Ladungsmuster. Drei Items (Item 36, Item 41, Item 47) mit hohen Nebenladungen ($a_{ij} > .30$) bei relativ geringen Hauptladungen ($a_{ij} < .60$) wurden deshalb sukzessive eliminiert. Weiterhin wurden mit dem Ziel der Aufrechterhaltung der angestrebten Faktorstruktur zwei Items (Item 50, Item 51) integriert. Diese zwei Items gehen ursprünglich auf Dweck zurück und dienen der Erfassung des Begabungskonzepts. Für die nun 14 Items ergibt die Hauptkomponenten-Analyse mit Extraktion und anschließender Varimax-Rotation der Faktoren mit einem Eigenwert > 1 eine Zwei-Faktoren-Lösung (Scree-Test, 46.31% Varianzaufklärung, $N = 411$, siehe Tabelle 5.19). Das jetzt interpretierbare Ladungsmuster legt nahe, den ersten Faktor als Testangst-Komponente und den zweiten Faktor als Selbstsicherheits-Komponente der einbezogenen Items aufzufassen. In der hier untersuchten Stichprobe und Itemauswahl scheint es wenig sinnvoll, zwischen einer Worry-Komponente und einer Emotionality-Komponente bei der Erfassung von Testangst zu unterscheiden. Dies mag daran liegen, dass alle Items zu körperlichen Sensationen, die mit Angsterleben (Emotionality) einhergehen können, aus dem Meijerschen Fragenbogen zur Testangst nicht vorgegeben wurden (z.B. „*I am so tense that my stomach is upset.*“). Eine Zusammenfassung der Itemvarianz zu einem Testangst-Faktor ist aber für die hier zu untersuchende inhaltliche Fragestellung (siehe Kapitel 4) ausreichend.

Tabelle 5.19

Kennwerte der faktoranalytischen Dimensionalitätsprüfung des Verfahrens zur Erfassung von Testangst
(in Anlehnung an Meijer, 2001)

Itemnummer, originale Skalenzugehörigkeit	Kommunalität	Faktor 1	Faktor 2
Item 39, Testangst, Emotionality	.70	.82	
Item 40, Testangst, Worry	.60	.78	
Item 37, Testangst, Emotionality	.54	.73	
Item 38, Testangst, Worry	.48	.69	
Item 44, Testangst, Self-Confidence	.49	.65	
Item 42, Testangst, Emotionality	.35	.59	
Item 46, Testangst, Worry	.30	.54	
Item 45, Testangst, Emotionality	.31	.47	.29
Item 34, Testangst, Worry	.25	.44	.24
Item 48, Testangst, Worry	.16	.36	
Item 51, Self-Confidence	.62		.78
Item 52, Self-Confidence	.55		.74
Item 43, Testangst, Self-Confidence	.57		.74
Item 35, Testangst, Self-Confidence	.60		.73
Eigenwerte der rotierten Faktoren		3.95	2.54
Durch Faktoren aufgeklärter Varianzanteil		28.18	18.12

Anmerkung: Ladungen < |.20| werden nicht aufgeführt.

Die Skalensummen der neu gebildeten Subskalen sind normal verteilt (Testangst-Skala: $Z = .909$, $p = .380$, $N = 412$, Begabungskonzept-Skala: $Z = .946$, $p = .333$, $N = 412$). Die Bestimmung der *Schwierigkeit* der Subskalen zeigt, dass die Mehrzahl der Schüler die Items zur Erfassung der schulbezogenen Testangst ($\bar{p}_i = .64$) als auch des schulbezogenen Begabungskonzepts ($\bar{p}_i = .63$) in Richtung hoher Merkmalsausprägung beantwortete.

Alle den entsprechenden Faktoren zugeordneten Items korrelieren jeweils höher mit der eigenen Skalensumme (Trennschärfe) als mit der Summe der jeweils anderen Skala (Fremdtrennschärfe; Tabelle 5.20). Die Fremdtrennschärfen liegen durchgehend unter dem vorgeschlagenen Kriterium ($r_{it} = .30$) und sind damit vernachlässigbar.

Beide Subskalen erfüllen das Homogenitätskriterium (Subskala „Testangst“: $\bar{r}_{ii} = .33$; Subskala „Begabungskonzept“: $\bar{r}_{ii} = .45$) und weisen akzeptable interne Konsistenzen auf (Subskala „Testangst“: Cronbachs $\alpha = .83$; Subskala „Begabungskonzept“: Cronbachs $\alpha = .77$).

Tabelle 5.20

Item- und Skalenanalyse der Items zur Erfassung der schulbezogenen Testangst (in Anlehnung an Meijer, 2001)

Itemnummer, Skalenzugehörigkeit	Trennschärfe	Fremdtrennschärfe	Homogenität (innerhalb der Subskala)	Schwierigkeit	N
Item 39, Testangst	.74	.27	.45	.62	413
Item 40, Testangst	.65	.15	.41	.73	406
Item 37, Testangst	.61	.22	.39	.67	413
Item 38, Testangst	.53	.10	.36	.73	413
Item 44, Testangst	.59	.29	.37	.63	413
Item 42, Testangst	.44	.09	.27	.56	413
Item 46, Testangst	.44	.14	.30	.57	410
Item 45, Testangst	.43	.27	.26	.50	413
Item 34, Testangst	.39	.24	.27	.56	413
Item 48, Testangst	.32	.15	.24	.79	413
Item 51, Begabungskonzept	.58	.21	.45	.61	413
Item 52, Begabungskonzept	.54	.18	.44	.67	412
Item 43, Begabungskonzept	.58	.27	.45	.60	413
Item 35, Begabungskonzept	.57	.27	.44	.62	413

Zusammenfassend kann anhand der Kennwerte der Item- und Skalenanalyse, der faktoriellen Dimensionalitätsprüfung und Reliabilitätsbestimmung der Subskalen die hier vorgeschlagene zweifaktorielle Struktur des Verfahrens zur Erfassung von Testangst akzeptiert und als inhaltlich interpretierbar bewertet werden.

Situationsspezifische Leistungsangst

In Anlehnung an je ein Item der ursprünglich konzipierten Emotionality-Komponente (Item 37) sowie der Worry-Komponente (Item 40) des Verfahrens zur Erfassung von Testangst (Meijer, 2000) wurde jeweils ein testspezifisches Item formuliert.

Die Antworthäufigkeiten beider Items (Item 7, Item 8) zur Erfassung der situationsspezifischen Leistungsangst vor der ersten Reasoning-Testung sind nicht normal verteilt (Tabelle 5.21). Anhand der zwar akzeptablen *Itemschwierigkeiten* ist ebenfalls erkennbar, dass beide Items eher ablehnend beantwortet wurden (Tabelle 5.21). Die Mehrheit der Testanden gab demnach bei Konfrontation mit den Beispielitems für die im Test zu lösenden Aufgaben an, eher „entspannt“ bzw. „optimistisch“ zu sein, „so gut als eben möglich“ im anschließenden Test abzuschneiden. Die Item-Interkorrelation ist dabei substantiell und spricht dafür, dass beide Items gemeinsame systematische Merkmalsvarianz erfassen ($r_s = .52, p < .001, N = 410$).

Tabelle 5.21

Kennwerte der Verteilung und Itemschwierigkeit des Items zur Erfassung der situationsspezifischen Leistungsangst (in Anlehnung an Meijer, 2001)

	<i>Median</i>	<i>Interquartilbereich</i>	<i>Z-Wert</i>	<i>Schwierigkeit</i>	<i>N</i>
Item 7 (Emotionality)	3.2	1.0-4.9	1.965*	.44	410
Item 8 (Worry)	3.5	1.5-4.8	1.478*	.48	410

Anmerkung: * $p < .05$

Die im Sinne einer Validitätsüberprüfung zu betrachtenden Korrelationen zwischen den Items zur situationsspezifischen Leistungsangst mit der Subskala „Testangst“ zeigen, dass hier gemeinsame Merkmalsvarianz erfasst wird. Sowohl die substantielle Item-Interkorrelation ($r_s = .52$, $p < .001$, $N = 410$) als auch die vergleichbare Zusammenhangsstärke beider situationsspezifischen Items zur Subskala „Testangst“ (Item 7: $r_s = .35$, $p < .001$, $N = 402$; Item 8: $r_s = .39$, $p < .001$, $N = 402$) sprechen gegen die Notwendigkeit, auch auf situationsspezifischer Ebene zwischen einer Worry-Komponente und einer Emotionality-Komponente der Testangst zu unterscheiden.

5.4.2.5 Schulische Motivation

Des Weiteren wurden die Lehrer gebeten, jeden ihrer Schüler hinsichtlich seiner schulischen *Leistungsmotivation* auf einer fünfstufigen Skala einzuschätzen. Ziel war hier, die Selbsturteile der Schüler um Fremdurteile zu ergänzen. Um diesen Aufwand für den Lehrer entsprechend gering zu halten, wurde auf eine Skala zurückgegriffen, welche lediglich die Zuordnung eines Schülers zu einem von fünf Verhaltenstypen fordert (Elliott, Hufton, Illushin & Frase, 2001). Die Skala zielt bewusst auf im Klassenraum durch den Lehrer beobachtbares Verhalten der Schüler und soll eine Einschätzung der Motivationslage im Sinne von schulischem Engagement, Interesse, Anstrengungsbereitschaft der Schüler unabhängig von Selbsturteilen der Schüler ermöglichen.

Das Urteilsverhalten der Lehrer fällt eher wohlwollend aus, was sich in der geringen Schwierigkeit der Skala ausdrückt ($\bar{p}_i = .63$, $N = 414$). Somit kann die Zuordnungshäufigkeit auch nicht normal verteilt sein ($Z = 4.122$, $p < .001$, $N = 414$). Immerhin knapp die Hälfte der Schüler wurde als im Schulalltag typischerweise sehr hoch (19.3%) bzw. überdurchschnittlich (25.8%) motiviert, und dagegen nur ein vergleichsweise geringer Anteil der Schüler als oft (12.6%) oder konsistent (4.6%) uninteressiert eingestuft.

5.4.3 Zusammenfassung zu den psychometrischen Eigenschaften der eingesetzten Verfahren

Die Aufgabenanalyse der figuralen Reasoning-Items ergab im Gegensatz zu anderen Untersuchungen nach Beckmann und Guthke (1999) nur mäßige Itemkennwerte sowie Reliabilitätsindizes. Als plausible Ursachen für die nur mäßige psychometrische Qualität der figuralen Reasoning-Tests bieten sich der vergleichsweise geringe Stichprobenumfang und natürlich die Veränderung der Testprozedur an. Eine antizipierte Konsequenz daraus ist, dass aufgrund der geringen Reliabilität der figuralen Reasoning-Itemsets potentielle Feedback-Effekte auf die Leistung schwieriger zu identifizieren sein werden. So kann mit einem wenig reliablen Messinstrument nur der Nachweis eines starken (generellen sowie differentiellen) Effekts von Richtig/Falsch-Feedback auf die figurale Reasoning-Leistung erwartet werden. Die psychometrische Qualität der Tests zur Erfassung von numerischer Reasoning-Fähigkeit ist dagegen zufrieden stellend.

Da die eingesetzten persönlichkeitsdiagnostischen Verfahren teilweise verändert, bzw. für die vorliegende Untersuchung erst entwickelt wurden, war eine genauere Überprüfung ihrer psychometrischen Qualität notwendig. Demnach sollten Testwerte mit Zurückhaltung interpretiert werden, die mittels der Skalen zur Erfassung der Ursachen für schulischen Erfolg/Misserfolg (Duda & Nicholls, 1992) erhoben wurden. Zudem waren die Testrohrewerte einiger Verfahren (z.B. Subskala Aufgabenorientierung der MOS) nicht normal verteilt. Eine Normalverteilung der Testrohrewerte kann jedoch nicht bei jedem zu messenden Konstrukt erwartet werden. So ist zum Beispiel bekannt, dass das Konstrukt Ängstlichkeit nicht normal verteilt ist (Amelang & Bartussek, 2001). Insgesamt fielen die Kennwerte der Item- und Skalenanalysen der eingesetzten persönlichkeitsdiagnostischen Verfahren jedoch gut bis zufrieden stellend aus.

Interessant war die Gegenüberstellung der verschiedenen Skalen zur Erfassung der motivationalen Orientierung oder auch Zielorientierung. Beide Verfahren ergaben akzeptable Itemkennwerte sowie Reliabilitäts- und (faktorielle) Validitätskennwerte (MOS), so dass beide psychodiagnostischen Zugänge, die ein eindimensionales oder ein zweidimensionales Modell der motivationalen Orientierung repräsentieren, für die Beantwortung der inhaltlichen Fragestellungen genutzt werden können. Zudem werden auf diese Weise potentiell Bedingungen geschaffen, die eine bessere Generalisierbarkeit der Befunde ermöglichen. Dabei ist jedoch zu beachten, dass beide verschiedenen diagnostischen Zugänge zu teilweise unterschiedlichen Ergebnissen führten. So gaben Schüler an, eher Lernziele zu verfolgen, wenn sie aufgefordert wurden, sich für eine der

beiden Zielkategorien zu entscheiden (Dweck, 2000). Werden jedoch beide Zielkategorien zugelassen (MOS, Duda & Nicholls, 1992), berichteten Schüler durchaus auch Leistungsziele, und zwar sowohl im schulischen als auch im situationsspezifischen Kontext. Somit wird deutlich, dass beide Ziele auch gleichzeitig verfolgt werden (zweidimensionales Modell). Wie die in Abschnitt 2.2.5 dargestellten Befunde zum Zusammenhang von Zielorientierung und Leistung belegen, sind Leistungsdefizite unter evaluativen Leistungsbedingungen insbesondere bei einer *Leistungszielorientierung* (nach Nicholls Ich-Orientierung) und gleichzeitig geringem Begabungskonzept zu erwarten. Deshalb soll dem zweidimensionalen Modell der Vorzug gegeben werden.

6 Ergebnisse

Nachdem die psychometrische Qualität der eingesetzten Verfahren überprüft ist und als akzeptabel bewertet werden kann (Abschnitt 5.4), widmet sich das nächste Kapitel der Prüfung der inhaltlichen Hypothesen.

6.1 Generelle Feedback-Effekte

Zunächst soll der Frage nach einem generellen Feedback-Effekt nachgegangen werden. Dabei wird angenommen, dass die unter der Bedingung mit itemspezifischem Richtig/Falsch-Feedback erbrachte Reasoning-Leistung höher ist als die unter Non-Feedback-Bedingungen ($H_A: \mu_{F-} < \mu_{F+}$, *generelle Feedback-Hypothese*). Die Untersuchung des generellen Feedback-Effekts wird jeweils getrennt für den figuralen und numerischen Reasoning-Test vorgenommen. Auf Zwischen-Gruppen-Vergleiche wird dabei mit dem Begriff *interindividuelle Feedback-Hypothese* Bezug genommen, auf Inner-Gruppen-Vergleiche mit dem Begriff *intraindividuelle Feedback-Hypothese*.

Um zu prüfen, ob Personen ihre Leistungen unter Feedback-Bedingungen im Vergleich zu ihren Leistungen unter feedbackfreien Testbedingungen verbessern (intraindividuelle Feedback-Hypothese), müssen Reihenfolgeeffekte kontrolliert werden (siehe Versuchsplan, Tabelle 5.6). So ist der Anteil der Varianz zwischen zwei Messzeitpunkten zu prüfen, der durch die Wiederholungstestung bedingt ist (Paralleltesteffekt).

6.1.1 Numerische Reasoning-Testleistung

Versuchsgruppe 1 (F- F+) und Kontrollgruppe (F- F-) unterscheiden sich erwartungsgemäß nicht in ihrer Leistung – operationalisiert als Anzahl korrekter Antworten – zum ersten Messzeitpunkt ($F_{1/76} = 1.256$, $p = .266$, $N = 78$). Um nun zu prüfen, ob Feedback leistungsförderlich ist, wird die Veränderung des Leistungsverhaltens der in ihrem Ausgangsniveau gleichen Versuchs- und Kontrollgruppe vom ersten Messzeitpunkt zum zweiten Messzeitpunkt gegenübergestellt (*intraindividuelle Feedback-Hypothese*). Dabei wird erwartet, dass sich Versuchsgruppe 1 (F- F+), die zum zweiten Messzeitpunkt Feedback erhalten hat, in ihrer Leistung im Vergleich zur Kontrollgruppe (F- F-) verbessert. Ein solcher leistungsförderlicher Feedback-Effekt sollte sich in einer signifikanten Wechselwirkung im Rahmen einer einfaktoriellen Varianzanalyse mit Messwiederholung zeigen.

Die Analyse ergibt keine signifikante Wechselwirkung von experimenteller Bedingungsvariation (Feedback/kein Feedback) und Messzeitpunkt ($F_{1/76} = 2.488, p = .119$). Die Nullhypothese kann nicht zurückgewiesen werden ($H_0: \mu_{F-} = \mu_{F+}$). Ein Paralleltesteffekt kann ebenso wenig festgestellt werden (Hauptwirkung Messwiederholung: $F_{1/76} = .611, p = .437, N = 78$). Insgesamt ist unter den gegebenen Stichprobenverhältnissen ein mittlerer und starker Feedback-Effekt (genauer ein Effekt von $f \geq .32$) mit hinreichender statistischer Sicherheit²¹ ($1 - \beta \geq .80$) auszuschließen.

Für einen interindividuellen Gruppenvergleich soll die Leistung der Versuchsgruppe 2 (F+ F-), die zum ersten Messzeitpunkt Feedback erhalten hat, mit der Leistung der Gruppen, die zum ersten Messzeitpunkt kein Feedback erhalten haben (Kontrollgruppe, Versuchsgruppe 1) verglichen werden (*interindividuelle Feedback-Hypothese*). Die Varianzanalyse ergibt auch hier keinen signifikanten Haupteffekt der experimentellen Bedingung (Feedback/kein Feedback zum Messzeitpunkt 1; $F_{1/104} = .047, p = .829$). Die Nullhypothese muss beibehalten werden ($H_0: \mu_{F-} = \mu_{F+}$). Es wird erwartungsgemäß auch für den interindividuellen Leistungsvergleich ein mittlerer und starker Feedback-Effekt (genauer $f \geq .28$) mit akzeptabler statistischer Sicherheit ($1 - \beta \geq .80$) ausgeschlossen.

Weder unter einer intraindividuellen Perspektive noch unter einer interindividuellen Perspektive wurde ein Feedback-Effekt auf die numerische Reasoning-Leistung festgestellt. Konsequenterweise treten keine Reihenfolgeeffekte der Intervention (Feedback zum ersten versus zweiten Testzeitpunkt) auf. Leistungsverbesserungen aufgrund einer Messwiederholung können ebenfalls ausgeschlossen werden (kein Paralleltesteffekt). Tabelle 6.1 stellt die numerische Reasoning-Leistung in den experimentellen Gruppen nochmals gegenüber.

²¹ Zur Berechnung der statistischen Power wurde das Programm *G•Power* verwendet (Buchner, Faul & Erdfelder, 1997).

Tabelle 6.1

Mittelwerte und Standardabweichungen der *numerischen Reasoning*-Leistung (Anzahl Richtige) in den experimentellen Gruppen

	Messzeitpunkt 1		Messzeitpunkt 2	
	\bar{x}	s	\bar{x}	s
Versuchsgruppe 1 ($N = 51$)	7.29	2.81	7.84	2.73
Versuchsgruppe 2 ($N = 28$)	7.68	2.31	6.96	2.50
Kontrollgruppe ($N = 27$)	8.04	2.74	7.85	3.21

Anmerkung: Versuchsgruppe 1 (F–F+), Versuchsgruppe 2 (F+ F–), Kontrollgruppe (F–F–)

6.1.2 Figurale Reasoning-Testleistung

Die analoge Analyse der Feedback-Effekte auf die figurale Reasoning-Leistung ergab im Großen und Ganzen ein konsistentes Befundmuster. So ließen sich ebenfalls keine performanzbezogenen positiven Effekte der experimentellen Bedingungsvariation nachweisen. Zur Prüfung eines Feedback-Effekts wurde wie in Abschnitt 6.1.1 die Veränderung im Leistungsverhalten vom ersten zum zweiten Messzeitpunkt in Versuchsgruppe 1 (F– F+) und Kontrollgruppe (F– F–) verglichen (*intraindividuelle Feedback-Hypothese*). Auch hier unterschieden sich beide Gruppen erwartungsgemäß nicht hinsichtlich ihrer feedbackfreien Testleistung zum ersten Messzeitpunkt ($F_{1/105} = .326, p = .569$). Die Varianzanalyse mit Messwiederholung ergab keine signifikante Wechselwirkung von experimenteller Bedingungsvariation (Feedback/kein Feedback) und Messzeitpunkt ($F_{1/105} = .463, p = .498$). Die Nullhypothese wird folglich auch hier beibehalten ($H_0: \mu_{F-} = \mu_{F+}$). Unter den gegebenen Stichprobenverhältnissen ist mit hinreichender Sicherheit ($1 - \beta \geq .80$) davon auszugehen, dass performanzbezogenes Richtig/Falsch-Feedback keinen moderaten oder etwa starken Effekt (genauer keinen Effekt von $f \geq .28$) auf das Leistungsverhalten hat.

Im Sinne eines interindividuellen Leistungsvergleichs wurden wie in Abschnitt 6.1.1 die Leistungen der experimentellen Gruppen verglichen, die zum ersten Messzeitpunkt einen Test mit Feedback (Versuchsgruppe 2) versus ohne Feedback (Versuchsgruppe 1 und Kontrollgruppe) bearbeitet haben ($F_{1/140} = 3.509, p = .063, \textit{interindividuelle Feedback-Hypothese}). Der sich im Ergebnis andeutende Trend ist in gewisser Weise kontra-intuitiv, d.h. die Leistungen scheinen unter Feedback-Bedingungen sogar tendenziell schlechter zu werden (siehe dazu die in Abschnitt 2.2.3 referierten Befunde von Kluger & DeNisi, 1996). Unter den gegebenen Stichprobenverhältnissen kann jedoch ein$

mittlerer bis starker *negativer* Feedback-Effekt (genauer $f \geq .24$) mit hinreichender statistischer Sicherheit ($1 - \beta \geq .80$) ausgeschlossen werden.

Ein Vergleich der Veränderung im Leistungsverhalten vom ersten zum zweiten Messzeitpunkt zwischen *Versuchsgruppe 2* (F+ F-) und Kontrollgruppe (F- F-) ergibt keine signifikante Wechselwirkung von Feedback-Bedingung (Feedback/kein Feedback) und Messzeitpunkt (*intraindividuelle Feedback-Hypothese*). Das Angebot von Feedback im figuralen Reasoning-Test beeinflusst nicht die Leistung in einem nachfolgenden feedbackfreien Paralleltest. Unter den gegebenen Stichprobenverhältnissen ist jedoch lediglich ein starker Transfereffekt (genauer $f \geq .42$) von Feedback mit hinreichender statistischer Sicherheit ($1 - \beta \geq .80$) auszuschließen. Der signifikante Haupteffekt des Messzeitpunkts weist aber auf einen leistungsförderlichen Effekt der *Wiederholungsmessung* für beide experimentellen Gruppen hin (Haupteffekt Messwiederholung: $F_{1/65} = 13.163$, $p = .001$; Wechselwirkung: $F_{1/65} = 2.622$, $p = .11$). Offenbar begünstigt das in diesem Test realisierte Multiple-Choice-Antwortformat einen Paralleltesteffekt. Eine Leistungsverbesserung bei Messwiederholung findet sich indessen nicht für die *Versuchsgruppe 1* (F- F+) verglichen mit der Kontrollgruppe (F- F-; Haupteffekt Messwiederholung: $F_{1/105} = 2.145$, $p = .146$). Die zum ersten Messzeitpunkt beobachtete leistungshemmende Wirkung von Feedback wird hier möglicherweise durch einen leistungsförderlichen Paralleltesteffekt überlagert (Reihenfolgeeffekt der Intervention).

Insgesamt ergeben sich für den figuralen Reasoning-Test bestenfalls Hinweise auf einen schwachen generellen, jedoch negativen Feedback-Effekt. Zum einen zeigen Testanden in der Feedback-Bedingung tendenziell geringere Leistungen als Testanden in der feedbackfreien Bedingung (*interindividuelle Feedback-Hypothese*). Zum anderen lässt sich die in einer feedbackfreien Testung beobachtete Leistungsverbesserung bei Wiederholungsmessung in einer Testbedingung mit Feedback zum zweiten Messzeitpunkt nicht ausmachen (*intraindividuelle Feedback-Hypothese*). Das Angebot von Richtig/Falsch-Feedback verhindert offensichtlich die Leistungsverbesserung bei Messwiederholung mit parallelen Testverfahren. Tabelle 6.2 gibt die Mittelwerte und Standardabweichungen der figuralen Reasoning-Leistung in den experimentellen Gruppen wieder.

Tabelle 6.2

Mittelwerte und Standardabweichungen der *figuralen* Reasoning-Leistung (Anzahl Richtige) in den experimentellen Gruppen

	Messzeitpunkt 1		Messzeitpunkt 2	
	\bar{x}	s	\bar{x}	s
Versuchsgruppe 1 (N = 75)	8.05	2.04	8.21	1.73
Versuchsgruppe 2 (N = 35)	7.29	1.62	8.43	1.79
Kontrollgruppe (N = 32)	7.81	1.89	8.25	2.27

Anmerkung: Versuchsgruppe 1 (F- F+), Versuchsgruppe 2 (F+ F-), Kontrollgruppe (F- F-)

Zusammenfassend ist festzuhalten: Richtig/Falsch-Feedback hatte keinen Effekt auf die *numerische* Reasoning-Testleistung. Auch ergeben sich bestenfalls Hinweise auf einen schwachen Feedback-Effekt bei Analyse der *figuralen* Reasoning-Testleistung. Das etwas unterschiedliche Befundmuster lässt sich zum einen auf die Reasoning-Domäne (figural versus numerisch), zum anderen auf das unterschiedliche Antwortformat (Multiple-Choice versus freies Antwortformat) der Reasoning-Tests zurückführen. Möglicherweise fördert Feedback bei Multiple-Choice-Antwortformat eine Ratestrategie und führt deshalb tendenziell zu geringeren Leistungen. Insgesamt konnte aber anhand der hier vorgenommenen Analysen entgegen den ursprünglichen Erwartungen kein genereller *leistungsförderlicher* Effekt von Richtig/Falsch-Feedback festgestellt werden (*generelle Feedback-Hypothese*).

Zusammenhangsanalysen der Leistungen zu beiden Messzeitpunkten legen jedoch Rangplatzverschiebungen von Test 1 zu Test 2 in den Versuchsgruppen nahe. So lassen sich zumindest auf deskriptiver Ebene in den Versuchsgruppen, die entweder zum ersten Messzeitpunkt (Versuchsgruppe 2) oder zum zweiten Messzeitpunkt (Versuchsgruppe 1) Feedback erhalten haben, die Leistungen bei Wiederholungsmessung schlechter anhand der Leistungen der Ersttestung vorhersagen als in den Kontrollgruppen (siehe Tabelle 6.3). Dieser Befund gilt unabhängig vom Materialbereich (Figurenfolgen, Zahlenfolgen) und legt daher differentielle Feedback-Effekte nahe (entspricht der Paralleltest-Reliabilität, siehe differentielle Diagnostizierbarkeit, Jäger, 1978).

Tabelle 6.3

Zusammenhangsanalyse der numerischen sowie der figuralen Reasoning-Leistung zu beiden Messzeitpunkten in den experimentellen Gruppen

Materialbereich	experimentelle Bedingung	<i>r</i>	<i>N</i>
Zahlenfolgen	Versuchsgruppe 1 (F- F+)	.72**	51
	Versuchsgruppe 2 (F+ F-)	.73**	28
	Kontrollgruppe (F- F-)	.85**	27
Figurenfolgen	Versuchsgruppe 1 (F- F+)	.45**	75
	Versuchsgruppe 2 (F+ F-)	.47**	35
	Kontrollgruppe (F- F-)	.64**	32

Anmerkung: F- = kein Feedback; F+ = Feedback; ** = $p < .01$

Im folgenden Abschnitt soll deshalb der Frage nachgegangen werden, ob außerintellektuelle Merkmalskonstellationen subgruppenspezifische und somit differentielle Feedback-Effekte moderieren. So wäre es theoretisch möglich, dass sich differentielle Effekte unter einer generellen Perspektive gegenseitig kompensieren, was das Ausbleiben von Feedback-Effekten in den vorangegangenen Analysen erklären würde.

6.2 Differentielle Feedback-Effekte

In Abschnitt 6.1 wurde gezeigt, dass die Vorhersagegenauigkeit der Reasoning-Leistung in der ersten Testsitzung bezogen auf die Reasoning-Leistung in der zweiten Testsitzung in den Versuchsgruppen tendenziell geringer ist als in den Kontrollgruppen. Es wurde daraus der Schluss gezogen, dass in den Versuchsgruppen differentielle Feedback-Effekte wirken. Im Sinne der differentiellen Diagnostizierbarkeit (Jäger, 1978) liegt hier die Suche nach möglichen, auch außerintellektuellen Moderatoren des Zusammenhangs von Fähigkeit und Leistung nahe. Zum Beispiel könnte ein *testängstlicher* Schüler auf Misserfolg anzeigendes Feedback mit ressourcenbeanspruchenden Worry-Kognitionen reagieren und verglichen mit einer feedbackfreien Testbedingung seine Leistung verschlechtern. Dagegen sollte ein wenig ängstlicher Schüler Feedback produktiv nutzen und höhere Leistungen erreichen als in einer feedbackfreien Testbedingung. Eine zentrale Annahme beim dynamischen Testen ist, dass die verbesserte Leistung unter diesen „leistungsoptimierenden Testbedingungen“ einen angemesseneren Indikator für die „wahre“ Leistungsfähigkeit darstellt (siehe Kapitel 3).

Die beobachtete Reduktion des Anteils an aufgeklärter Leistungsvarianz in der Feedback-Bedingung ließe sich aber auch auf eine feedbackbedingte Erhöhung der *Fehlervarianz* innerhalb der Leistungsvarianz zurückführen. Zur Prüfung dieser Annahme

wurden die Testleistungen – operationalisiert als Anzahl Richtige – in den Bedingungen mit und ohne Feedback mit den zusätzlich vorliegenden Leistungsvariablen (Intelligenztest-Leistungen, Schulleistungen, siehe Kapitel 5) auf ihre Zusammenhänge getestet.

Tabelle 6.4 zeigt, dass keine höheren Zusammenhänge von Testleistung und Außenkriterien in den feedbackfreien Testbedingungen verglichen mit den Feedback-Bedingungen auftreten. Somit kann die Annahme Feedback führe lediglich zur Erhöhung des Anteils an unsystematischer Varianz (Fehlervarianz) innerhalb der Leistungsvarianz zurückgewiesen werden.

Tabelle 6.4

Korrelationen der numerischen und figuralen Reasoning-Leistung mit Außenkriterien (Schulleistungen, Intelligenztest-Leistungen) in den feedbackfreien Testbedingungen sowie in den Feedback-Bedingungen zum ersten und zweiten Messzeitpunkt

		Feedback				kein Feedback			
		Test 1	N_{Test1}	Test 2	N_{Test2}	Test 1	N_{Test1}	Test 2	N_{Test2}
Zahlenfolgen	Intelligenztest	.39*	28	.63**	51	.60**	78	.42**	55
	Mathematiknote	.71**	28	.80**	51	.73**	78	.64**	55
	Gesamtnote	.61**	28	.66**	51	.67**	78	.63**	55
Figurenfolgen	Intelligenztest	.28	35	.42**	75	.30**	107	.48**	67
	Mathematiknote	.37*	35	.56**	75	.36**	107	.47**	67
	Gesamtnote	.25	35	.54**	75	.39**	107	.48**	67

Anmerkung: Die Schulleistungen liegen als T -Normen vor. * = $p < .05$; ** = $p < .01$

In den relativ niedrigen Korrelationen zwischen den *figuralen* Reasoning-Leistungen und den Schulleistungen sowie Intelligenztest-Leistungen, verglichen mit den entsprechenden Korrelationen der *numerischen* Reasoning-Leistung (Tabelle 6.4), spiegelt sich die geringe Reliabilität des Messinstruments zur Erfassung von figuralem Reasoning wieder (siehe Abschnitt 5.4). Die Einschränkung der Korrelationen trifft konsistenter Weise sowohl für die Feedback- als auch für die feedbackfreie Testbedingung zu.

Da eine Induktion von Fehlervarianz durch Feedback ausgeschlossen werden kann, soll geprüft werden, welche anderen Variablen zusätzliche Leistungsvarianz verursachen. Speziell soll geklärt werden, ob sich Variablen finden lassen, die die unterschiedliche Steigung der Regression der abhängigen Variablen (Reasoning-Leistung in Test 2) auf die unabhängige Variable (Reasoning-Leistung in Test 1) in den Versuchsgruppen verglichen mit der Kontrollgruppe bedingen. Solche Variablen werden als *Moderatorvariablen* bezeichnet (z.B. Aguinis, 2004; Bartussek, 1970). Eine Moderatorvariable

ist als eine weitere unabhängige Variable operationalisiert, die den Zusammenhang (Richtung und Stärke) zwischen einer unabhängigen und einer abhängigen Variable bestimmt. Moderatorvariablen zeigen unter welchen Bedingungen, d.h. bei welcher Ausprägung der Moderatorvariablen ein Zusammenhang zwischen zwei Variablen besteht. Davon abzugrenzen sind *Mediatorvariablen*, die in einem kausalen Zusammenhang zur unabhängigen Variablen stehen. Letztere beschreiben mögliche Ursachen für einen Zusammenhang zwischen zwei Variablen (deshalb auch *Process-Variables*). Lässt sich ein Moderatoreffekt ausmachen, sollte der zugrunde liegende Prozess über eine Mediatoranalyse aufgeklärt werden (z.B. Baron & Kenny, 1986).

6.2.1 Feedback-Effekte und intellektuelle Leistungsfähigkeit

6.2.1.1 Numerische Reasoning-Testleistung

In einem regressionsanalytischen Ansatz lässt sich durch Hinzunahme einer Dummy-Variablen (Feedback/kein Feedback) der neben dem Prädiktor (Leistung in Test 1) durch die experimentelle Bedingungsvariation (Feedback/kein Feedback) zusätzlich aufklärbare Varianzanteil in der abhängigen Variablen (Leistung in Test 2) schätzen (Modell 1). Darüber hinaus ist somit prüfbar, ob die experimentelle Bedingungsvariation (Feedback/kein Feedback) die Regression der Leistung zum Messzeitpunkt 2 auf die Leistung zum Messzeitpunkt 1 moderiert (Modell 2, siehe vorhergehender Abschnitt). Bei Nachweis eines Moderatoreffekts für die Variable Feedback/kein Feedback ließen sich unterschiedliche Anstiege der Regressionsgeraden für die Versuchs- und Kontrollgruppe auf das Wirken des Moderators (Feedback/kein Feedback) zurückführen. Gleichzeitig wird die Abhängigkeit eines potentiellen Feedback-Effekts von der feedbackfreien Leistung (Leistung in Test 1) des Testanden geprüft, da diese als Prädiktor in der Regressionsgleichung berücksichtigt wird (*Ausgangswerthypothese*, siehe Kapitel 4).

In die Analyse gehen die Versuchsgruppe 1 (F– F+) sowie die Kontrollgruppe (F– F–) ein. Zur Vorhersage der Testleistung zum zweiten Messzeitpunkt werden folgende Prädiktoren berücksichtigt: die Testleistung zum ersten Messzeitpunkt (feedbackfreie Testleistung), als potentieller Moderator die experimentelle Bedingung (Feedback/kein Feedback zum Messzeitpunkt 2) sowie der Produktterm des potentiellen Moderators (Feedback/kein Feedback zum Messzeitpunkt 2) und des Prädiktors (Leistung in Test

1). Ein Moderatoreffekt der Variablen Feedback/kein Feedback sollte sich in einem signifikant von Null verschiedenen Beta-Gewicht des Produktterms zeigen ($H_0: \beta_{\text{Produkt}} = 0$).

Tabelle 6.5 zeigt die Kennwerte der entsprechenden Regressionskoeffizienten in Modell 1 (einfache Regression, $R^2 = .64$, $F_{2/76} = 69.194$, $p < .001$) und Modell 2 (multiple moderierte Regression; $R^2 = .65$, $F_{3/76} = 47.51$, $p < .001$). Die durch die Prädiktoren aufgeklärte Varianz wird, wenn nicht anders vermerkt, generell als adjustiertes R^2 angegeben. In der multiplen moderierten Regression entspricht die Prüfung der Signifikanz der Effektgröße ΔR^2 anhand des entsprechenden F -Wertes der Signifikanzprüfung des Beta-Gewichts des Produktterms anhand des t -Wertes (z.B. Aguinis, 2004). Deshalb werden die F -Werte für die Effektgröße ΔR^2 hier generell nicht angegeben, sobald das Beta-Gewicht des Produktterms und der t -Wert aufgeführt sind.

Tabelle 6.5

Kennwerte der Regressionskoeffizienten zweier Regressionsanalysen zur Vorhersage der *numerischen Reasoning-Testleistung* zum zweiten Messzeitpunkt der Versuchsgruppe 1 (F–F+) und der Kontrollgruppe (F– F–)

	Prädiktoren	B	SE	β	t	p
Modell 1	Leistung in Test 1 (L)	.852	.072	.816	11.763**	.000
	experimentelle Bedingung zu Test 2 (F) (Feedback/kein Feedback)	.761	.416	.127	1.829	.071
Modell 2	Leistung in Test 1 (L)	.996	.123	.953	8.120**	.000
	experimentelle Bedingung zu Test 2 (F) (Feedback/kein Feedback)	2.458	1.244	.409	1.976	.052
	Produktterm (LxF)	-.219	.151	-.312	-1.446	.152

Anmerkung: Modell 1 = einfache Regression; Modell 2 = multiple moderierte Regression; Kodierung der Dummy-Variable: Feedback = 1, kein Feedback = 0; Aufgrund eines Residuums von mehr als drei Standardabweichungen wurde ein Fall als Ausreißer bewertet (Fall 107, standardisiertes Residual: -3.498) und aus der Regressionsanalyse ausgeschlossen.

Ein genereller Feedback-Effekt bleibt aus (Modell 1: $\Delta R^2 = .02$, $p = .071$, $N = 77$, *intraindividuelle Feedback-Hypothese*). Modell 2 ($\Delta R^2 = .01$, $p = .152$, $N = 77$) zeigt, dass die experimentelle Bedingungsvariation (Feedback/kein Feedback) den Zusammenhang von Leistung in Test 1 und Leistung in Test 2 nicht moderiert. Es besteht also keine Wechselwirkung von *Leistungsfähigkeit* (feedbackfreie Leistung zum ersten Messzeitpunkt) und experimenteller Bedingungsvariation (Feedback/kein Feedback zum zweiten Messzeitpunkt) bezogen auf die Leistung zum zweiten Messzeitpunkt (*Ausgangswerthypothese*). Die Nullhypothese wird beibehalten ($H_0: \beta_{\text{Produkt}} = 0$). Unter den gegebenen Stichprobenverhältnissen kann mit hinreichender statistischer Sicherheit ($1 - \beta \geq .80$)

jedoch nur ein mittlerer und starker differentieller Effekt (genauer $f^2 \geq .11$) der Ausgangsleistung ausgeschlossen werden.

Auch unter Beachtung der *Ausgangsleistung*, d.h. der feedbackfreien Leistung zum ersten Messzeitpunkt, lässt sich in der vorliegenden Stichprobe kein leistungsförderlicher Richtig/Falsch-Feedback-Effekt von psychologisch relevanter Größe auf die numerische Reasoning-Leistung feststellen.

6.2.1.2 Figurale Reasoning-Testleistung

Die hier verfolgte Analysestrategie ist analog. In die Regressionsanalysen gehen die Daten der Versuchsgruppe 1 (F- F+) und der Kontrollgruppe (F- F-) ein. Folgende Prädiktoren werden berücksichtigt: die Testleistung zum ersten Messzeitpunkt (feedbackfreie Testleistung), als potentieller Moderator die experimentelle Bedingung zum zweiten Messzeitpunkt (Feedback/kein Feedback) und der Produktterm aus Prädiktor (Testleistung zum ersten Messzeitpunkt) und Moderator (experimentelle Bedingung). Es soll die Leistung zum zweiten Messzeitpunkt vorhergesagt werden. Tabelle 6.6 zeigt die Kennwerte der Regressionskoeffizienten in Modell 1 ($R^2 = .25$, $F_{2/106} = 18.302$, $p < .001$) und Modell 2 ($R^2 = .62$, $F_{3/106} = 59.390$, $p < .001$).

Tabelle 6.6

Kennwerte der Regressionskoeffizienten zweier Regressionsanalysen zur Vorhersage der *figuralen Reasoning*-Leistung zum zweiten Messzeitpunkt der Versuchsgruppe 1 (F–F+) und Kontrollgruppe (F–F–)

	Prädiktoren	B	SE	β^{22}	t	p
Modell 1	Leistung in Test 1 (L)	.486	.08	.511	6.049**	.000
	experimentelle Bedingung zu Test 2 (F) (Feedback/kein Feedback)	-.154	.348	-.037	-.442	.659
Modell 2	Leistung in Test 1 (L)	.241	.062	.253	3.120**	.000
	experimentelle Bedingung zu Test 2 (F) (Feedback/kein Feedback)	-7.250	.735	-1.760	-9.864**	.000
	Produktterm (LxF)	.871	.085	1.859	10.246**	.000

Anmerkung: Modell 1 = einfache Regression; Modell 2 = multiple moderierte Regression; Kodierung der Dummy-Variable: Feedback = 1, kein Feedback = 0

Anhand Modell 1 wird das Ausbleiben eines generellen Treatmenteffekts deutlich ($\Delta R^2 = .001$, $p = .659$, $N = 107$, *intraindividuelle Feedback-Hypothese*). Unter Beachtung der Ausgangsleistung (feedbackfreie Leistung) lässt sich jedoch ein Effekt von Richtig/Falsch-Feedback ausmachen (Modell 2: $\Delta R^2 = .37$, $p < .001$, $N = 107$, $f^2 = .03$). Die Voraussetzung homogener Fehlervarianzen auf den Moderatorstufen²³ ist dabei erfüllt (*Error Variance Ratio* 1 : 1.1; Bartlett's $M = .66$, $p = .42$; Aguinis, 2004).

²² Beta-Gewichte von größer Null gelten als Hinweis auf Verletzung der Voraussetzung nichtkorrelierter Prädiktoren in Regressionsanalysen (Multikollinearitätsproblem). Durch Hinzunahme des Produktterms aus beiden Prädiktoren innerhalb einer multiplen moderierten Regression muss jedoch immer eine Interkorrelation der Prädiktoren bestehen. Dieses Problem bleibt in der Literatur zur multiplen moderierten Regression gewissermaßen ungeklärt (schon Jäger, 1978). Auch jüngere Veröffentlichungen (z.B. Aguinis, 2004; Baron & Kenny, 1986) behandeln die Problematik der Multikollinearität im Gegensatz zu anderen Fragen zur multiplen moderierten Regression, wie z.B. die Homoskedastizität von Typ I und II (Homogenität der Fehlervarianzen in der Gesamtstichprobe bzw. bei kategorialen Moderatoren in den Substichproben, d.h. auf den Moderatorstufen), nicht oder nur kurz und verweisen dabei auf Cronbach (1987). Cronbach (1987) stellte fest, dass Multikollinearität nicht die statistische Power der multiplen moderierten Regression reduziert. Aguinis (2004) frohlockte sogar: "Fortunately, the concerns about the detrimental effects of multicollinearity on MMR [multiple moderierte Regression, d.V.] have proven to be unfounded." (S. 77).

²³ Die Prüfung der Homogenität der Fehlervarianzen auf den Moderatorstufen wurde mit einer von Aguinis, Petersen & Pierce (1999) zur Verfügung gestellten Software (web: <http://carbon.cudenver.edu/~haguinis/mmr/>) vorgenommen, welche auf Basis der Standardabweichung des Kriteriums Y sowie der Korrelationen zwischen Prädiktor X und Kriterium Y in den anhand des hier kategorialen Moderators gebildeten Subgruppen sowie der Substichprobengröße Bartlett's M einschließlich des p -Wertes berechnet. Dieser wird herangezogen, um die Homogenität der Fehlervarianzen zu prüfen. Zudem wird das Verhältnis der größten und kleinsten Fehlervarianz in den Subgruppen bestimmt und mit einem von DeShon und Alexander (1996, zitiert in Aguinis, 2004) vorgeschlagenen Kriterium verglichen. Das besagt, dass der F -Test in der multiplen moderierten Regression beeinträchtigt wird, sobald die Fehlervarianz in einer der gebildeten Subgruppen mehr als 1.5 mal größer ist als die Fehlervarianz in einer anderen Subgruppe. Letztere Faustregel wird dem Bartlett-Test vorgezogen, wenn eine Normalverteilung der Daten nicht angenommen werden kann (vgl. Aguinis, 2004).

Zur graphischen Veranschaulichung der Wechselwirkung von Leistungsfähigkeit und Feedback auf die Testleistung wurden gruppenspezifisch unter- und überdurchschnittliche Leistungsscores für den Prädiktor (L_1) in die Regressionsgleichung (Gleichung 1) eingesetzt und für beide Moderatorstufen (Feedback/kein Feedback) der Kriteriumswert (L_2) berechnet (siehe Aguinis, 2004; Cohen, Cohen, West & Aiken, 2003).

$$L_2 = 6.37 + .241L_1 - 7.25F + .871(L \times F) \quad (1)$$

mit:

L_2 = Leistung zum Messzeitpunkt 2

L_1 = Leistung zum Messzeitpunkt 1

F = experimentelle Bedingung (Feedback/kein Feedback; Moderator)

($L \times F$) = Produktterm (Leistung zum Messzeitpunkt 1 x experimentelle Bedingung)

Abbildung 6.1 zeigt die figurale Reasoning-Leistung für Testanden mit unter- versus überdurchschnittlichen Testleistungen ($\bar{x} = 7.98$, $s = 1.99$) in Test 1 jeweils unter einer feedbackfreien Testbedingung und einer Testbedingung mit Feedback zum zweiten Messzeitpunkt. Es wird deutlich, dass insbesondere die bereits unter feedbackfreien Bedingungen überdurchschnittlich leistungsfähigen Testanden von Feedback profitierten. Sie konnten ihre Leistung um durchschnittlich 1.5 Aufgaben verbessern. Unterdurchschnittlich leistungsfähige Testanden verschlechterten sich hingegen weiter unter Feedback und zwar um durchschnittlich 2 Aufgaben. Es handelt sich hierbei um einen schwachen Effekt ($f^2 = .03$).

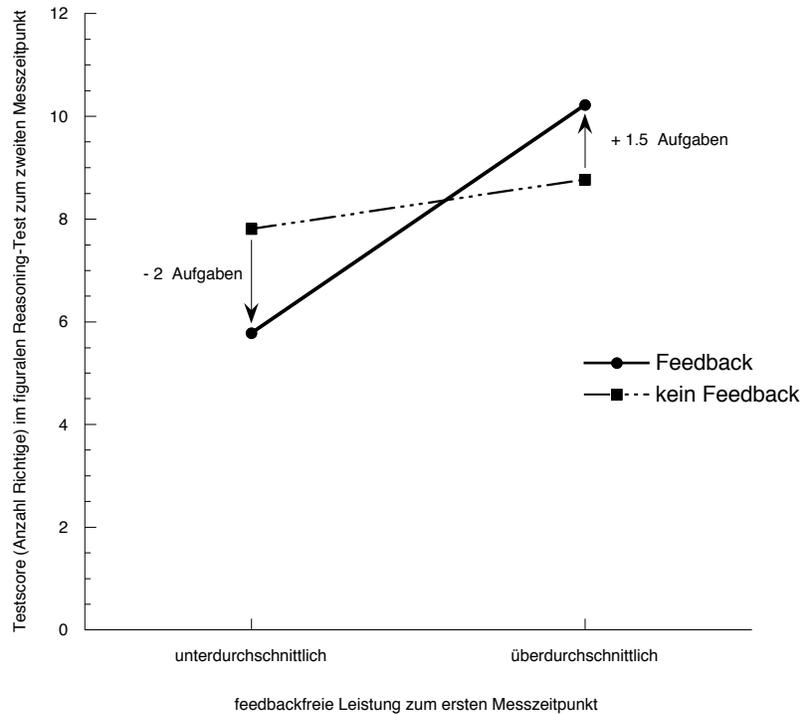


Abbildung 6.1 Figurale Reasoning-Leistung unter- versus überdurchschnittlich leistungsfähiger Testanden in einer Testbedingung mit versus ohne Feedback.

Zusammenfassend ist festzuhalten: Richtig/Falsch-Feedback ist nicht generell leistungsförderlich (*generelle Feedback-Hypothese*). Zum einen lässt sich erwartungswidrig kein genereller positiver Feedback-Effekt ausmachen. Für den Materialbereich Figurenfolgen ergaben sich sogar Hinweise, dass Richtig/Falsch-Feedback die Reasoning-Leistung negativ beeinflusst (Abschnitt 6.1). Zum anderen legen die Analysen eine ebenfalls domainspezifische (Figurenfolgen) Abhängigkeit des Treatmenteffekts von der Leistungsfähigkeit des Testanden nahe (*differenzielle Feedback-Hypothese*). Leistungsfähige Testanden profitieren von Richtig/Falsch-Feedback, während wenig leistungsfähige Testanden sich unter Feedback weiter verschlechtern. Je höher also die Ausgangsleistung der Testanden, desto höher ist der zu erwartende Feedback-Profit. Die Abhängigkeit des Feedback-Effekts von der Leistungsfähigkeit (differenzieller Feedback-Effekt) trägt dazu bei, dass kein genereller Effekt von Richtig/Falsch-Feedback bei der Bearbeitung von figuralen Reasoning-Aufgaben festgestellt werden konnte. Unter 6.2.2.1.4 sollen Gründe diskutiert werden, warum sich das Befundbild für die verschiedenen Materialbereiche (Figuren- und Zahlenfolgen) unterscheidet.

Lassen sich daneben auch *nicht-intellektuelle* Merkmale ausmachen, anhand derer die Vorhersage der Leistung in einer Testbedingung mit Feedback verglichen mit einer Testbedingung ohne Feedback verbessert werden kann? Dieser Analyse widmet sich der nächste Abschnitt. Da die psychometrischen Eigenschaften der *numerischen* Reasoning-Tests günstiger ausfielen als die der figuralen Reasoning-Tests, werden diese jeweils vorrangig betrachtet.

6.2.2 Feedback-Effekte und außerintellektuelle Personmerkmale

Ziel der Arbeit ist es, potentielle Effekte von Feedback auf die Leistung aufzudecken. Sinnvollerweise gilt es daher Leistungen, die unter Feedback erbracht werden, Leistungen unter feedbackfreien Bedingungen gegenüberzustellen. Im Rahmen der vorliegenden Studie stehen im Prinzip vier Leistungsvariablen zur Verfügung:

- a) Leistungen in Intelligenzstatustests (MidYIS, CAT),
- b) Leistungen in Reasoning-Tests (Figurenfolgen, Zahlenfolgen) ohne Feedback,
- c) Schulleistungen,
- d) Leistungen in Reasoning-Tests (Zahlenfolgen, Figurenfolgen) mit Feedback.

Diese vier Variablen lassen sich hinsichtlich der Bedeutsamkeit von Feedback beim Zustandekommen der Leistung klassifizieren. Während bei a) und b) Feedback keine Rolle spielt, stellen Schulleistungen (c) Performanzwerte dar, die im Rahmen eines mehr oder weniger kontinuierlichen Feedback-Prozesses erbracht wurden. Dies trifft in viel stärkerem und kontrolliertem Maße auf die Feedback-Testleistungen (d) zu. Diese Perspektive soll die im Folgenden realisierte Analysestrategie bestimmen. Die Tatsache, dass vier verschiedene Leistungsvariablen in die Analyse einbezogen werden können, ermöglicht in gewissem Sinne eine Kreuzvalidierung der Befunde. Dies sollte vor allem stichprobengrößen-bezogene Limitierungen der hier vorgestellten Studie kompensieren und somit die Verallgemeinerbarkeit der Befunde günstig beeinflussen.

Zunächst soll also geprüft werden, ob mit den hier verwendeten Messinstrumenten die von Dweck (2000) bzw. Nicholls et al. (1990) postulierten Modellannahmen zum Zusammenhang von Zielorientierung und *Schulleistung* replizierbar sind. Gleiches gilt für die von Meijer (2001) vertretenen Annahmen zum Zusammenhang von Testangst und Leistung. Schulleistungen werden dabei als Leistungsindikatoren aus einer „natürlichen“ Leistungssituation mit Feedback verstanden. Die Ergebnisse werden dann den Analysen zur Vorhersage der klassischen *feedbackfreien Intelligenztest*-Leistungen

gegenübergestellt. Unterschiede in den Ergebnissen werden auf einen unterschiedlichen Anteil an Feedback beim Zustandekommen der Schul- und Intelligenzstatustest-Leistungen zurückgeführt. Im nächsten Auswertungsschritt wird geprüft, ob sich das Modell zur Vorhersage der Schulleistungen auch auf *Reasoning*-Testleistungen unter Feedback-Bedingungen übertragen lässt. Um zu klären, ob potentielle differentielle Leistungseffekte für eine Testbedingung mit Feedback charakteristisch sind, werden dieselben Analysen für die feedbackfreie Reasoningtest-Bedingung vorgenommen. Unterschiede in den Ergebnissen der Analysen werden auch hier auf den unterschiedlichen Anteil an Feedback beim Zustandekommen der Testleistung in den experimentellen Bedingungen zurückgeführt. Dazu wird zunächst die Zahlenfolgen-Bedingung betrachtet, da die Reliabilität dieser Tests akzeptabel war. Dann wird geprüft, ob sich die Befunde in der Figurenfolgen-Bedingung replizieren lassen.

6.2.2.1 Zielorientierung

Untersuchungen zum Konstrukt Zielorientierung dienten vordergründig der Analyse der motivationalen Faktoren, die *schulische* Leistungen (mit)bestimmen und wurden überwiegend im Schulsetting durchgeführt (z.B. Dweck, 1986; Nicholls, 1984, im deutschsprachigen Raum z.B. Broome, 2001; Dresel, 2001; Ziegler & Schober, 1999; siehe Abschnitt 2.2.5). In einem ersten Schritt soll deshalb die Wechselbeziehung zwischen schulbezogener Zielorientierung und schulbezogenem Begabungskonzept in Bezug auf die Schulleistungen untersucht werden. Es soll geklärt werden, ob sich in der vorliegenden Stichprobe mit den hier verwendeten Messinstrumenten theoriekonform Schulleistungen anhand der Wechselwirkung von Zielorientierung und Begabungskonzept vorhersagen lassen. Schulleistungen (Schulnoten) empfehlen sich für die Analyse der schulischen Leistungsfähigkeit unter Feedback-Bedingungen, da Schulleistungen das Resultat einer „natürlichen“ Leistungssituation mit kontinuierlichem Leistungsfeedback darstellen. Es soll dann geprüft werden, ob dieses Modell auch auf Testsituationen mit Feedback-Angebot übertragen werden kann.

Nach Dweck (2000, Dweck & Leggett, 1988) wird angenommen, dass leistungszielorientierte Personen mit geringem Begabungskonzept unter Misserfolgsbedingungen Leistungseinbußen aufweisen, im Gegensatz zu lernzielorientierten Personen oder Personen mit hohem Begabungskonzept (siehe auch z.B. Nicholls, 1984). Im Rahmen eines Moderatoransatzes lässt sich das Begabungskonzept hier als eine *Moderatorvariable* für den Zusammenhang von Zielorientierung und Leistung konzeptualisieren. Eine

Leistungszielorientierung bedingt also nicht per se Leistungsdefizite, sondern nur bei einer bestimmten Ausprägung des Begabungskonzepts (*Moderatorhypothese Begabungskonzept*).

6.2.2.1.1 Schulleistungen

Zur Prüfung dieser Hypothese werden als Prädiktoren die Variablen Leistungszielorientierung (Faktorwerte des Faktors Ich-Orientierung), als potentieller Moderator das Begabungskonzept (Faktorwerte) sowie deren Produktterm herangezogen, um die standardisierten Schulleistungen vorherzusagen. Zur Kontrolle des Leistungsniveaus geht außerdem die Leistung im Mathematikunterricht als ein wesentlicher Prädiktor von Schulleistungen in die Vorhersage ein. Eine Moderation der Beziehung von Leistungszielorientierung und Schulleistungen durch das Begabungskonzept unabhängig vom mathematischen Leistungsniveau sollte sich in einem von Null signifikant verschiedenen Beta-Gewicht des Produktterms zeigen ($H_0: \beta_{\text{Produkt}} = 0$). Tabelle 6.7 gibt einen Überblick über die Regressionskoeffizienten der entsprechenden Regressionsanalysen in den verschiedenen experimentellen Bedingungen.

Es lässt sich anhand des von Null signifikant verschiedenen Beta-Gewichts des Produktterms (BxZ) zeigen, dass das Begabungskonzept den Zusammenhang von Leistungszielorientierung und Schulleistung moderiert, und zwar in der Substichprobe, die die Zahlenfolgen bearbeitet hat ($\Delta R^2 = .02$, $R^2 = .63$, $F_{4/103} = 44.441$, $p < .001$, $N = 104$), als auch in der Gesamtstichprobe ($\Delta R^2 = .01$, $R^2 = .65$, $F_{4/241} = 112.135$, $p < .001$, $N = 242$).

Tabelle 6.7

Kennwerte der Koeffizienten der Regression von Schulleistungen auf die Leistung im Mathematikunterricht und die schulbezogene Leistungszielorientierung, moderiert durch das schulbezogene Begabungskonzept

<i>experimentelle Bedingung</i>	<i>Prädiktoren</i>	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Zahlenfolgen	Mathematiknote	.511	.043	.77	11.943**	.000
	Zielorientierung (Z)	-.815	.686	-.075	-1.187	.238
	Begabungskonzept (B)	.681	.658	.067	1.035	.303
	Produktterm (BxZ)	-1.060	.499	-.129	-2.125*	.036
Zahlen- und Figurenfolgen	Mathematiknote	.544	.027	.784	19.840**	.000
	Zielorientierung (Z)	-.515	.415	-.050	-1.289	.199
	Begabungskonzept (B)	.804	.421	.076	1.91	.057
	Produktterm (BxZ)	-.746	.357	-.08	-2.092*	.038

Anmerkung: In die Analysen gehen nur Personen mit vollständigen Datensätzen ein. Abhängige Variable: Schulleistungen (standardisiert). In der gemeinsamen Analyse der Zahlen- und Figurenfolgen ist Fall 103 eliminiert, da der Standardfehler des Residuals = -3.245 beträgt. Für die Merkmale Leistungszielorientierung und Begabungskonzept wurden Faktorwerte herangezogen.

Da zufällig mehr Schülerinnen als Schüler die Zahlenfolgen bearbeitet haben, wurde zudem geprüft, ob der Zusammenhang in der hier untersuchten Stichprobe insbesondere für Schüler und weniger für Schülerinnen gilt (Tabelle 6.8). Eine nach Geschlecht getrennte Analyse der Moderatorwirkung der Variable Begabungskonzept deckt auf, dass der gefundene Moderatoreffekt tatsächlich auf die männliche Stichprobe zurückzuführen ist (männliche Substichprobe: $\Delta R^2 = .03$, $R^2 = .66$, $F_{4/124} = 60.665$, $p < .001$, $N = 125$; weibliche Substichprobe: $R^2 = .66$, $F_{4/117} = 57.975$, $p < .001$, $N = 118$). Die Regressionskoeffizienten unterscheiden sich dabei signifikant (Konfidenzintervall für die Differenz der Regressionskoeffizienten für $\alpha = 5\%$: .08 bis 3.16, Cohen et al., 2003). Insofern stellt das Geschlecht einen weiteren Moderator der Beziehung von Zielorientierung und Schulleistung dar (siehe Moderatoranalysen für kategoriale Moderatoren, Aguinis, 2004; Baron & Kenny, 1986).

Tabelle 6.8

Kennwerte der Koeffizienten der Regression von Schulleistungen auf die Leistung im Mathematikunterricht und die schulbezogene Leistungszielorientierung, moderiert durch das schulbezogene Begabungskonzept – geschlechtsspezifische Analyse in der Gesamtstichprobe

Substichprobe	Prädiktoren	B	SE	β	t	p
männlich (N = 125)	Mathematiknote	.512	.038	.748	13.624**	.000
	Zielorientierung (Z)	.761	.670	.067	1.136	.258
	Begabungskonzept (B)	1.680	.546	.170	3.076**	.003
	Produktterm (BxZ)	-1.978	.570	-.207	-3.467**	.001
weiblich (N = 118)	Mathematiknote	.552	.039	.781	14.105**	.000
	Zielorientierung (Z)	-.615	.617	-.059	-.997	.321
	Begabungskonzept (B)	1.691	.721	.133	2.347*	.021
	Produktterm (BxZ)	-.359	.540	-.039	-.665	.507

Anmerkung: analog zu Tabelle 6.7

Abbildung 6.2 stellt die Wechselwirkung von Leistungszielorientierung und Begabungskonzept bei der Vorhersage der Schulleistungen für durchschnittlich leistungsfähige Schüler dar.

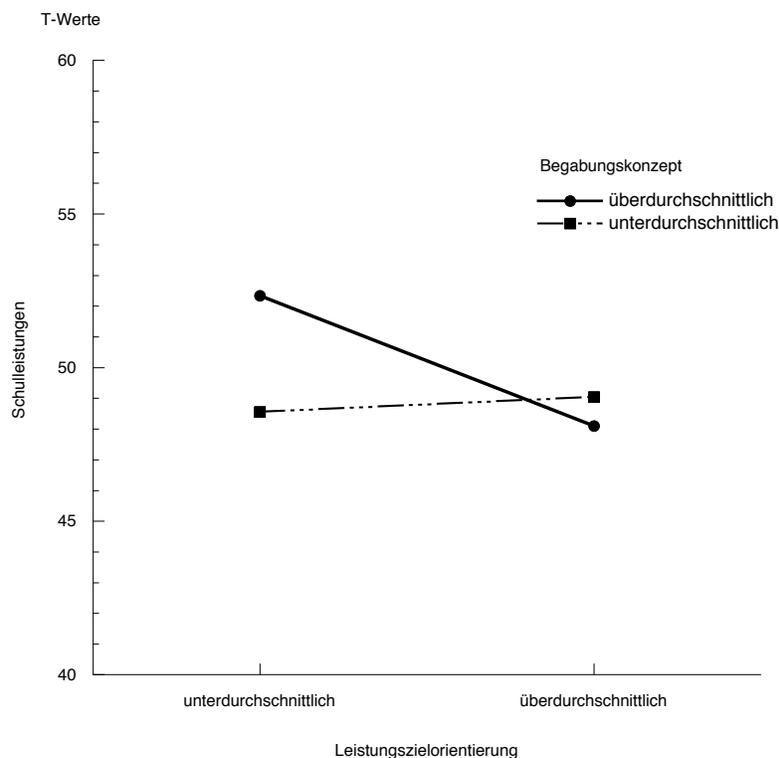


Abbildung 6.2 Schulleistungen (alle Fächer) unter- versus überdurchschnittlich leistungszielorientierter Schüler mit unter- versus überdurchschnittlichem Begabungskonzept bei durchschnittlicher schulischer Leistungsfähigkeit (Leistung im Mathematikunterricht).

Die schulische Leistungsfähigkeit wurde durch Konstanthalten der Mathematikleistung (T -Norm = 50) bei Berechnung des Kriteriumswertes anhand der Regressionsgleichung für Schüler mit gruppenspezifisch unter- versus überdurchschnittlichen Testwerten in den Prädiktoren Begabungskonzept ($\bar{x} = .10, s = 1.09$) und Leistungszielorientierung ($\bar{x} = .03, s = 1.02$) kontrolliert (analog zu Abschnitt 6.2.1.2). Die Gruppenunterschiede lassen sich folglich nicht auf eine unterschiedliche Leistungsfähigkeit der Schüler zurückführen. Theoriekonform zeigten nicht leistungszielorientierte Schüler mit überdurchschnittlichem Begabungskonzept die höchsten Schulleistungen. Die nach Dweck (1986) und Nicholls (1984) definierte Risikogruppe – hohe Leistungszielorientierung bei geringem Begabungskonzept – fiel hier jedoch nicht durch Leistungsdefizite auf. Eine protektive Wirkung eines überdurchschnittlichen Begabungskonzepts bei überdurchschnittlicher Leistungszielorientierung ließ sich ebenfalls nicht feststellen. Erwartungsgemäß kann aber geschlussfolgert werden: Je eher Schüler ihren Fähigkeiten vertrauten und je seltener sie Leistungsziele verfolgten, desto besser waren ihre Schulleistungen.

6.2.2.1.2 Intelligenztest-Leistungen

Hinsichtlich der Vorhersage der Leistungsvarianz in einem Intelligenzstatustest (CAT, MidYIS)²⁴ findet sich dagegen keine Wechselwirkung von Zielorientierung (Faktorwerte des Faktors Ich-Orientierung) und Begabungskonzept (Faktorwerte). Dieser Befund gilt für die Stichprobe, die die Zahlenfolgen bearbeitet hat ($R^2 = .42, F_{4/102} = 19.707, p < .001, N = 103$; Tabelle 6.9) als auch für die Gesamtstichprobe ($R^2 = .44, F_{4/237} = 46.630, p < .001, \Delta R^2 = .001, F_{1/233} = .591, p = .443, N = 238$). Unter den gegebenen Stichprobenverhältnissen kann in diesem Kontext mit hinreichender statistischer Sicherheit ($1 - \beta \geq .80$) ein mittlerer bis schwacher differentieller Effekt (genauer ein Effekt von $f^2_{\text{Zahlenfolgen}} \geq .08$ und $f^2_{\text{Gesamt}} \geq .04$) ausgeschlossen werden. Tabelle 6.9 zeigt die Ergebnisse der Regressionsanalyse für die Substichprobe der Bedingung Zahlenfolgen.

²⁴ Für beide Klassenstufen liegen jeweils Testwerte in verschiedenen Intelligenztests vor (CAT, MidYIS, siehe Abschnitt 5.3). Da die Leistungen in beiden Tests als IQ-Skala (100 +/- 15, nationale Normen) normiert sind, wurde für die Analysen eine gemeinsame Variable gebildet.

Tabelle 6.9

Kennwerte der Koeffizienten der Regression der intellektuellen Leistung auf die Schulleistung und die schulbezogene Leistungszielorientierung, moderiert durch das schulbezogene Begabungskonzept

Prädiktoren	B	SE	β	t	p
Schulleistung	.685	.085	.647	8.034**	.000
Leistungszielorientierung (Z)	-1.219	.904	-.105	-1.349	.180
Begabungskonzept (B)	.764	.872	.071	.876	.383
Produktterm (BxZ)	-.162	.672	-.018	-.240	.810

Anmerkung: Die Signifikanzprüfung der Beta-Gewichte ändert sich auch dann nicht, wenn anstatt der durchschnittlichen Schulleistung (alle Fächer), lediglich die Leistung im Mathematikunterricht herangezogen wird.

Als Kriterium wurde hier die Leistung in einem anfallenden Intelligenzstatustest herangezogen, die traditionell ohne Feedback erbracht wurde. Im Gegensatz dazu resultieren Schulleistungen aus der Auseinandersetzung des Lehrers mit dem Schüler und umgekehrt, also in einer Leistungssituation mit kontinuierlichem Leistungsfeedback. Vor dem Hintergrund dieser Überlegungen scheint es nur konsequent, dass Zielorientierung und Begabungskonzept in ihrer Wechselbeziehung keinen Vorhersagebeitrag für die Leistungen in traditionellen, feedbackfreien Intelligenztests leisten, auch dann nicht, wenn im Sinne der Konsistenz (vorangestellter Abschnitt) die Leistungsfähigkeit anhand der Variable Schulleistung kontrolliert wird.

6.2.2.1.3 Numerische Reasoning-Testleistung

Im Rahmen der hier zu untersuchenden Fragestellung ist nun interessant zu prüfen, ob sich *Reasoning*-Testleistungen (Zahlenfolgen, Figurenfolgen) ebenfalls durch eine Wechselwirkung von Zielorientierung und Begabungskonzept vorhersagen lassen. Dabei wird erwartet, dass gerade in der *Feedback*-Bedingung eine solche Wechselbeziehung von Leistungszielorientierung und Begabungskonzept im Leistungsverhalten zum Ausdruck kommt. Dahinter steht die Annahme, dass durch Feedback im Testverlauf Erfolgs- und Misserfolgserleben ausgelöst wird. Ein leistungshinderlicher Umgang mit Misserfolg wird im Rahmen der Theorie zur Zielorientierung (Dweck, 1986; Nicholls, 1984) von leistungszielorientierten Schülern mit geringem Begabungskonzept erwartet (siehe Abschnitt 2.2.5).

Zur Prüfung der Hypothese, dass der Zusammenhang der Variablen Zielorientierung und Reasoning-Leistung durch die Variable Begabungskonzept in einer Testbedingung mit Feedback moderiert wird (*Moderatorhypothese Begabungskonzept*), nicht aber

in einer feedbackfreien Testbedingung, wird die *Versuchsgruppe 1* (F– F+) betrachtet. Neben der Leistungszielorientierung und dem Begabungskonzept, sowie deren Produktterm geht als weiterer Prädiktor und zur Kontrolle der Leistungsfähigkeit die Leistung in einem Intelligenzstatustest²⁵ in die Regressionsanalyse ein. Das Kriterium ist die numerische Reasoning-Leistung in der *Feedback*-Bedingung (Test 2). Um die Ergebnisse einer feedbackfreien Testbedingung gegenüberzustellen, soll mit denselben Prädiktoren ebenso die *feedbackfreie* Reasoning-Leistung zum ersten Messzeitpunkt vorhergesagt werden. Die Ergebnisse beider Regressionsanalysen werden verglichen. Ein Moderatoreffekt sollte sich in den Regressionsanalysen jeweils in einem signifikant von Null verschiedenen Beta-Gewicht des Produktterms (BxZ) zeigen ($H_0: \beta = 0$).

Die Tabellen 6.10 und 6.11 zeigen, dass bei Verwendung der trait-näheren *schulbezogenen* Persönlichkeitsmaße das Begabungskonzept (Faktorwerte) den Zusammenhang von Zielorientierung (Faktorwerte des Faktors Ich-Orientierung) und numerischer Reasoning-Testleistung nicht moderiert, und zwar weder in der feedbackfreien Bedingung (Test 1: $R^2 = .45$, $F_{4/48} = 10.806$, $p < .001$, $N = 49$, Tabelle 6.11) noch in der Feedback-Bedingung (Test 2: $R^2 = .42$, $F_{4/48} = 9.683$, $p < .001$, $N = 49$; Tabelle 6.10).

²⁵ Anstatt der Leistung im ersten feedbackfreien Test (Versuchsgruppe 1, F– F+) wurde hier als Indikator für die feedbackfreie Leistungsfähigkeit die Leistung in einem Intelligenzstatustest herangezogen. Auf diese Weise können dieselben Variablen als Prädiktoren in den beiden Regressionsanalysen berücksichtigt werden, die zur Vorhersage der *feedbackbedingten* Reasoning-Leistung in Test 2 bzw. zur Vorhersage der *feedbackfreien* Reasoning-Leistung in Test 1 dienen sollen. Die Ergebnisse der Regressionsanalysen werden verglichen. Unterschiede in den Ergebnissen lassen sich dann einzig auf das Kriterium (feedbackbedingte versus feedbackfreie Reasoning-Leistung) zurückführen, da ansonsten dieselben Variablen in die Analysen eingehen.

Tabelle 6.10

Kennwerte der Koeffizienten der Regressionen der numerischen Reasoning-Leistung unter *Feedback* in Versuchsgruppe 1 (F-F+) auf die intellektuelle Leistungsfähigkeit sowie die Leistungszielorientierung, moderiert durch das Begabungskonzept unter einer schulbezogenen sowie situationsspezifischen Perspektive

Prädiktoren		<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
schulbezogene Maße	Intelligenztest-Leistung	.132	.024	.626	5.441**	.000
	Leistungszielorientierung (Z)	.811	.320	.298	2.535*	.015
	Begabungskonzept (B)	-.077	.287	-.034	-.270	.789
	Produktterm (BxZ)	-.028	.224	-.014	-.126	.900
situationsspezifische Maße vor Test 2	Intelligenztest-Leistung	.133	.022	.618	6.075**	.000
	Leistungszielorientierung (Z)	1.161	.544	.593	2.135*	.038
	Begabungskonzept (B)	1.308	.468	.894	2.792*	.008
	Produktterm (BxZ)	-.203	.101	-.936	-2.012*	.050

Anmerkung: Die schulbezogenen Maße stellen Faktorwerte dar. Als situationsspezifische Maße gehen die Items „I feel pretty confident that I shall be able to solve most of the problems“ versus „I do not feel confident that I shall be able to solve most of the problems.“ (Begabungskonzept) sowie „I think I will have done well on this test if I get more correct answers than other students.“ versus „I do not think that I will have done well unless I get more correct answers than other students.“ (Leistungszielorientierung, Nicholls), die jeweils vor der Testsitzung appliziert wurden, in die Analyse ein.

Werden jedoch die auf die konkrete Testsituation bezogenen Maße zur *situations-spezifischen* Zielorientierung und zum *situationsspezifischen* Begabungskonzept herangezogen, die jeweils unmittelbar vor der Reasoning-Testung erfasst wurden, lässt sich ein Moderatoreffekt aufdecken, und zwar erwartungsgemäß nur in der Feedback-Bedingung ($\Delta R^2 = .04$, $R^2 = .50$, $F_{4/49} = 13.123$, $p < .001$, $N = 50$, Tabelle 6.10). Die Nullhypothese kann also abgelehnt werden ($H_0: \beta_{\text{Produkt}} = 0$).

Für die *feedbackfreie* Bedingung derselben Stichprobe (Test 1) kann auch unter Verwendung der situationsspezifischen Maße kein entsprechender Moderatoreffekt nachgewiesen werden ($R^2 = .54$, $F_{4/49} = 15.38$, $p < .001$, $N = 50$; Tabelle 6.11). Die Regressionskoeffizienten unterscheiden sich dabei signifikant (Konfidenzintervall der Differenz der Regressionskoeffizienten für $\alpha = 5\%$: .06 bis .59, Cohen et al., 2003). Unter Beachtung der gegebenen Stichprobenverhältnisse kann mit hinreichender statistischer Sicherheit ($1 - \beta \geq .80$) jedoch lediglich ein mittlerer und starker Effekt (genauer $f^2 \geq .17$) der Wechselwirkung von Begabungskonzept und Leistungszielorientierung auf die Reasoning-Testleistung in der feedbackfreien Bedingung ausgeschlossen werden.

Tabelle 6.11

Kennwerte der Koeffizienten der Regressionen der numerischen Reasoning-Leistung in der *feedbackfreien* Testbedingung in Versuchsgruppe 1 (F- F+) auf die intellektuelle Leistungsfähigkeit sowie die Leistungszielorientierung, moderiert durch das Begabungskonzept unter einer schulbezogenen sowie situationsspezifischen Perspektive

Prädiktoren		<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
schulbezogene Maße	Intelligenztest-Leistung	.113	.024	.522	4.661**	.000
	Leistungszielorientierung (Z)	.567	.321	.202	1.768	.084
	Begabungskonzept (B)	.598	.287	.253	2.081*	.043
	Produktterm (BxZ)	.093	.224	.046	.415	.680
situationsspezifische Maße erhoben vor Test 1	Intelligenztest-Leistung	.115	.022	.529	5.163**	.000
	Leistungszielorientierung (Z)	-.528	.479	-.363	-1.103	.276
	Begabungskonzept (B)	.236	.367	.128	.643	.524
	Produktterm (BxZ)	.124	.091	.555	1.361	.180

Anmerkung: analog Tabelle 6.10

Es fällt weiter auf, dass mittels der *situationsspezifischen* Maße eine höhere Varianzaufklärung im Kriterium möglich wird. Dieser Befund ist aus ökonomischer Perspektive interessant, da die situationsspezifischen Maße je aus nur einem Item bestehen, während die schulbezogenen Maße Skalenwerte (je 8 Items per Skala) darstellen.

Auch unter Verwendung des Dweckschen Items zur Erfassung der situationsspezifischen Zielorientierung als eindimensionales Konstrukt („*What is important to me is getting a higher score on the test than everyone else.*“ versus „*What is important to me is learning how to solve the problems.*“) lässt sich in diesem Kontext ein Moderatoreffekt der Variablen Begabungskonzept nachweisen ($R^2 = .50$, $F_{4/49} = 13.100$, $p < .001$, $\Delta R^2 = .04$, $F_{1/45} = 4.284$, $p < .05$, $N = 50$). Das Begabungskonzept moderiert also den Zusammenhang von Zielorientierung und Leistung in einer Testbedingung mit Feedback, unabhängig von der Konzeptualisierung des Konstrukts Zielorientierung.

Es soll nun geprüft werden, ob der anhand der *situationsspezifischen* Maße gefundene Moderatoreffekt in der Feedback-Bedingung in Versuchsgruppe 1 (F- F+) tatsächlich auf das Treatment (Feedback) und nicht auf die *Wiederholungsmessung* zurückzuführen ist. Dazu wird die Kontrollgruppe (F- F-) gemeinsam mit der Versuchsgruppe 2 (F+ F-) betrachtet. Beide Gruppen haben zum zweiten Messzeitpunkt eine feedbackfreie Testversion bearbeitet. Wenn eine Moderatorfunktion des Begabungskonzepts bezogen auf den Zusammenhang von Zielorientierung und Leistung allein durch die Testwiederholung bedingt ist, sollte zum zweiten Messzeitpunkt, also in der feedbackfreien Testbedingung, ein entsprechender Effekt auch in der Versuchsgruppe 2 (F+ F-) und

Kontrollgruppe (F– F–) auftreten. Beide Gruppen werden zu diesem Zweck gemeinsam betrachtet ($N = 54$).

Tabelle 6.12 zeigt die Koeffizienten der Regressionsanalyse für diese beiden Gruppen ($R^2 = .23$, $F_{4/53} = 4.985$, $p < .05$, $N = 54$). Anhand des statistisch nicht bedeutsamen Beta-Gewichts des Produktterms (BxZ) lässt sich schlussfolgern, dass in einer feedbackfreien Testbedingung auch bei Wiederholungsmessung kein Moderatoreffekt der Variablen Begabungskonzept auftritt. Es kann aber unter den gegebenen Stichprobenverhältnissen lediglich ein mittlerer und starker Effekt (genauer ein Effekt von $f^2 \geq .15$) der Wechselwirkung von Begabungskonzept und Leistungszielorientierung auf die feedbackfreie Testleistung mit hinreichender statistischer Sicherheit ($1 - \beta \geq .80$) ausgeschlossen werden.

Tabelle 6.12

Kennwerte der Koeffizienten der Regression der numerischen Reasoning-Leistung in der *feedbackfreien* Testbedingung bei Wiederholungsmessung auf die intellektuelle Leistungsfähigkeit sowie die Leistungszielorientierung, moderiert durch das Begabungskonzept unter einer situationsspezifischen Perspektive

Prädiktoren	B	SE	β	t	p
Intelligenztest-Leistung	.09	.034	.340	2.619*	.012
Leistungszielorientierung (Z)	.17	.641	.010	.026	.970
Begabungskonzept (B)	.546	.582	.329	.938	.353
Produktterm (BxZ)	.006	.125	.024	.047	.963

Die Variable Begabungskonzept moderiert hypothesenkonform den Zusammenhang von Leistungszielorientierung und Reasoning-Leistung in einer Testung mit Feedback, nicht jedoch in einer feedbackfreien Testung. Die Annahme, dass dieser Moderatoreffekt auf die Wiederholungsmessung zurückzuführen ist, ist nicht gerechtfertigt.

Abbildung 6.3 bietet eine graphische Veranschaulichung der Wechselwirkung von situationsspezifischer Leistungszielorientierung und situationsspezifischem Begabungskonzept bezogen auf die numerische Reasoning-Leistung. Dazu wurden analog zu Abschnitt 6.2.1.2 gruppenspezifisch unter- und überdurchschnittliche Testwerte für die Prädiktoren Leistungszielorientierung ($\bar{x} = 4.54$, $s = 1.39$) und Begabungskonzept ($\bar{x} = 4.65$, $s = 1.86$) in die Regressionsgleichung eingesetzt. Die intellektuelle Leistungsfähigkeit wurde dabei mit dem gruppenspezifisch durchschnittlichen Testwert (IQ = 104.8) konstant gehalten.

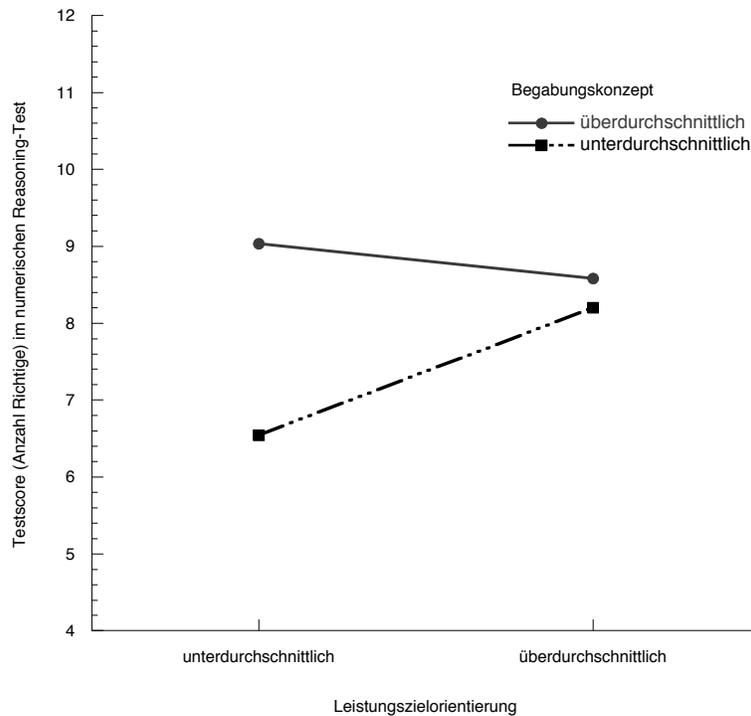


Abbildung 6.3 Numerische Reasoning-Leistung unter- versus überdurchschnittlich leistungszielorientierter Testanden mit unter- versus überdurchschnittlichem Begabungskonzept bei durchschnittlicher intellektueller Leistungsfähigkeit unter *Feedback*-Bedingungen.

Für einen Vergleich zeigt Abbildung 6.4 die *nicht* signifikante Wechselwirkung von situationsspezifischer Leistungszielorientierung und situationsspezifischem Begabungskonzept bezogen auf die numerische Reasoning-Leistung derselben Stichprobe unter *feedbackfreien* Testbedingungen (Versuchsgruppe 1, F- F+, Test 1). Dazu wurden wieder gruppenspezifisch unter- und überdurchschnittliche Testwerte für die Prädiktoren Leistungszielorientierung ($\bar{x} = 4.3$, $s = 1.89$) und Begabungskonzept ($\bar{x} = 4.89$, $s = 1.5$) in die Regressionsgleichung eingesetzt. Die intellektuelle Leistungsfähigkeit wurde ebenfalls mit dem durchschnittlichen Testwert (IQ = 104.8) konstant gehalten. Die Pfeile zeigen das unter Feedback-Bedingungen erreichte Leistungsniveau der jeweiligen Subgruppen an.

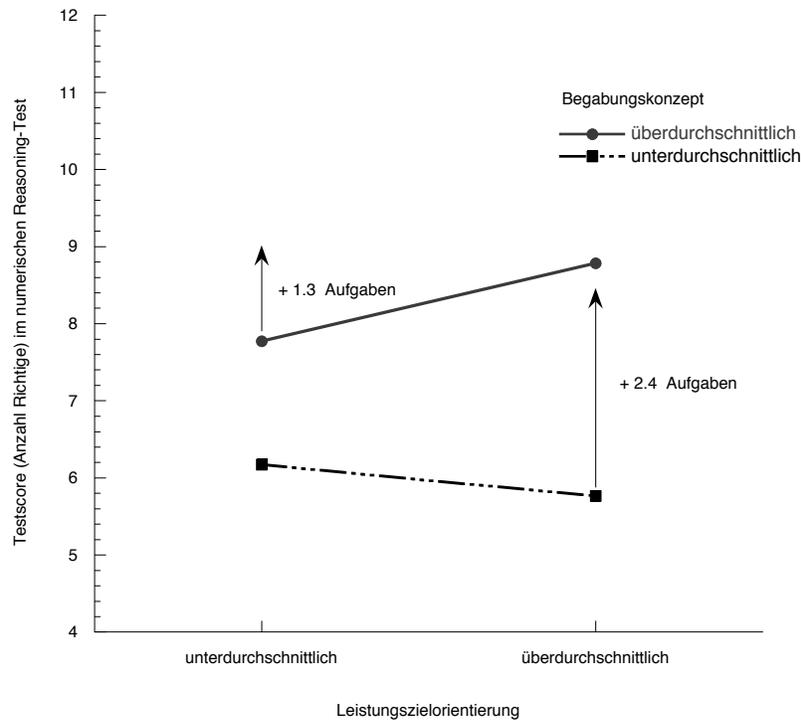


Abbildung 6.4 Numerische Reasoning-Leistung unter- versus überdurchschnittlich leistungszielorientierter Testanden mit unter- versus überdurchschnittlichem Begabungskonzept bei durchschnittlicher intellektueller Leistungsfähigkeit in der *feedbackfreien* Testung.

Hypothesenkonform ist ein hohes Begabungskonzept vor Testbeginn generell leistungsförderlich, und zwar unabhängig von der Zielorientierung (Abbildung 6.3 und Abbildung 6.4). Begabungskonzeptstarke Testanden, die keine Leistungsziele verfolgten, profitierten zusätzlich von Feedback (intraindividuell 1.3 Aufgaben, vgl. Abbildung 6.3 mit Abbildung 6.4). Auffällig ist jedoch der Feedback-Gewinn *leistungszielorientierter* Testanden mit unterdurchschnittlichem Begabungskonzept (intraindividuell 2.4 Aufgaben, interindividuell 1.7 Aufgaben). Dieser Befund steht auch im Widerspruch zu den Ergebnissen bei Vorhersage der Schulleistungen. Besitzt ein Testand bei Testbeginn ein geringes Begabungskonzept, scheint ihn unter Feedback-Bedingungen erwartungswidrig eine *wettbewerbsorientierte* motivationale Grundhaltung (Leistungszielorientierung: „*I think I will have done well on this test if I get more correct answers than other students.*“) vor einem Leistungsabfall zu schützen, und zwar unabhängig von seiner Leistungsfähigkeit.

6.2.2.1.4 Figurale Reasoning-Testleistung

Da in Abschnitt 6.2.2.1.1 das Geschlecht als ein Moderator des Zusammenhangs von schulbezogener Zielorientierung und Begabungskonzept bei Vorhersage der Schulleistung ausgemacht werden konnte, erfolgt die Analyse der figuralen Reasoning-Leistung

für die gesamte Stichprobe vollständiger Datensätze sowie für eine Zufallsstichprobe, in der das Verhältnis von Mädchen und Jungen mit dem der Zahlenfolgen-Stichprobe übereinstimmt. Es soll geprüft werden, ob sich die Ergebnisse für die Zahlenfolgen-Stichprobe in einer Zufallsstichprobe replizieren lassen, die in ihrer Zusammensetzung bezüglich Geschlecht der Zahlenfolgen-Stichprobe entspricht. Eine einfache Gegenüberstellung einer Stichprobe weiblicher und männlicher Testanden ist aufgrund des zu geringen Stichprobenumfangs beider Substichproben für eine Regressionsanalyse mit vier Prädiktoren nicht angemessen.

In der Gesamtstichprobe der Bedingung Figurenfolgen ($N = 73$) als auch in der potentielle Geschlechtseffekte kontrollierenden Substichprobe ($N = 61$) lässt sich kein Moderatoreffekt der Variablen Begabungskonzept auf den Zusammenhang von Zielorientierung und Reasoning-Leistung nachweisen (Substichprobe $N = 61$: $R^2 = .20$, $F_{4/60} = 4.631$, $p < .01$, $\Delta R^2 = .001$, $F_{1/56} = .079$, $p = .78$). Dieses Ergebnis besteht bei Verwendung der Maße zur Erfassung des *schulbezogenen* Begabungskonzepts und der *schulbezogenen* Zielorientierung (Gesamtstichprobe Figurenfolgen: $R^2 = .18$, $F_{4/72} = 4.943$, $p = .001$, $N = 73$, Tabelle 6.13), als auch bei Verwendung der entsprechenden *situationsspezifischen* Maße (Gesamtstichprobe Figurenfolgen: $R^2 = .14$, $F_{4/72} = 3.857$, $p = .007$, $N = 73$, Tabelle 6.13). Die Nullhypothesen müssen beibehalten werden ($H_0: \beta_{\text{Produkt}} = 0$). Unter den gegebenen Stichprobenverhältnissen ist in diesem Kontext ein mittlerer und starker Moderatoreffekt ($f^2_{\text{Substichprobe}} \geq .14$ and $f^2_{\text{Gesamt}} \geq .11$) der Variablen Begabungskonzept mit hinreichender statistischer Sicherheit ($1 - \beta \geq .80$) auszuschließen.

Tabelle 6.13

Kennwerte der Koeffizienten der Regressionen der figuralen Reasoning-Leistung unter *Feedback* in Versuchsgruppe 1 (F– F+) auf die intellektuelle Leistungsfähigkeit sowie die Leistungszielorientierung, moderiert durch das Begabungskonzept unter einer schulbezogenen sowie situationsspezifischen Perspektive

Prädiktoren		B	SE	β	t	p
schulbezogene Maße	Intelligenztest-Leistung	.067	.017	.463	4.050**	.000
	Leistungszielorientierung (Z)	.057	.191	.033	.298	.767
	Begabungskonzept (B)	-.0003	.219	.000	-.001	.999
	Produktterm (BxZ)	.414	.230	.205	1.795	.077
situationsspezifische Maße erhoben vor Test 2	Intelligenztest-Leistung	.063	.016	.431	3.818**	.000
	Leistungszielorientierung (Z)	-.136	.435	-.126	-.313	.755
	Begabungskonzept (B)	-.122	.421	-.095	-.289	.774
	Produktterm (BxZ)	.032	.080	.194	.404	.687

Anmerkung: analog Tabelle 6.10

Als eine potentielle Ursache für das Ausbleiben des Moderatoreffekts ließen sich die in diesem Kontext problematischen Reliabilitätskennwerte der Figurenfolgentests heranziehen, die das Aufdecken derartiger Effekte erschweren (siehe Abschnitt 5.4).

Analog zu den Analysen zum Zahlenfolgentest (vorangestellter Abschnitt) wurde weiter geprüft, ob sich bei der Vorhersage der *feedbackfreien* figuralen Reasoning-Leistung in Versuchsgruppe 1 (F- F+) ein Moderatoreffekt des Begabungskonzepts für den Zusammenhang von Zielorientierung und Leistung zum ersten Messzeitpunkt (kein Feedback) ausmachen lässt. Diese Analyse scheiterte schon im ersten Schritt, da bereits das moderatorenfreie Regressionsmodell die empirischen Daten nur unzureichend widerspiegelte (Versuchsgruppe 1 [F- F+] erster Messzeitpunkt: $R^2 = .07$, $F_{4/73} = 2.259$, $p = .072$; Versuchsgruppe 2 [F+ F-] und Kontrollgruppe [F- F-] zum zweiten Messzeitpunkt: $R^2 = .20$, $F_{4/65} = 5.087$, $p = .001$, $\Delta R^2 = .02$, $F_{1/61} = 1.287$, $p = .261$, $N = 66$).

Die Variable Begabungskonzept moderiert nicht den Zusammenhang von Zielorientierung und figuraler Reasoning-Leistung, und zwar weder unter Feedback- noch unter feedbackfreien Testbedingungen.

Zusammenfassend kann festgehalten werden: Hypothesenkonform moderiert das Begabungskonzept den Zusammenhang von Leistungszielorientierung und *numerischer* Reasoning-Leistung unter Feedback-Bedingungen, nicht jedoch unter feedbackfreien Testbedingung (*Moderatorhypothese Begabungskonzept*). Bezogen auf die *figurale* Reasoning-Leistung lässt sich ein solcher Moderatoreffekt weder in der Feedback-Bedingung, noch in der feedbackfreien Testbedingung ausmachen.

Wie lässt sich das unterschiedliche Befundmuster für die Testbedingungen Figuren- versus Zahlenfolgen erklären? Dazu sind inhaltliche sowie formale Aspekte zu bedenken. Zum einen muss an die in Abschnitt 5.4 dargestellte geringe Reliabilität der Figurenfolgen erinnert werden. Die deshalb zu erwartenden niedrigen Korrelationen der figuralen Reasoningtest-Leistung mit anderen Variablen dürften das Aufdecken eines Moderatoreffekts generell erschweren (siehe auch Aguinis, 2004 zu Reliabilitätsproblemen bei Moderatoranalysen). Als ein Hinweis für diese Annahme kann der deutlich geringere Anteil an aufgeklärter Varianz im Regressionsmodell zur Vorhersage der figuralen Reasoning-Leistung im Vergleich zum Regressionsmodell zur Vorhersage der numerischen Reasoning-Leistung gewertet werden (vgl. $R^2_{Figuren} = .14$ und Zahlen: $R^2_{Zahlen} = .50$).

Zum anderen lassen sich auch inhaltliche Gründe als potentielle Erklärung für das Ausbleiben des Moderator effekts bei Betrachtung des figuralen Materialbereichs im Gegensatz zum numerischen Materialbereich heranziehen. Möglicherweise besitzen Testanden zu figuralen Reasoning-Aufgaben ein wenig spezifisches Fähigkeits selbstbild (Begabungskonzept), da sie im Allgemeinen solcher Art Aufgaben seltener ausgesetzt sind als Aufgaben, die den Umgang mit Zahlen erfordern. Es ist auch daran zu erinnern, dass die Figurenfolgen für die hier untersuchte Stichprobe von relativ geringer Schwierigkeit waren (siehe Abschnitt 5.4), d.h. auch im Vergleich zu den Zahlenfolgen wenig Misserfolgserleben induziert haben. Zudem ist zu vermuten, dass computergestützte Aufgaben mit Multiple-Choice-Antwortformat (Figurenfolgen) im Gegensatz zu Aufgaben mit freiem Antwortformat (Zahlenfolgen) von vornherein als weniger selbstwertbedrohlich wahrgenommen werden, da bei Misserfolg eine Ratestrategie durch einfaches Anklicken einer Antwortalternative mit weniger psychischem Aufwand verbunden ist als das Eintippen einer geratenden Antwort bei freiem Antwortformat. Insgesamt war das Begabungskonzept der Testanden bei der Bearbeitung der Figurenfolgen in der hier untersuchten Stichprobe möglicherweise von geringer Relevanz, und kann demnach auch keine Moderatorfunktion für den Zusammenhang von Zielorientierung und Leistung innehaben.

6.2.2.2 Anstrengungsvermeidung

Im Folgenden soll der Frage nachgegangen werden, durch welchen *Prozess* die Wechselwirkung von Begabungskonzept und Zielorientierung bei Vorhersage der Leistung erklärt werden kann. Dweck und Leggett (1988) nennen mehrere Mechanismen, die zu einer Leistungsverschlechterung nach Misserfolg führen können. Im Mittelpunkt stehen dabei die investierte *Anstrengung* nach Misserfolg, sowie die Art der *Attribution* des Misserfolgs. So wird angenommen, dass leistungszielorientierte Personen mit geringem Begabungskonzept im Gegensatz zu lernzielorientierten Personen, oder Personen mit hohem Begabungskonzept Misserfolge ungünstig, also leistungshemmend attribuieren, d.h. internale und stabile Ursachen, wie mangelnde Begabung, für Misserfolge annehmen (siehe auch Nicholls et al., 1990 zu subjektiven Theorien über schulische Erfolge). Zum anderen wird angenommen, dass diese Personen nach Misserfolg weniger Anstrengungsbereitschaft zeigen und deshalb relativ geringe Leistungen erzielen. Dieses Verhaltensmuster wird von Dweck auch im Kontext „gelernter Hilflosigkeit“ interpretiert (siehe Abschnitt 2.2.5).

Es soll zunächst geprüft werden, ob auch in der vorliegenden Stichprobe theoriekonform Anstrengungsvermeidung und/oder Erfolgsattribution die Wechselwirkung von Begabungskonzept und Zielorientierung bei Vorhersage der Schulleistung erklären. In diesem Kontext werden die Variablen Anstrengungsvermeidung und Erfolgsattribution als *Mediatorvariablen* konzipiert (siehe S. 114). In einem weiteren Schritt wird untersucht, inwiefern sich die in einer Leistungstestsituation mit Feedback (Zahlenfolgen) gefundene Wechselwirkung von situationsspezifischer Zielorientierung und situationsspezifischem Begabungskonzept ebenfalls mit Hilfe der situationsspezifischen Anstrengungsvermeidung begründen lässt.

6.2.2.2.1 Schulleistungen

Damit die Anstrengungsvermeidung als potentieller *Mediator* aufgefasst werden kann, wird erwartet, dass sich die Variable schulbezogene Anstrengungsvermeidung anhand der Wechselwirkung von schulbezogenem Begabungskonzept und schulbezogener Zielorientierung vorhersagen lässt (Auswertungsschritt: Regression des Mediators auf die unabhängige Variable). Weiterhin sollte die Anstrengungsvermeidung selbst einen wichtigen Prädiktor für die Schulleistung darstellen (Auswertungsschritt: Regression der abhängigen Variable auf den potentiellen Mediator). Entscheidend ist jedoch, dass die Varianzaufklärung der Schulleistung (Kriterium) anhand der Wechselwirkung von Begabungskonzept und Zielorientierung unter Kontrolle der Mediatorvariable Anstrengungsvermeidung deutlich reduziert ist ($H_A = \beta_1 < \beta_2$, mit β_1 = standardisierter Regressionskoeffizient des Prädiktors unter Einbezug des Mediators in die Regressionsanalyse und β_2 = standardisierter Regressionskoeffizient des Prädiktors in der Regressionsanalyse ohne Mediator). In diesem Falle ließe sich die Wechselwirkung von Zielorientierung und Begabungskonzept bei der Vorhersage der Schulleistung auf eine damit einhergehende unterschiedliche Ausprägung an Anstrengungsvermeidung zurückführen (Anstrengungsvermeidung als Mediatorvariable).

Da die Größe des standardisierten Regressionskoeffizienten des Produktterms innerhalb einer multiplen moderierten Regression nicht interpretiert werden sollte (Aguinis, 2004), wird als ein strengeres Kriterium für das Auftreten eines Mediatoreffekts die statistische Signifikanz des Beta-Gewichts des Produktterms in beiden Regressionsmodellen herangezogen (Regressionsmodell ohne Einbezug des potentiellen Mediators als Prädiktor: $H_A: \beta_{\text{Produkt (BxZ)}} \neq 0$; Regressionsmodell mit Einbezug des potentiellen

Mediators als Prädiktor: $H_0: \beta_{\text{Produkt (BxZ)}} = 0$). Tabellen 6.14-6.16 geben einen Überblick über die entsprechenden Regressionsanalysen.

Tabelle 6.14

Kennwerte der Koeffizienten der Regression der schulbezogenen *Anstrengungsvermeidung* auf die schulbezogene Leistungszielorientierung, moderiert durch das schulbezogene Begabungskonzept

Prädiktoren	B	SE	β	t	p
Leistungszielorientierung (Z)	.054	.099	.055	.550	.584
Begabungskonzept (B)	-.196	.091	-.212	-2.140*	.035
Produktterm (BxZ)	-.184	.072	-.244	-2.549*	.012

Anmerkung: In die Analyse gehen Testwerte aus der Bedingung Zahlenfolgen ein ($N = 103$). Die Persönlichkeitsmaße stellen Faktorwerte dar.

Modellkonform gehen unterschiedliche Konstellationen von schulbezogenem Begabungskonzept und schulbezogener Zielorientierung mit einer unterschiedlichen Ausprägung an schulbezogener Anstrengungsvermeidung einher ($\Delta R^2 = .06$, $R^2 = .074$, $F_{3/102} = 3.702$, $p = .014$, $N = 103$; Tabelle 6.14, Abbildung 6.5).

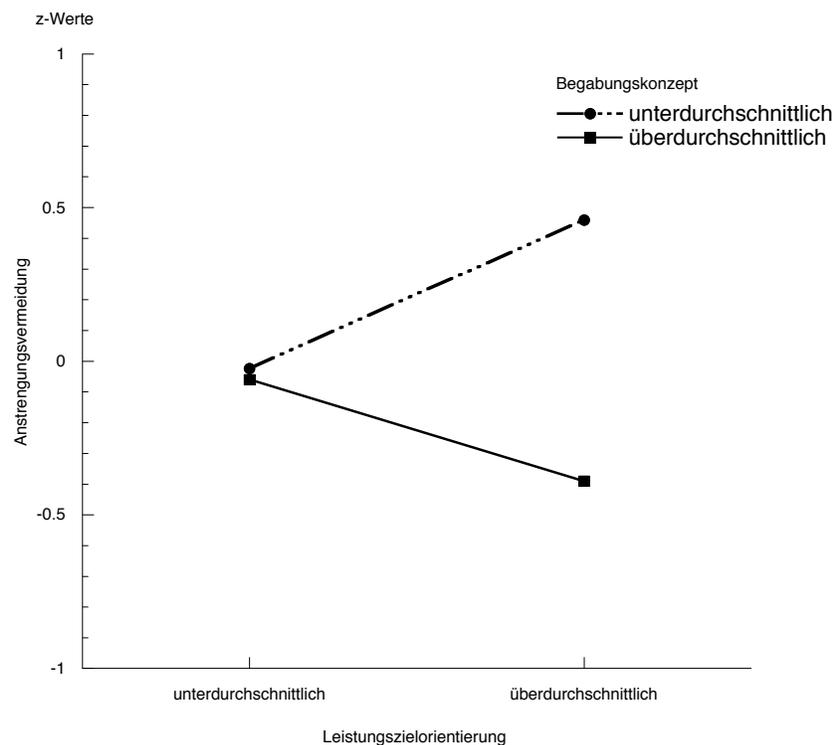


Abbildung 6.5 Anstrengungsvermeidung unter- versus überdurchschnittlich leistungszielorientierter Schüler mit unter- versus überdurchschnittlichem Begabungskonzept.

Zur Vorhersage der Schulleistungen ist jedoch die Hinzunahme einer weiteren Variablen erforderlich, nämlich das im Klassenraum tatsächlich gezeigte Anstrengungs-

verhalten. Die im Schulalltag gezeigte Anstrengung – operationalisiert als Fremdeinschätzung der schulischen Motivation durch den Lehrer – moderiert dabei den Zusammenhang von selbst-berichteter Anstrengungsvermeidung und Schulleistung, unabhängig davon, ob die Leistungsfähigkeit (Mathematikleistung) dabei kontrolliert wird ($\Delta R^2 = .02$, $R^2 = .63$, $F_{4/102} = 44.793$, $p < .001$, $N = 103$; Tabelle 6.15)²⁶.

Tabelle 6.15

Kennwerte der Koeffizienten der Regression der *Schulleistungen* auf die schulbezogene Anstrengungsvermeidung, moderiert durch die im Klassenraum gezeigte Anstrengung

Prädiktoren	B	SE	β	t	p
Mathematikleistung	.466	.043	.703	10.917**	.000
Verhalten im Klassenraum (V)	-1.738	.639	-.173	-2.719*	.008
Anstrengungsvermeidung (A)	-3.803	1.758	-.387	-2.162*	.033
Produktterm (VxA)	1.480	.636	.414	2.326*	.022

Anmerkung: In die Analyse gehen Testwerte aus der Bedingung Zahlenfolgen ein ($N = 103$). Die Persönlichkeitsmaße stellen Faktorwerte dar.

Unter Kontrolle der Wechselwirkung beider Variablen (berichtete Anstrengungsvermeidung und gezeigtes Anstrengungsverhalten) lässt sich die Schulleistung nicht mehr signifikant anhand der Wechselwirkung von Begabungskonzept und Zielorientierung vorhersagen ($R^2 = .64$, $F_{7/101} = 26.08$, $p < .001$, $N = 102$; Tabelle 6.16). Unter den gegebenen Stichprobenverhältnissen kann mit hinreichender Sicherheit ($1 - \beta \geq .80$) allerdings nur ein mittlerer und starker Effekt (genauer $f^2 \geq .08$) der Wechselwirkung von Begabungskonzept und Leistungszielorientierung auf die Schulleistung bei Kontrolle der Anstrengungsregulation ausgeschlossen werden.

²⁶ Bei Betrachtung der Korrelationskoeffizienten zwischen den Prädiktoren sowie dem Kriterium fällt nur für die Variable Anstrengungsvermeidung auf, dass die Korrelation zu einem Prädiktor, nämlich dem Moderator Verhalten im Klassenraum, etwas größer ausfällt ($r = .17$, $p < .05$) als zum Kriterium ($r = -.14$, $p = .08$). Die Korrelationskoeffizienten unterscheiden sich jedoch nicht signifikant. Es kann somit angenommen werden, dass die Variable Anstrengungsvermeidung keine Supressorvariable im multiplen Regressionsmodell darstellt (siehe Bartussek, 1970). Die für den Produktterm im Regressionsmodell (Tabelle 6.15) angegebene Kollinearitäts-Toleranz beträgt .114. Multikollinearität muss deshalb ebenfalls nicht angenommen werden (siehe aber auch zum Problem der Multikollinearität bei einer multiplen moderierten Regression Fußnote Nummer 22).

Tabelle 6.16

Kennwerte der Koeffizienten der Regression der *Schulleistungen* auf die Mathematikleistung, die Leistungszielorientierung moderiert durch das Begabungskonzept, sowie die Anstrengungsvermeidung, moderiert durch die im Klassenraum gezeigte Anstrengung

<i>Prädiktoren</i>	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>
Mathematikleistung	.466	.045	.703	10.324**	.000
Leistungszielorientierung (Z)	-.693	.691	-.064	-1.003	.319
Begabungskonzept (B)	.296	.652	.029	.453	.651
Produktterm (BxZ)	-.931	.547	-.108	-1.701	.092
Verhalten im Klassenraum (V)	-1.575	.654	-.157	-2.408*	.018
Anstrengungsvermeidung (A)	-4.062	1.793	-.410	-2.265*	.026
Produktterm (VxA)	1.461	.646	.404	2.262*	.026

Anmerkung: In die Analyse gehen Testwerte aus der Bedingung Zahlenfolgen ein ($N = 103$). Die Persönlichkeitsmaße stellen Faktorwerte dar.

In Abbildung 6.6 wird noch einmal deutlich, dass eine geringe Anstrengungsvermeidung für die Leistungsvorhersage nicht ausreichend ist. Entscheidend ist, ob die selbst berichtete *Anstrengungsbereitschaft* auch in beobachtbares *Anstrengungsverhalten* im Klassenraum umgesetzt wird. Erst das gemeinsame Auftreten von Anstrengungsbereitschaft und Anstrengung im Schulalltag zahlt sich – bei gleicher Leistungsfähigkeit – in höheren Schulleistungen aus (beachte die Kontrolle der Leistungsfähigkeit in allen Regressionsanalysen zur Vorhersage der Schulleistungen).

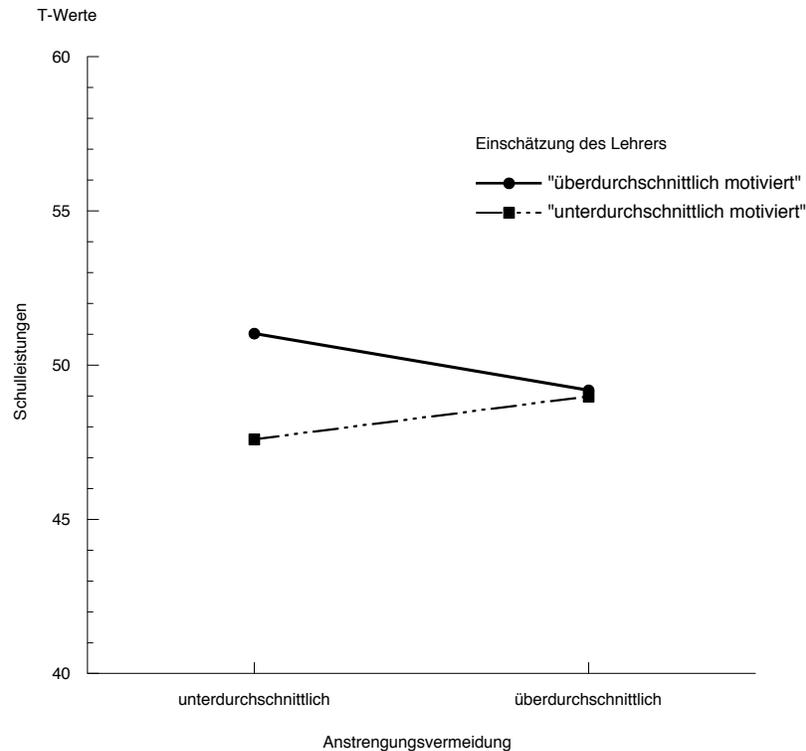


Abbildung 6.6 Schulleistungen unter- versus überdurchschnittlich anstrengungsvermeidender Schüler (Selbsteinschätzung) mit unter- versus überdurchschnittlichem Anstrengungsverhalten im Klassenraum (Fremdeinschätzung)

Auch eine niedrig ausgeprägte Anstrengungsvermeidung kann also mit relativ geringer Performanz einhergehen (Abbildung 6.6). Dieser etwas überraschende Befund könnte als Ausdruck einer wenig realistischen Selbstwahrnehmung der Schüler verstanden werden, die sich als anstrengungsbereit beschrieben haben, durch den Lehrer jedoch als unterdurchschnittlich motiviert beurteilt wurden, und auch relativ niedrigere Schulleistungen zeigten. Da Anstrengungsbereitschaft sozial erwünscht ist, lässt sich eine Divergenz von Selbst- und Fremdeinschätzung auch mit einem sozial erwünschten Antwortverhalten der Testanden erklären. Die beobachteten Korrelationen zwischen Anstrengungsvermeidung und Leistung sind folglich eher gering.

Zusammenfassend wird festgehalten: Die Anstrengungsregulation – Wechselwirkung von berichteter Anstrengungsvermeidung und im Schulalltag gezeigter Anstrengung – stellt einen Mediator, also eine Ursache dar, für die Wechselwirkung von Begabungskonzept und Zielorientierung bei Vorhersage der Leistung. Dabei scheint die berichtete Anstrengungsvermeidung nur ein Aspekt zu sein. Entscheidend ist, inwiefern diese auch tatsächlich im schulischen Alltag in für den Lehrer beobachtbares Verhalten

umgesetzt wird. Die Anstrengungsregulation kann somit als ein Prozess ausgemacht werden, der zur Erklärung von schulischen Leistungsdefiziten leistungszielorientierter Personen mit geringem Begabungskonzept beiträgt (Abbildung 6.5).

Abbildung 6.2 (Seite 124) weist allerdings darauf hin, dass in der vorliegenden Stichprobe die vermeintliche „Risikogruppe“, nämlich leistungszielorientierte Schüler mit geringem Begabungskonzept, keine Leistungseinbußen zeigte, verglichen z.B. mit leistungszielorientierten Schülern mit hohem Begabungskonzept (siehe auch Abbildung 6.6). Erwartungskonform wiesen Schüler aber – auch unter Kontrolle ihrer Leistungsfähigkeit – umso höhere Schulleistungen auf, je seltener sie Leistungsziele verfolgten und höher ihr schulbezogenes Begabungskonzept ausfiel.

In Abbildung 6.7 werden die bisherigen Befunde in einem Modell zusammengefasst. Es werden die Moderatoreffekte (Begabungskonzept, schulische Motivation), sowie der Mediatoreffekt (Anstrengungsvermeidung) bei der Vorhersage der Schulleistungen dargestellt.

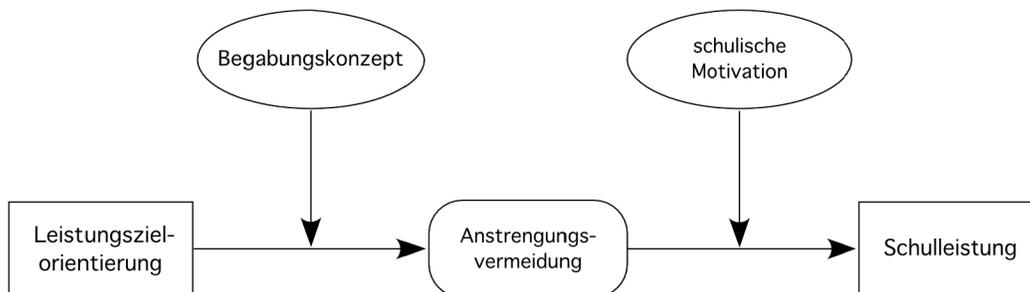


Abbildung 6.7 Modell der Moderatorvariablen (Ellipsen) sowie der Mediatorvariablen (gerundetes Rechteck) bei der Vorhersage der Variable Schulleistung.

Bisher wurde nur die Stichprobe der Testanden betrachtet, die die *Zahlenfolgen* bearbeitet haben, da sich nur in dieser Stichprobe auf situationsspezifischer Ebene ein Moderatoreffekt der Variablen Begabungskonzept für den Zusammenhang von Leistungszielorientierung und Schulleistung nachweisen lässt. Folglich wird die potentielle Mediatorfunktion der Variablen Anstrengungsvermeidung auf situationsspezifischer Ebene ebenfalls nur für diese Stichprobe untersucht. Es soll an dieser Stelle deshalb nur sehr kurz auf die Ergebnisse für die *Gesamtstichprobe*, die entweder Zahlen- oder Figu-

renfolgen bearbeitet haben, zur Überprüfung der Generalisierbarkeit der bisherigen Befunde, eingegangen werden.

Auch in der Gesamtstichprobe stellt das Begabungskonzept erwartungsgemäß eine Moderatorvariable für den Zusammenhang von Zielorientierung und Anstrengungsvermeidung dar ($R^2 = .04$, $F_{3/242} = 3.115$, $p < .05$, $\Delta R^2 = .02$, $F_{1/239} = 4.035$, $p < .05$, $N = 243$). Unter Kontrolle der Anstrengungsregulation lässt sich die Schulleistung nicht mehr signifikant anhand der Wechselwirkung von Begabungskonzept und Leistungszielorientierung vorhersagen ($R^2 = .68$, $F_{7/239} = 68.736$, $p < .001$, $\Delta R^2 = .004$, $F_{1/232} = 2.882$, $p = .091$, $N = 240$). Die Anstrengungsregulation stellt somit auch in der Gesamtstichprobe der Testanden, die entweder Zahlenfolgen oder Figurenfolgen bearbeitet haben, eine Mediatorvariable für den Zusammenhang von Zielorientierung sowie Begabungskonzept auf der einen Seite und Schulleistung auf der anderen Seite dar.

6.2.2.2 Reasoning-Testleistung

Zur Prüfung der Frage, ob die Variable Anstrengungsvermeidung im Modell von Zielorientierung, Begabungskonzept und *numerischer* Reasoning-Leistung eine Mediatorfunktion innehat, wird wieder die unter Feedback-Bedingungen zum zweiten Messzeitpunkt erbrachte Leistung der *Versuchsgruppe 1* (F- F+) betrachtet. Das unmittelbar vor der Testung erfasste situationsspezifische Begabungskonzept steht hier in keinem Zusammenhang zur situationsspezifischen Anstrengungsvermeidung ($\beta = .081$, $t = .565$, $p = .574$, $N = 50$), moderiert aber den Zusammenhang von Zielorientierung und Anstrengungsvermeidung²⁷ ($R^2 = .18$, $F_{3/49} = 4.626$, $p = .007$, $\Delta R^2 = .13$, $F_{1/46} = 7.525$, $p = .009$, $N = 50$; Regression des potentiellen Mediators auf die unabhängige Variable). Erwartungsgemäß zeigten leistungszielorientierte Testanden mit geringem Begabungskonzept die höchste Ausprägung an Anstrengungsvermeidung (7-stufige Skala; siehe Abbildung 6.8). Dennoch profitierte gerade diese Testandengruppe hypothesenkonträr von Richtig/Falsch-Feedback (siehe Abbildungen 6.3 und 6.4).

²⁷ Die Multikollinearität der Prädiktoren kann hier nicht völlig ausgeschlossen werden (die Toleranz beträgt .05, ist aber $> .01$). Bei Betrachtung der Korrelationsmatrix der Prädiktoren sowie des Kriteriums fällt auf, dass die Variable Begabungskonzept stärker mit dem Prädiktor Zielorientierung korreliert ist ($r = .32$, $p < .05$) als mit dem Kriterium ($r = .08$, $p = .287$). Die Differenz ist jedoch nicht signifikant ($t_{47} = -1.55$, $p = .128$, $N = 50$). Dennoch sollten die Parameter des Regressionsmodells mit Vorsicht interpretiert werden, da eine Suppressorwirkung der Variablen Begabungskonzept nicht ausgeschlossen werden kann (siehe auch Problem der Multikollinearität bei einer multiplen moderierten Regression).

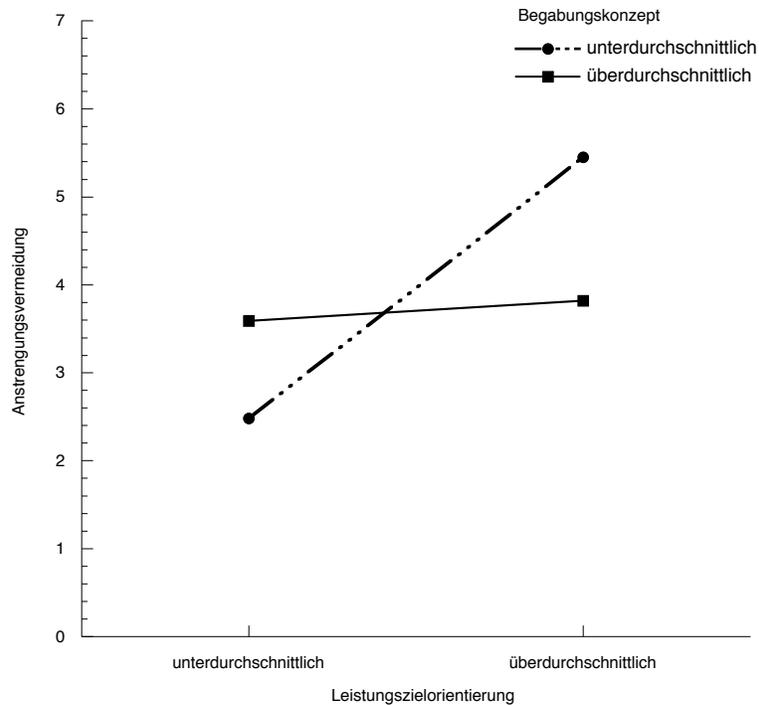


Abbildung 6.8 Situationsspezifische Anstrengungsvermeidung unter- versus überdurchschnittlich situationsspezifisch leistungszielorientierter Testanden mit unter- versus überdurchschnittlichem situationsspezifischem Begabungskonzept in der Feedback-Bedingung.

Die vor der Testung berichtete Anstrengungsvermeidung trägt allerdings nicht zur Vorhersage der Reasoning-Testleistung in der Feedback-Bedingung bei und kann auch deshalb nicht als Mediator im Modell von Zielorientierung, Begabungskonzept und Leistung fungieren ($\beta = -.039$, $t = -.268$, $p = .790$, $N = 50$, Regression der abhängigen Variablen auf die potentielle Mediatorvariable). Unter den gegebenen Stichprobenverhältnissen ist ein mittlerer und starker Effekt ($f^2 \geq .17$) der Anstrengungsvermeidung auf die Reasoning-Leistung mit hinreichender statistischer Sicherheit ($1 - \beta \geq .80$) auszuschließen. Das Ausmaß an Anstrengungsvermeidung vor Testbeginn erklärt demnach nicht die Leistungsunterschiede zwischen Personen mit hoher oder niedriger Ausprägung in den Merkmalen Leistungszielorientierung und Begabungskonzept.

Der geringe Zusammenhang zwischen Anstrengungsvermeidung und Leistung fiel auch bei der Analyse der Schulleistungen auf (Abschnitt 6.2.2.2.1). Eine weitere Variable, die als Indikator für Anstrengung den Zusammenhang zwischen selbst berichteter situationsspezifischer Anstrengungsvermeidung und Reasoning-Leistung aufklärt (siehe Analyse der Schulleistungen), liegt für das situationsspezifische Modell leider nicht vor.

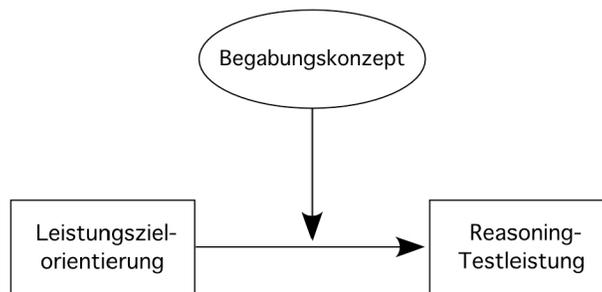


Abbildung 6.9 Begabungskonzept als Moderator des Zusammenhangs von Leistungszielorientierung und numerischer Reasoning-Testleistung in einer Testbedingung mit Feedback.

Abbildung 6.9 zeigt den Moderatoreffekt der Variablen Begabungskonzept für den Zusammenhang der Variablen Leistungszielorientierung und numerischer Reasoning-Testleistung. Anstrengungsvermeidung kann als Mediatorvariable ausgeschlossen werden.

6.2.2.3 Attribution von Erfolg/Misserfolg

Nach Dweck (2000) sowie Nicholls et al. (1990) werden neben der Anstrengungsvermeidung ungeeignete Ursachenzuschreibungen für Erfolg/Misserfolg für geringe Leistungen nach rückgemeldetem Misserfolg verantwortlich gemacht. Insbesondere leistungszielorientierte Personen mit geringem Begabungskonzept sollten Misserfolg ungünstig attribuieren (internal und stabil) und deshalb nach Misserfolg Leistungsdefizite zeigen.

In der vorliegenden Untersuchung konnten jedoch keine Hinweise gefunden werden, die diese theoretischen Annahmen stützen. Tabelle 6.17 stellt die Regressionsanalysen zur Überprüfung eines potentiellen Mediatoreffekts der Variablen Attribution auf Fähigkeit sowie Attribution auf Anstrengung für die Zusammenhangsstruktur der Variablen Begabungskonzept, Zielorientierung und Leistung dar. Die Ergebnisse zeigen konsequent, dass der Attribution von Erfolg/Misserfolg keine Mediatorfunktion zugeschrieben werden kann. Wie die Regressionskoeffizienten in Tabelle 6.17 belegen, weist weder die Variable Attribution auf Anstrengung noch die Variable Attribution auf Fähigkeit – als Facetten der Attribution von Erfolg/Misserfolg – die zur Analyse von Mediatoreffekten notwendige Zusammenhangsstruktur auf. Tabelle 6.17 stellt die Ergebnisse für die Zahlenfolgen-Stichprobe dar. Die für die Gesamtstichprobe (Figuren- oder Zahlenfolgen) berechneten Ergebnisse sind analog.

Tabelle 6.17

Koeffizienten der Regressionsanalysen zur Bestimmung eines Mediatoreffekts der Variablen der Attribution von schulischen Erfolgen/Misserfolgen in der Zahlenfolgen-Stichprobe

Modell	unabhängige Variablen	abhängige Variable	R ²	F	df	p	N
Regression des potentiellen Mediators auf den Prädiktor	Begabungskonzept (B)	Attribution auf Fähigkeit (Mediator)	.16	7.538**	3/102	.001	103
	Leistungszielorientierung (Z)						
	Produktterm (BxZ)						
Regression des Kriteriums auf den potentiellen Mediator	Begabungskonzept (B)	Attribution auf Anstrengung (Mediator)	.01	1.286	3/102	.283	103
	Leistungszielorientierung (Z)						
	Produktterm (BxZ)						
Regression des Kriteriums auf den potentiellen Mediator	Attribution auf Fähigkeit	Schulleistung (Kriterium)	.02	3.086	1/103	.086	104
	Attribution auf Anstrengung	Schulleistung (Kriterium)	.01	.239	1/103	.626	104

Anmerkung: Die Persönlichkeitsmaße gingen als Faktorwerte in die Analysen ein.

6.2.2.4 Testangst

Testangst gilt als ein bedeutsamer Prädiktor für Testleistungen. Meijer und Elshout (2001) stellten dabei insbesondere die medierende Rolle des *Begabungskonzepts* heraus. Danach geht Testangst mit einem geringen Begabungskonzept einher. Ein hohes Begabungskonzept sollte dagegen vor Testangst und in Folge vor Leistungsdefiziten schützen. Wird ein geringes Begabungskonzept als ursächlich für Testangst aufgefasst, lässt sich die Variable Testangst auch als *Mediatorvariable* für den Zusammenhang von Begabungskonzept und Testleistung konzipieren. Nach Morris et al. (1981) bedingen vor allem Worry-Kognitionen – die kognitive Komponente von Testangst – Leistungseinbußen. Worry-Kognitionen binden kognitive Ressourcen, die dann für die zu bearbeitende Aufgabe nicht mehr zur Verfügung stehen (siehe Abschnitt 2.2.6).

Eine Testsituation mit unmittelbarer Ergebnisrückmeldung ist geprägt durch Erfolgs- und Misserfolgserleben und kann infolgedessen testangstauslösend wirken. Es wird deshalb erwartet, dass die Variable *Testangst* insbesondere in der *Feedback*-Bedingung einen wichtigen Prädiktor für Testleistungen darstellt. Weiterhin wird angenommen, dass die Ausbildung von Testangst im Testverlauf neben der habituellen Angstneigung durch das aufgabenspezifische *Begabungskonzept* (spezifisches Fähigkeits-Selbstkonzept) bestimmt ist. Da durch Leistungsrückmeldungen im Testverlauf eine Anpassung des Begabungskonzepts an das tatsächliche Leistungsniveau möglich wird bzw. extern angebotene Information über das eigene Fähigkeitsniveau genutzt werden kann, sollte das vor der Testung berichtete Begabungskonzept in einer Feedback-Bedin-

gung verglichen mit einer feedbackfreien Bedingung an Vorhersagekraft verlieren. Der dennoch anzunehmende kausale Zusammenhang von – vor Testbeginn erfasstem – Begabungskonzept und Testleistung sollte in der *Feedback*-Bedingung im Gegensatz zur feedbackfreien Testbedingung insbesondere durch die hier stärkere Wirkung der Variablen Testangst bestimmt werden (*Mediatorhypothese Testangst*). Leistungsdefizite von Personen mit geringem Begabungskonzept ließen sich dann auf die hohe Ausprägung an Testangst in dieser Personengruppe zurückführen.

Zunächst soll der Zusammenhang von Begabungskonzept und Testangst wieder anhand der Vorhersage der *Schulleistungen* analysiert werden. In einem zweiten Schritt werden dann beide Variablen als Prädiktoren bei der Vorhersage der Reasoning-Testleistungen näher betrachtet und geprüft, ob das Modell auf Testsituationen mit Feedback-Angebot übertragen werden kann. Zur Feststellung eines *Mediatoreffekts* der Variablen Testangst wird erwartet, dass der Beitrag der Variablen Begabungskonzept (Prädiktor) zur Vorhersage der Leistung (Kriterium) unter Kontrolle der Variablen Testangst (potentieller Mediator) deutlich reduziert ist ($H_A: \beta_1 < \beta_2$, mit β_1 = standardisierter Regressionskoeffizient des Prädiktors unter Einbezug des potentiellen Mediators Testangst in die Regressionsanalyse und β_2 = standardisierter Regressionskoeffizient des Prädiktors ohne Einbezug des potentiellen Mediators Testangst in die Regressionsanalyse).

6.2.2.4.1 Schulleistungen

Die schulbezogene Test- oder Prüfungsangst²⁸ ($R^2 = .09$, $F_{1/103} = 11.00$, $\beta = -.312$, $p = .001$) erweist sich als ein bedeutsamer Prädiktor für die Schulleistungen. Immerhin sind 9% der Leistungsvarianz (Schulleistung) allein anhand der berichteten Testangst vorher-sagbar. Es lässt sich weiterhin theoriekonform zeigen, dass das schulbezogene Begabungskonzept und die schulbezogene Testangst negativ korreliert sind ($R^2 = .03$, $F_{1/103} = 3.983$, $\beta = -.194$, $p = .049$; Regression des potentiellen Mediators auf den Prädiktor). Der Beitrag des Begabungskonzepts zur Vorhersage der Schulleistungen ($R^2 = .05$, $F_{1/104} = 6.652$, $\beta = .246$, $p = .011$) ist zwar kleiner, wenn die Variable Testangst einbezogen, also kontrolliert wird ($R^2 = .11$, $F_{2/103} = 7.64$, $p = .001$, $\beta = .188$, $t = 1.99$, $p = .049$; Regression der abhängigen Variablen auf die unabhängige Variable und den potentiellen

²⁸ Als Testwerte wurden hier für alle Analysen die Summenwerte der Subskala Testangst (Meijer, 1996, 2001) sowie zur Erfassung des Begabungskonzepts der Mittelwert der beiden Items nach Dweck (2000) herangezogen.

Mediator). Die Reduktion des Vorhersagebeitrags des Begabungskonzepts ist jedoch statistisch nicht signifikant ($B_1 = 1.852$; $B_2 = 1.43$; KI für $\alpha = 5\%$: -1.58 bis 2.42, Cohen, et al., 2003). Die Nullhypothese muss beibehalten werden ($H_0: \beta_1 = \beta_2$). Zudem nimmt die insgesamt aufgeklärte Varianz zu (vgl. $R^2 = .05$ und $R^2 = .11$), so dass geschlussfolgert werden muss, dass die Variable Testangst eher einen zur Variable Begabungskonzept *zusätzlichen* Prädiktor bei Vorhersage der Schulleistungen darstellt. Der Zusammenhang von schulbezogenem Begabungskonzept und Schulleistung kann demnach nicht eindeutig auf die in Abhängigkeit des Begabungskonzepts unterschiedlich ausgeprägte schulbezogene Testangst zurückgeführt werden (*Mediatorhypothese Testangst*).

Dieses Befundmuster ergibt sich auch bei Betrachtung der *Gesamtstichprobe* der Testanden ($N = 244$), die entweder Figurenfolgen oder Zahlenfolgen bearbeitet haben. Auch hier ist der Beitrag des Begabungskonzepts bei der Vorhersage der Schulleistungen ($R^2 = .06$, $F_{1/244} = 15.973$, $\beta = .248$, $p < .001$) unter Kontrolle der Variablen Testangst zwar reduziert ($R^2 = .08$, $F_{2/243} = 11.887$, $p < .001$, $\beta = .203$, $t = 3.21$, $p = .002$). Die Reduktion des Vorhersagebeitrags des Begabungskonzepts ist jedoch ebenfalls statistisch nicht signifikant ($B_1 = 1.89$; $B_2 = 1.557$; KI: -1.0 bis 1.66; Cohen, et al., 2003). Testangst kann daher nicht eindeutig als Ursache für den beobachteten Zusammenhang von Begabungskonzept und Schulleistung angesehen werden.

6.2.2.4.2 Intelligenzstatustest-Leistungen

Mit Hilfe der Variablen schulbezogene Testangst lässt sich in der Zahlenfolgen-Stichprobe die Leistung in einem Intelligenzstatustest (CAT, MidYIS) nicht präzisieren ($R^2 = .02$, $F_{1/100} = 3.03$, $\beta = -.172$, $p = .09$), wohl aber mit Hilfe des schulbezogenen Begabungskonzepts ($R^2 = .06$, $F_{1/101} = 7.349$, $\beta = .262$, $p = .008$). Da die Intelligenzstatustest-Leistung durch den potentiellen Mediator schulbezogene Testangst nicht vorhersagbar ist, kann die Testangst keine Mediatorvariable für den Zusammenhang von schulbezogenem Begabungskonzept (Prädiktor) und intellektueller Leistung (Kriterium) darstellen. Prüfungs- oder Testangst sollte somit in der hier untersuchten Stichprobe nicht als Ursache für die beobachteten Leistungsdefizite in einem Intelligenzstatustest von Personen mit geringem schulbezogenem Begabungskonzept interpretiert werden.

Bei Betrachtung der *Gesamtstichprobe* ($N = 244$) ist die Befundlage weniger eindeutig. Die Testangst stellt hier einen (schwachen) Prädiktor der Intelligenzstatustest-Leistungen dar. 3% der Leistungsvarianz lassen sich anhand der Testangst signifikant

vorhersagen ($F_{1/238} = 9.275$, $\beta = -.194$, $p < .05$). Die Kontrolle der Variablen Testangst bei Vorhersage der Leistungen im Intelligenzstatustest führt zu einer (leichten) Reduktion des Vorhersagebeitrags des schulbezogenen Begabungskonzepts ($R^2 = .08$, $F_{2/238} = 10.901$, $p < .001$, $\beta = .224$, $t = 3.477$, $p = .001$, vgl. ohne Kontrolle der Variablen Testangst: $\beta = .258$, $t = 4.112$, $p < .001$), die aber ebenfalls statistisch nicht signifikant ist ($B_1 = 1.936$; $B_2 = 1.69$; KI: -1.1 bis 1.58 ; Cohen, et al., 2003; Regression der Intelligenzstatustest-Leistung auf das Begabungskonzept und den potentiellen Mediator Testangst). Eine Mediatorwirkung der Variablen Testangst kann somit nicht angenommen werden.

Zusammenfassend ist festzuhalten: In der vorliegenden Untersuchung ließen sich *Schulleistungen* erwartungsgemäß anhand der Test- oder Prüfungsangst vorhersagen. Die Testangst fungiert jedoch nicht eindeutig als Mediatorvariable für den Zusammenhang von schulbezogenem Begabungskonzept und Schulleistung und kann somit nicht als eine Ursache für Leistungsdefizite wenig selbstkonzeptstarker Schüler bewertet werden. Hinsichtlich der feedbackfreien *Intelligenzstatustest*-Leistungen stellt die Test- oder Prüfungsangst einen nur schwachen Prädiktor der Leistung dar. Eine völlige Unabhängigkeit der beiden Variablen Testangst und Intelligenzstatustest-Leistung ist jedoch auch aufgrund der zahlreichen in der Literatur referierten Befunde zum Konstrukt Testangst nicht zu erwarten (siehe z.B. Meijer, 1996, 2001; Morris et al., 1981; Wiedl, Bethge & Bethge, 1982).

Test- oder *Prüfungsangst* wirkt sich vor allem negativ auf (feedbackbedingte) Schulleistungen aus, ein hohes *Begabungskonzept* ist jedoch sowohl für Schulleistungen als auch feedbackfreie Intelligenzstatustest-Leistungen leistungsförderlich. Tabelle 6.18 stellt die standardisierten Regressionskoeffizienten (β) aus den einzelnen oben genannten Regressionsanalysen zur Vorhersage der Schulleistungen bzw. der Intelligenzstatustest-Leistung in der Zahlenfolgen-Stichprobe nochmals gegenüber.

Tabelle 6.18

Zusammenhang von schulbezogenem Begabungskonzept und Testangst mit Schulleistungen und Intelligenzstatustest-Leistungen

	<i>Schulleistung</i> (N = 105)	<i>Intelligenzstatustest-Leistung</i> (N = 103)
Begabungskonzept	.25*	.26*
Testangst	-.31**	-.17

Anmerkung: Als Zusammenhangsmaß wurden die standardisierten Regressionskoeffizienten (β) aus den Regressionsanalysen (je ein Prädiktor) herangezogen (siehe Text). * = $p < .05$, ** = $p < .01$

An dieser Stelle soll kurz darauf hingewiesen werden, dass ein empirisch nachgewiesener Zusammenhang zwischen Begabungskonzept und Leistung selbstverständlich nicht kausal interpretiert werden kann. Theoretisch wird jedoch ein geringes Begabungskonzept – latente Eigenschaft – als antezedent für Leistungsdefizite – Verhalten – unter Misserfolgsbedingungen angesehen (siehe Theorie zum Begabungskonzept nach Meyer, 1973, 1984; im Rahmen von Zieltheorien z.B. nach Dweck & Leggett, 1988). Die Annahme, dass sich in einem geringen Begabungskonzept die Misserfolgserfahrung in der Lerngeschichte des Testanden widerspiegelt, ist dessen ungeachtet gleichermaßen gerechtfertigt.

6.2.2.4.3 Numerische Reasoning-Testleistung

6.2.2.4.3.1 Situationsspezifische Selbsteinschätzung

Für die Prüfung einer potentiellen Mediatorfunktion der situationsspezifischen Testangst wird wieder *Versuchsgruppe 1* (F– F+) betrachtet. Die vor Testbeginn erfasste Testangst²⁹ stellt einen wesentlichen Prädiktor für die Testleistung in der *Feedback*-Bedingung ($\beta = -.350$, $t = -2.616$, $p = .012$) dar, gleiches gilt für das vor der Testung erfasste Begabungskonzept³⁰ ($\beta = .334$, $t = 2.483$, $p = .016$), selbst wenn die intellektuelle Leistungsfähigkeit durch Einbezug der Variable Intelligenzstatustest-Leistung als weiterer Prädiktor kontrolliert wird. Immerhin 11% der Leistungsvarianz lassen sich in der Feedback-Bedingung allein anhand der vor Testbeginn berichteten Testangst vorhersagen ($F_{1/50} = 6.841$, $p = .012$), 9% anhand des vor Testbeginn angegebenen Begabungskonzepts ($F_{1/50} = 6.167$, $p = .016$). Weiterhin zeigt sich auch situationsspezifisch, dass Testangst und Begabungskonzept erwartungsgemäß negativ korreliert sind ($\beta = -.600$, $t = -5.25$, $p < .001$). Unter Kontrolle der Variablen Testangst ist die Vorhersage der Testleistung anhand des Prädiktors Begabungskonzept indessen nicht mehr möglich ($R^2 = .11$, $F_{2/50} = 4.125$, $p = .022$, $\beta = .194$, $t = 1.166$, $p = .250$, $N = 51$; Regression der abhängigen Variablen auf die unabhängige Variable und den potentiellen Mediator). Die Nullhypo-

²⁹ *Testangst*: Mittelwert der Antworten zu Item „I am afraid I may not do as well on this test as I could.“ versus „I am pretty optimistic that I will do as well on this test as I can.“ zur Erfassung der Komponente Worry und Item „Just now I am nervous about doing these problems.“ versus „Just now I am relaxed about doing these problems.“ zur Erfassung der Komponente Emotionality.

³⁰ *Begabungskonzept*: Mittelwert der Antworten zu Item „I feel pretty confident that I shall be able to solve most of the problems.“ versus „I do not feel confident that I shall be able to solve most of the problems.“ zur Erfassung von Konfidenz und Item „The more problems I tackle, the easier it will be to solve them.“ versus „I do not think, that tackling more problems will help me to solve them.“ zur Erfassung von Lernfähigkeit.

these kann abgelehnt werden ($H_0: \beta_1 = \beta_2$). Unter den gegebenen Stichprobenverhältnissen lässt sich ein mittlerer und starker Effekt (genauer ein Effekt von $f^2 \geq .17$) der Variablen Begabungskonzept auf die Testleistung bei Kontrolle der Variablen Testangst mit hinreichender statistischer Sicherheit ($1 - \beta \geq .80$) ausschließen. Die situationspezifische Testangst kann somit als Mediator des Zusammenhangs von situationspezifischem Begabungskonzept und Testleistung, und zwar nur in der *Feedback*-Bedingung, interpretiert werden (*Mediatorhypothese Testangst*).

In der *feedbackfreien* Testbedingung der gleichen Stichprobe (Versuchsgruppe 1, F– F+, Test 1) bleibt unter Kontrolle der Variablen Testangst der Zusammenhang von situationspezifischem Begabungskonzept und Testleistung dagegen bestehen ($R^2 = .35$, $F_{2/50} = 14.685$, $p < .001$, $\beta = .611$, $t = 4.793$, $p < .001$, $N = 51$). Handelt es sich bei der beobachteten Mediatorwirkung der Variablen Testangst in der *Feedback*-Bedingung (Versuchsgruppe 1, F– F+, Test 2) um einen einfachen Effekt der *Testwiederholung*, sollte sich zum zweiten Messzeitpunkt in der *feedbackfreien* Bedingung (Versuchsgruppe 2, F+ F– und Kontrollgruppe, F– F–) ebenfalls ein entsprechender Mediatoreffekt der Variable Testangst finden lassen. Bei Kontrolle der Variablen Testangst stellt das Begabungskonzept hier aber weiterhin einen signifikanten Prädiktor der *feedbackfreien* Testleistung dar ($R^2 = .14$, $F_{2/50} = 5.308$, $p = .008$, $\beta = .342$, $t = 2.691$, $p = .01$, $N = 55$).

Abbildung 6.10 zeigt ein Modell der Mediatorfunktion der Variable Testangst bei Vorhersage der Reasoning-Leistung anhand des vor Testbeginn erfassten Begabungskonzepts unter *Feedback*-Bedingungen.

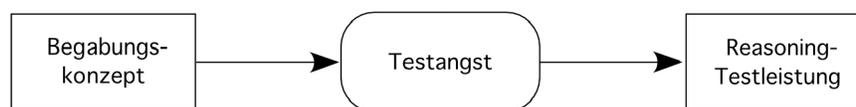


Abbildung 6.10 Testangst als Mediator des Zusammenhangs von Begabungskonzept und numerischer Reasoning-Testleistung unter *Feedback*-Bedingungen.

Im Folgenden soll die Zusammenhgangsstruktur von Testangst, Begabungskonzept und Reasoning-Leistung noch etwas genauer analysiert werden. Mit dem Begriff *intraindividuelle Perspektive* wird dabei auf den Vergleich performanzbezogener differentieller Effekte in der *Feedback*-Bedingung und der *feedbackfreien* Bedingung in einer Substichprobe, nämlich *Versuchsgruppe 1* (F– F+), Bezug genommen. Mit dem Begriff

interindividuelle Perspektive soll sich auf die Kontrastierung der differentiellen Effekte in den Feedback- versus Non-Feedback-Bedingungen zu beiden Messzeitpunkten in der Gesamtstichprobe bezogen werden.

Situationsspezifisches Begabungskonzept: Die Annahme, dass der Zusammenhang von Begabungskonzept und Testleistung in der Feedback-Bedingung geringer ist als in der feedbackfreien Bedingung, scheint zwar deskriptiv bei Betrachtung der Beta-Gewichte in Versuchsgruppe 1 (F– F+) zum ersten und zweiten Messzeitpunkt nahe liegend (vgl. $\beta = .616$ und $\beta = .334$, Tabelle 6.19, *intraindividuelle Perspektive*); allerdings findet sich auch in der *feedbackfreien* Testbedingung in Versuchsgruppe 2 (F+ F–) und Kontrollgruppe (F– F–) zum zweiten Testzeitpunkt ein vergleichbar niedriges Beta-Gewicht ($\beta = .363$, Tabelle 6.20). Die zwar auch statistisch signifikante Reduktion der Beta-Gewichte von Test 1 (F–) zu Test 2 (F+) in Versuchsgruppe 1 (siehe Tabelle 6.19, $B_1 = 1.259$; $B_2 = .602$; KI: 0 bis 1.32, Cohen et al., 2003), d.h. der beobachtete geringere Zusammenhang von Begabungskonzept und Testleistung in der Feedback-Bedingung (Test 2, *intraindividuelle Perspektive*), lässt sich demnach nicht eindeutig auf die Feedback-Intervention zurückführen, sondern auch auf die Wiederholungsmessung (*interindividuelle Perspektive*).

Tabelle 6.19

Regression der numerischen Reasoning-Testleistung der *Versuchsgruppe 1* (F– F+, $N = 51$) auf die situationsspezifische Testangst sowie auf das situationsspezifische Begabungskonzept in einer feedbackfreien Testbedingung sowie in einer Testbedingung mit Feedback – *intraindividuelle* Perspektive

		R^2	F	df	β	t	p
Testangst	kein Feedback (Test 1)	.06	4.416*	1/50	-.288	-2.101*	.041
	Feedback (Test 2)	.11	6.841*	1/50	-.35	-2.616*	.012
Begabungskonzept	kein Feedback (Test 1)	.37	29.972**	1/50	.616	5.475**	.000
	Feedback (Test 2)	.09	6.167*	1/50	.334	2.483*	.016

Anmerkung: Die Regressionsanalysen wurden jeweils getrennt für die Merkmale Testangst und Begabungskonzept vorgenommen. Es gehen die Mittelwerte aus beiden Items zur Erfassung der Testangst bzw. des Begabungskonzepts vor der Testsitzung ein. * $p < .05$; ** $p < .01$

Auch stellt das Treatment (Feedback/kein Feedback) keine *Moderatorvariable* bei Vorhersage der Testleistung zum zweiten Messzeitpunkt anhand des vor Testbeginn berichteten Begabungskonzepts dar (*interindividuelle Perspektive*, Nichtsignifikanz des Produktterms bzw. ΔR^2 : $R^2 = .10$, $F_{3/105} = 4.995$, $p < .05$, $\Delta R^2 = .002$, $F_{1/102} = .211$, $p = .647$, $N = 106$; das Vorgehen entspricht Abschnitt 6.2.1), unabhängig davon ob die intellektu-

elle Leistungsfähigkeit der Testanden mit der Variablen Intelligenzstatustest-Leistung kontrolliert wird. Die Homogenität der Fehlervarianzen ist dabei insgesamt sowie auf den Moderatorstufen gegeben (*Error Variance Ratio*: 1:1.02; Bartlett's $M = .0041$, $p = .95$). Unter den gegebenen Stichprobenverhältnissen muss ein mittlerer und starker Modereffekt ($f^2 \geq .08$) des Treatments mit hinreichender statistischer Sicherheit ($1 - \beta \geq .80$) ausgeschlossen werden. Der Zusammenhang zwischen situationsspezifischem Begabungskonzept und Testleistung unterscheidet sich demnach bei Messwiederholung nicht bedeutsam in Abhängigkeit der experimentellen Bedingungsvariation (Feedback/kein Feedback, siehe auch Tabellen 6.19 und 6.20). In einer Feedback- als auch in einer feedbackfreien Testbedingung geht ein hohes situationsspezifisches Begabungskonzept tendenziell mit hohen Testleistungen einher.

Tabelle 6.20

Regression der numerischen Reasoning-Testleistung der *Versuchsgruppe 2* (F+ F-) und *Kontrollgruppe* (F- F-; $N = 55$) auf die situationsspezifische Testangst sowie auf das situationsspezifische Begabungskonzept in einer *feedbackfreien* Testbedingung bei Testwiederholung – *interindividuelle* Perspektive

	R^2	F	df	β	t	p
Testangst	.04	3.022	1/54	-.232	-1.738	.088
Begabungskonzept	.12	8.028	1/54	.363	2.833*	.007

Anmerkung: analog zu Tabelle 6.19

Situationsspezifische Testangst: Bei einer Wiederholungstestung sollte, im Sinne einer Gewöhnung an die Testsituation, die Leistung des Testanden weniger von Testangst beeinflusst sein. Dieser Erwartung gemäß spielt in der feedbackfreien Testbedingung bei Testwiederholung die situationsspezifische Testangst (Emotionalität- und Worry-Komponente) keine Rolle bei der Vorhersage der Leistung (siehe Nicht-Signifikanz der Beta-Gewichte in der feedbackfreien Testbedingung zum zweiten Messzeitpunkt, Tabelle 6.20). Eine solche Reduktion des Zusammenhangs von Testangst und Leistung zum zweiten Messzeitpunkt, bzw. ein solcher angstreduzierender „Gewöhnungseffekt“, tritt jedoch in einer Testbedingung mit *Feedback* nicht auf (vgl. $\beta = -.29$ und $\beta = -.35$, *intraindividuelle Perspektive*). Testangst leistet hier weiterhin einen bedeutenden Beitrag zur Vorhersage der Testleistung (siehe Tabelle 6.19).

Allerdings lässt sich lediglich für den *ersten* Testzeitpunkt nachweisen, dass in einer Testbedingung mit *Feedback* Worry-Kognitionen ausgelöst werden (Tabelle 6.21). Nur unter *Feedback*-Bedingungen berichteten die Testanden nach der Testung ein höhe-

res Ausmaß an Worry-Kognitionen als vor der Testung. Auch nimmt unter feedbackfreien Testbedingungen, nicht jedoch unter Feedback-Bedingungen, die emotionale Komponente der Angst im Testverlauf ab. Zum *zweiten* Messzeitpunkt lässt sich aber eine Zunahme an Worry-Kognitionen unter Feedback-Bedingungen nicht mehr nachweisen ($M_1 = 4.4$, $M_2 = 3.3$, $Z = -.744$, $p = .457$, $N = 51$). Bei Testwiederholung wird demnach unter Feedback-Bedingungen nicht in höherem Ausmaß Testangst – speziell Worry-Kognitionen – berichtet als unter feedbackfreien Testbedingungen (Messwiederholungseffekt).

Tabelle 6.21

Situationsspezifische Testangst (Mediane) vor sowie nach der Reasoning-Testung (Zahlenfolgen) in einer feedbackfreien versus Feedback-Bedingung zum *ersten* Testzeitpunkt

	Komponente der Testangst	vor der Testung	nach der Testung	Z	p	N
Feedback	Emotionality	1.9	1.2	-.14	.889	28
	Worry	2.9	4.4	-2.382*	.017	28
kein Feedback	Emotionality	2.5	1.8	-3.219**	.001	78
	Worry	3.6	3.4	-.77	.441	78

Anmerkung: In die Analysen gehen die Items „I am afraid I may not do as well on this test as I could.“ versus „I am pretty optimistic that I will do as well on this test as I can.“ (Worry) und Item „Just now I am nervous about doing these problems.“ versus „Just now I am relaxed about doing these problems.“ (Emotionality) ein.

Die Interpretation der unterschiedlichen Zusammenhänge zwischen Testangst und Testleistung (siehe Signifikanzniveau der Zusammenhangsmaße, Tabelle 6.20) in den experimentellen Bedingungen muss zudem mit Vorsicht vorgenommen werden, da das Treatment (Feedback/kein Feedback) keine *Moderatorvariable* bei Vorhersage der Testleistung zum zweiten Messzeitpunkt anhand der vor Testbeginn berichteten Testangst darstellt (*interindividuelle Perspektive*, Nichtsignifikanz des Produktterms bzw. ΔR^2 : $R^2 = .06$, $F_{3/105} = 3.399$, $p < .05$, $\Delta R^2 = .001$, $F_{1/102} = .087$, $p = .769$, $N = 106$; das Vorgehen entspricht Abschnitt 6.2.1). Die Homogenität der Fehlervarianzen ist dabei insgesamt, sowie auf den Moderatorstufen gegeben (*Error Variance Ratio*: 1:1.2; Bartlett's $M = .43$, $p = .51$). Unter den gegebenen Stichprobenverhältnissen ist ein mittlerer und starker Moderatoreffekt ($f^2 \geq .08$) des Treatments mit hinreichender statistischer Sicherheit ($1 - \beta \geq .80$) auszuschließen. Der Zusammenhang zwischen situationsspezifischer Testangst – Worry-Komponente und Emotionality-Komponente – und Testleistung unterscheidet sich demnach bei Messwiederholung nicht signifikant in Abhängigkeit der experimen-

tellen Bedingungsvariation Feedback/kein Feedback (siehe auch Tabellen 6.19 und 6.20; *interindividuelle Perspektive*). In einer Feedback- als auch in einer feedbackfreien Testbedingung geht Testangst mit geringen Testleistungen einher.

Zusammenfassend ist festzuhalten: Es kann gezeigt werden, dass der in allen Testbedingungen bestehende Zusammenhang von situationsspezifischem Begabungskonzept und Leistung nur in der Feedback-Bedingung auf die Wirkung von Testangst zurückzuführen ist. Eine solche *Mediatorfunktion* kommt der Variablen Testangst in der feedbackfreien Testbedingung nicht zu. Dennoch kann Testangst für sich genommen auch in einer Testbedingung ohne Feedback einen Prädiktor für die Leistung darstellen, insbesondere wenn es sich um eine Ersttestung handelt (siehe Signifikanz des Beta-Gewichts in Test 1, Versuchsgruppe 1, F- F+).

Wird die Testangst als situationsübergreifendes Trait verstanden, lässt sich das Begabungskonzept auch als *Moderatorvariable* für den Zusammenhang von Testangst und Leistung unter Feedback konzipieren. Die Testangst sollte in einem solchen Moderatormodell nur leistungshemmend wirken, wenn sie mit einem geringen Begabungskonzept einhergeht, also nur auf bestimmten Moderatorstufen. Diese Erwartung wird jedoch durch die in der vorliegenden Studie gegebenen empirischen Tatsachen nicht gestützt. Testangst moderiert nicht den Zusammenhang von Begabungskonzept und Reasoning-Leistung unter Feedback-Bedingungen, unabhängig davon ob die intellektuelle Leistungsfähigkeit – operationalisiert als Intelligenzstatustest-Leistung – dabei kontrolliert wird (Nichtsignifikanz des Produktterms bzw. ΔR^2 , situationspezifische Maße: $R^2 = .46$, $F_{4/50} = 11.720$, $p < .001$, $\Delta R^2 = .007$, $F_{1/46} = .644$, $p = .426$, $N = 51$)³¹. Es kann ein mittlerer und starker Moderatoreffekt (genauer $f^2 \geq .17$) der Variablen Begabungskonzept für den Zusammenhang von Testangst und Leistung mit akzeptabler statistischer Sicherheit ($1 - \beta \geq .80$) ausgeschlossen werden. Ein *Moderatormodell* stellt also erwartungsgemäß – im Gegensatz zu einem *Mediatormodell* – kein angemessenes Modell der

³¹ Hier wurde analog zur Prüfung des Mediatormodells Versuchsgruppe 1 (F- F+, $N = 51$) betrachtet. Der Befund gilt auch für das Modell zur Vorhersage der *Schulleistungen* durch das schulbezogene Begabungskonzept. Die Testangst im Schulkontext erweist sich auch hier nicht als Moderator (Nichtsignifikanz des Produktterms für die Gesamtstichprobe: $R^2 = .09$, $F_{3/243} = 8.897$, $p < .001$, $\Delta R^2 = .01$, $F_{1/240} = 2.744$, $p = .099$, $N = 244$; auch nicht unter Kontrolle der Leistungsfähigkeit, $R^2 = .61$, $F_{4/243} = 96.15$, $p < .001$, $\Delta R^2 = .002$, $F_{1/239} = 1.134$, $p = .288$, $N = 244$).

empirischen Daten für die Zusammenhgangsstruktur der Variablen Begabungskonzept, Testangst und Reasoning-Leistung dar.

Die bisherigen Analysen bezogen sich hauptsächlich auf die unmittelbar vor der Testung erhobene Angstaussprägung (situationsspezifische Angst). Nun soll sich auf die unmittelbar nach absolvierter Testung erfragte Angstaussprägung konzentriert werden. Während ersteres als Erwartungsangst charakterisiert werden kann, reflektiert letzteres eher eine treatmentbedingte Veränderung des Angstniveaus. Dazu werden die Zusammenhänge zwischen Testangst, Begabungskonzept, Anstrengung und Reasoning-Leistung in den experimentellen Bedingungen Feedback/kein Feedback gegenübergestellt. Hierzu ist Folgendes anzumerken: Die Analyse von Moderatoren kann mittels verschiedener statistischer Methoden vorgenommen werden. In der vorliegenden Arbeit wurde die moderierte multiple Regression bevorzugt (siehe Kapitel 4). Der Vergleich von Korrelationskoeffizienten in den nach einem Moderator (z.B. Feedback/kein Feedback) gebildeten Subgruppen stellt eine weitere Methode³² zur Aufdeckung von Moderatorvariablen dar (siehe Jäger, 1978).

Die Tabellen 6.22 (*intraindividuelle Perspektive*) und 6.23 (*interindividuelle Perspektive*) stellen noch einmal die Zusammenhänge – hier als Korrelationskoeffizienten – des jeweils vor und nach der Testung berichteten situationsspezifischen Begabungskonzepts, der Testangst und Anstrengungsvermeidung für den Materialbereich Zahlenfolgen in den experimentellen Bedingungen Feedback/kein Feedback als Zusammenschau gegenüber.

Bei Vergleich der *Versuchsgruppe 1* (F– F+) in den Testbedingungen mit versus ohne Feedback (*intraindividuelle Perspektive*, Tabelle 6.22) fallen wieder (siehe oben) die

³² Die Methode der multiplen moderierten Regression wird hier genutzt, um die Bedeutsamkeit der unterschiedlichen Korrelationsmuster in den Feedback- und feedbackfreien Bedingungen abzuschätzen. Die multiple moderierte Regression ist der Berechnung von Gruppen-Korrelationskoeffizienten (auch als Fraktionierungsmethode bezeichnet, Jäger, 1978) zur Aufdeckung von Moderatorvariablen vorzuziehen (Bartussek, 1970; Baron & Kenny, 1986). Als ein Nachteil der Fraktionierungsmethode wird beispielsweise der größere Stichprobenfehler in den Substichproben genannt. Baron und Kenny (1986) stellen die Abhängigkeit der Korrelationskoeffizienten, im Gegensatz zu Regressionskoeffizienten, von der möglicherweise in den Subgruppen, d.h. auf den Moderatorstufen, unterschiedlichen Varianz des Prädiktors heraus. Unterschiedliche Ergebnisse beider Methoden sind demnach denkbar (zu unterschiedlichen Ergebnissen der Signifikanzprüfung von Unterschieden zwischen Korrelationskoeffizienten im Gegensatz zu Regressionskoeffizienten siehe auch Cohen et al., 2003, S. 47). Regressionskoeffizienten gelten als stabilere Maße, da sie weniger als Korrelationskoeffizienten von der Varianz des Prädiktors in der Population abhängen.

geringeren Zusammenhänge zwischen der *vor* der Testung angegebenen Leistungserwartung (Begabungskonzept) und der in einem Test mit Feedback tatsächlich gezeigten Leistung auf. Der zum zweiten Messzeitpunkt in der Feedback-Bedingung in Versuchsgruppe 1 (F– F+) beobachtete geringere Zusammenhang von Begabungskonzept und Leistung kann jedoch nicht eindeutig auf das Feedback zurückgeführt werden, sondern sollte eher als Effekt der Messwiederholung interpretiert werden. Die Bewertung dieses Befundes wurde bereits vorgenommen.

Die Leistungseinschätzungen der Testanden *nach* der Testung stimmen dagegen unabhängig von der experimentellen Bedingungsvariation Feedback/kein Feedback hoch mit ihrer tatsächlichen Leistung überein. Das heißt, nach der Testung haben Testanden in der Regel ein realistisches Bild über die gezeigte Leistung, interessanterweise unabhängig davon, ob sie Feedback erhalten haben oder nicht. Dies unterstützt die Annahme, dass Richtig/Falsch-Feedback zu Recht als informationsarmes Feedback charakterisiert werden kann. Unter Abschnitt 6.2.3 soll noch genauer auf diesen Befund eingegangen werden.

Tabelle 6.22

Zusammenhang der situationsspezifischen Selbsteinschätzung in der Feedback- versus feedbackfreien Testbedingung mit der numerischen Reasoning-Testleistung, Versuchsgruppe 1 (F– F+, N = 51) – intraindividuelle Perspektive

		Test 1 (kein Feedback)	Test 2 (Feedback)
vor der Testung	Begabungskonzept (Fähigkeit)	.57**	.35*
	Begabungskonzept (Lernen)	.47**	.18
	Testangst (Worry)	-.19	-.29*
	Testangst (Emotionality)	-.31*	-.31*
nach der Testung	Begabungskonzept (Fähigkeit)	.56**	.64**
	Testangst (Worry)	-.48**	-.70**
	Testangst (Emotionality)	-.29*	-.41**
	Anstrengung	.23	.44**

Anmerkung: Spearmans r_s , * = $p < .05$, ** = $p < .01$

Bei Betrachtung von Tabelle 6.22 lässt sich vermuten, dass in der Feedback-Bedingung im Vergleich zur feedbackfreien Bedingung beobachtete Leistungseinbußen bei hoher Testangst speziell auf die *Worry*-Komponente der Testangst zurückgehen (vgl. vor Testbeginn: Spearmans $r_s = -.19$ und $r_s = -.29$, Tabelle 6.22).

Unter einer *interindividuellen* Perspektive (Tabelle 6.23) zeigt sich aber, dass vor als auch nach der Testung angegebene Worry-Kognitionen nicht nur in einer Testbedin-

gung mit Feedback, sondern durchaus auch in einer feedbackfreien Testbedingung mit geringen Leistungen einhergehen (siehe auch Tabelle 6.19 und Tabelle 6.20). Auffällig bleiben jedoch die besonders hohen negativen Zusammenhänge der Testleistung mit den in der Feedback-Bedingung im Anschluss an die Testung berichteten Worry-Kognitionen, verglichen mit einer feedbackfreien Testung. Die Korrelationskoeffizienten unterscheiden sich sowohl zum ersten als auch zum zweiten Messzeitpunkt signifikant (Test 1: $Z = -2.37, p = .01$; Test 2: $Z = -2.22, p = .01$).

Nur für die Ersttestung kann jedoch geschlussfolgert werden, dass Feedback vermehrt zu Worry-Kognitionen und infolge zu Leistungseinbußen führt (siehe Medianvergleich Tabelle 6.21). Bei Wiederholungsmessung zeigen sich niedrige Testleistungen auch in der Ausprägung der nach dem Test berichteten Emotionality-Komponente der Testangst, und zwar ausschließlich in der Feedback-Bedingung.

Tabelle 6.23

Zusammenhang der situationsspezifischen Selbsteinschätzungen und der numerischen Reasoningtest-Leistung in der Feedback- versus feedbackfreien Testbedingung zum ersten und zweiten Messzeitpunkt – interindividuelle Perspektive

			Feedback	N_{F+}	kein Feedback	N_{F-}
Test 1	vor der Testung	Begabungskonzept (Fähigkeit)	.53**	28	.49**	77
		Testangst (Worry)	-.18	28	-.26*	77
	nach der Testung	Begabungskonzept (Fähigkeit)	.64**	28	.58**	77
		Testangst (Worry)	-.75**	28	-.40**	77
Test 2	vor der Testung	Begabungskonzept (Fähigkeit)	.35*	51	.38**	55
		Testangst (Emotionality)	-.31*	51	.05	55
		Testangst (Worry)	-.29*	51	-.36**	55
	nach der Testung	Begabungskonzept (Fähigkeit)	.64**	51	.58**	55
		Testangst (Emotionality)	-.41**	51	-.09	55
		Testangst (Worry)	-.70**	51	-.40**	55
		Anstrengung	.44**	51	.15	55

Anmerkung: Spearmans r_s , * = $p < .05$, ** = $p < .01$

Die während der Testung investierte *Anstrengung* geht mit einer erfolgreichen Bearbeitung eines Tests mit Feedback einher (Tabelle 6.22). In der feedbackfreien Testbedingung findet sich dagegen kein Zusammenhang von Anstrengung und Leistung. Dieser Befund zeigt sich auch bei einem interindividuellen Gruppenvergleich (Tabelle 6.23), trifft aber nur für die Wiederholungstestung zu.

6.2.2.4.3.2 Schulbezogene Selbsteinschätzung

Werden die „trait-näheren“ schulbezogenen Selbsteinschätzungen – erhoben in Sitzung 2 – betrachtet, stellt sich heraus, dass in der *Feedback*-Bedingung der *Versuchsgruppe 1* (F–F+) weder das *schulbezogene* Begabungskonzept (Mittelwerte beider Items nach Dweck) noch die *schulbezogene* Testangst (Summenwerte der Skala Testangst) zur Vorhersage der Testleistung beiträgt (Begabungskonzept: $\beta = .207$, $t = 1.469$, $p = .148$, $N = 50$; Testangst: $\beta = -.261$, $t = -1.855$, $p = .07$, $N = 49$).

In der *feedbackfreien* Testbedingung derselben Stichprobe (Versuchsgruppe 1, F–F+) lässt sich lediglich das schulbezogene Begabungskonzept, nicht aber die schulbezogene Testangst, eindeutig als Prädiktor für die Testleistung ausmachen (Begabungskonzept: $\beta = .400$, $t = 3.027$, $p = .004$, $N = 50$; Testangst: $\beta = -.265$, $t = -1.887$, $p = .07$, $N = 49$). Immerhin 14% der Varianz der Testleistung werden allein anhand des schulbezogenen Begabungskonzepts erklärt ($F_{1/49} = 9.162$, $p = .004$). Da Testangst nicht zur Vorhersage der Testleistung beiträgt – weder in der *Feedback*-Bedingung, noch in der *feedbackfreien* Bedingung – kann der *schulbezogenen* Testangst keine *Mediatorfunktion* für den beobachteten Zusammenhang von schulbezogenem Begabungskonzept und Testleistung zugeschrieben werden.

Zur Prüfung, ob es sich dabei um einen *Wiederholungstestungseffekt* handelt, werden wieder die *Versuchsgruppe 2* (F+ F–) und die *Kontrollgruppe* (F– F–) zum zweiten Messzeitpunkt – jeweils die *feedbackfreie* Testbedingung – betrachtet. Auch hier erweist sich das schulbezogene Begabungskonzept als ein wesentlicher Prädiktor für die Testleistung ($\beta = .316$, $t = 2.426$, $p = .019$, $N = 55$), nicht jedoch die schulbezogene Testangst ($\beta = -.171$, $t = -1.261$, $p = .213$, $N = 55$). Somit erhärtet sich der Befund, dass die schulbezogene Testangst nicht als erklärende Variable für den Zusammenhang von Begabungskonzept und Testleistung herangezogen werden kann.

Zusammenfassend kann festgehalten werden: Anhand der Instrumente zur Erfassung sowohl der schulbezogenen Testangst als auch des schulbezogenen Begabungskonzepts lässt sich hypothesenkonform zeigen, dass das *schulbezogene Begabungskonzept* nur in der *feedbackfreien* Testbedingung zur Vorhersage der Testleistung beiträgt, nicht aber in einer Test-Bedingung mit *Feedback*. Die *schulbezogene Testangst* steht in den hier betrachteten Substichproben (Versuchsgruppe 1 zu Test 1 und Test 2, Versuchsgruppe 2 und Kontrollgruppe zu Test 2) in keinem Zusammenhang zur Testleistung, unabhängig

von der experimentellen Bedingungsvariation, und kann somit auch keinen Mediator für den Zusammenhang von Begabungskonzept und Testleistung darstellen.

Die Tabellen 6.24 und 6.25 zeigen die Zusammenhänge von schulbezogener Selbsteinschätzung (Begabungskonzept, Anstrengungsvermeidung, Testangst) und numerischer Reasoning-Testleistung in den experimentellen Gruppen Feedback/kein Feedback im Überblick. Bei gemeinsamer Betrachtung beider Tabellen wird noch einmal deutlich, dass das schulbezogene *Begabungskonzept*, unabhängig vom Messzeitpunkt – also zur Ersttestung als auch zur Messwiederholung – in der feedbackfreien Testbedingung, nicht jedoch in der Feedback-Bedingung, mit der Reasoning-Leistung korreliert ist (siehe oben). Demnach stimmt die *schulbezogene* Selbsteinschätzung tendenziell mit der tatsächlichen Leistung der Schüler in der feedbackfreien Testbedingung überein, nicht jedoch in einer Testbedingung mit Feedback.

Tabelle 6.24

Zusammenhang von *schulbezogenem* Begabungskonzept, Anstrengungsvermeidung und numerischer Reasoningtest-Leistung in der Feedback- versus feedbackfreien Testbedingung, *Versuchsgruppe 1* (F– F+, $N = 50$) – *intraindividuelle* Perspektive

	Test 1 (kein Feedback)	Test 2 (Feedback)
Anstrengungsvermeidung	-.27	-.36**
Begabungskonzept	.49**	.21

Anmerkung: Korrelationskoeffizienten nach Pearson. Die Personmerkmale gehen als Faktorwerte in die Analyse ein; * = $p < .05$, ** = $p < .01$

Die schulbezogene *Anstrengungsvermeidung* scheint mit geringen Leistungen in einer Testvariante mit Feedback einherzugehen, nicht jedoch in einer Bedingung ohne Feedback (Tabelle 6.24). Anhand der vorliegenden Daten lässt sich dies aber nur für eine Paralleltestung mit Feedback schlussfolgern, da zum ersten Testzeitpunkt in der Feedback-Bedingung kein entsprechender Zusammenhang zwischen schulbezogener Anstrengungsvermeidung und Testleistung besteht (Tabelle 6.25). Die schulbezogene *Testangst* korreliert nur mit einer unter Feedback-Bedingung entstandenen Testleistung aus einer Ersttestung. Zum zweiten Testzeitpunkt steht die schulbezogene Testangst in keinem Zusammenhang mehr zur Leistung (Tabelle 6.25).

Tabelle 6.25

Zusammenhang von *schulbezogenem* Begabungskonzept, Testangst, Anstrengungsvermeidung und numerischer Reasoningtest-Leistung in der Feedback- versus feedbackfreien Testbedingung zum ersten und zweiten Messzeitpunkt – *interindividuelle* Perspektive

		Feedback	N_{F+}	kein Feedback	N_{F-}
Test 1	Begabungskonzept	.26	28	.39**	77
	Testangst	-.40*	28	-.13	77
	Anstrengungsvermeidung	-.25	28	-.21	77
Test 2	Begabungskonzept	.21	50	.36**	55
	Testangst	-.21	50	-.15	55
	Anstrengungsvermeidung	-.36**	50	-.17	55

Anmerkung: Korrelationskoeffizienten nach Pearson. Die Personenmerkmale gehen als Faktorwerte in die Analyse ein; * = $p < .05$, ** = $p < .01$

Die Interpretation der unterschiedlichen Zusammenhänge zwischen schulbezogenem *Begabungskonzept* und Testleistung (siehe Signifikanzniveau der Zusammenhangsmaße, Tabelle 6.24 und Tabelle 6.25) in den experimentellen Bedingungen Feedback/kein Feedback muss jedoch wieder mit Vorsicht vorgenommen werden, da auch hier das Treatment (Feedback/kein Feedback) keine *Moderatorvariable* bei Vorhersage der Testleistung zum zweiten Messzeitpunkt anhand des schulbezogenen Begabungskonzepts (Faktorwerte) darstellt (*interindividuelle Perspektive*, Nichtsignifikanz des Produktterms bzw. ΔR^2 : $R^2 = .07$, $F_{3/103} = 3.735$, $p < .05$, $\Delta R^2 = .01$, $F_{1/100} = 1.018$, $p = .315$, $N = 104$; das Vorgehen entspricht Abschnitt 6.2.1), unabhängig davon ob die intellektuelle Leistungsfähigkeit der Testanden dabei kontrolliert wird. Die Homogenität der Fehlervarianzen ist dabei insgesamt sowie auf den Moderatorstufen gegeben (*Error Variance Ratio*: 1:1.05; Bartlett's $M = .02$, $p = .87$). Unter den gegebenen Stichprobenverhältnissen muss ein mittlerer und starker Moderatoreffekt ($f^2 \geq .08$) des Treatments mit hinreichender statistischer Sicherheit ($1 - \beta \geq .80$) ausgeschlossen werden. Der Zusammenhang zwischen schulbezogenem Begabungskonzept und Testleistung unterscheidet sich demnach bei Wiederholungsmessung nicht signifikant in Abhängigkeit der experimentellen Bedingungsvariation Feedback/kein Feedback.

Insgesamt ergeben sich Hinweise darauf, dass in feedbackfreien Testbedingungen vor allem ein hohes schulbezogenes Begabungskonzept leistungsförderlich ist, für Testbedingungen mit Richtig/Falsch-Feedback scheint dagegen die schulbezogene Anstrengungsvermeidung, sowie bei Ersttestung die schulbezogene Testangst leistungsbeeinträchtigend zu wirken.

6.2.2.4.4 Figurale Reasoning-Testleistung

6.2.2.4.4.1 Situationsspezifische Selbsteinschätzung

Das situationsspezifische Begabungskonzept und die situationsspezifische Testangst stehen in *keinem* Zusammenhang zur figuralen Reasoning-Leistung, weder in der *Feedback*-Bedingung (Begabungskonzept: $\beta = .119$, $t = 1.023$, $p = .310$, $N = 75$; Testangst: $\beta = .054$, $t = .460$, $p = .647$, $N = 75$) noch in der *feedbackfreien* Bedingung der *Versuchsgruppe 1* (F- F+; Begabungskonzept: $\beta = .01$, $t = .088$, $p = .930$, $N = 75$; Testangst: $\beta = .016$, $t = .137$, $p = .891$, $N = 75$). Eine Mediator- oder Moderatorfunktion der Variablen Testangst für den Zusammenhang von Begabungskonzept und figuraler Reasoning-Leistung lässt sich folglich weder in der Feedback-Bedingung noch in der feedbackfreien Testbedingung ausmachen. Zur Klärung des Ausbleibens des Mediatoreffekts sollen die Zusammenhänge von Begabungskonzept und Testangst in den Bedingungen Zahlenfolgen und Figurenfolgen verglichen werden (Versuchsgruppe 1, F- F+, Tabelle 6.26).

Tabelle 6.26

Zusammenhang zwischen situationsspezifischem Begabungskonzept und situationsspezifischer Testangst bei Bearbeitung der *numerischen* versus *figuralen* Reasoning-Aufgaben in der Feedback-Bedingung der *Versuchsgruppe 1* (F- F+)

	Zahlenfolgen (N = 51)		Figurenfolgen (N = 75)	
	Begabungskonzept vor Test 1	Testangst vor Test 2	Begabungskonzept vor Test 1	Testangst vor Test 2
Begabungskonzept vor Test 2	.56**	-.58**	.51**	-.14

Anmerkung: * = $p < .05$, ** = $p < .01$

Tabelle 6.26 zeigt, dass in der Bedingung *Figurenfolgen* unter Feedback-Bedingungen kein Zusammenhang zwischen Testangst und Begabungskonzept besteht, wohl aber in der Bedingung *Zahlenfolgen*. Die signifikanten Zusammenhänge zwischen den vor der ersten und vor der zweiten Testung erfragten Einschätzungen der eigenen Fähigkeit weisen auf Stabilität im Antwortverhalten der Testanden hin. Der ausbleibende Zusammenhang zwischen Begabungskonzept und Testangst in der Bedingung *Figurenfolgen* lässt sich natürlich nicht auf die mangelnde Reliabilität der figuralen Reasoning-Items zurückführen. Vielmehr ist davon auszugehen, dass bei Konfrontation mit den figuralen Items wenig selbstkonzeptstarke Testanden nicht vermehrt Testangst berichteten, wie bei Konfrontation mit numerischen Reasoning-Items beobachtet. Folglich lässt

sich auch keine *Mediatorfunktion* der Variablen Testangst für den Zusammenhang von Begabungskonzept und figuraler Reasoning-Leistung feststellen.

Wird die *feedbackfreie* Leistung der Versuchsgruppe 2 (F+ F-) und Kontrollgruppe (F- F-) betrachtet (jeweils Test 2), fällt auf, dass die Testangst ($R^2 = .03$, $F_{1/66} = 2.937$, $p = .091$) sowie das Begabungskonzept ($R^2 = .03$, $F_{1/66} = 3.243$, $p = .076$) ebenfalls nicht zur Vorhersage der feedbackfreien figuralen Reasoning-Leistung beitragen. Eine Mediator- oder Moderatorfunktion der Variablen Testangst für den Zusammenhang von Begabungskonzept und figuraler Reasoning-Leistung lässt sich somit auch bei *Wiederholungsmessung* in der feedbackfreien Testung nicht ausmachen, unabhängig davon, ob die intellektuelle Leistungsfähigkeit – operationalisiert als Intelligenzstatustest-Leistung – dabei kontrolliert wird (*Moderatormodell*: Nichtsignifikanz der Produktterms: $\Delta R^2 = .01$, $F_{1/63} = .699$, $p = .406$, $N = 67$; *Mediatormodell*: keine Verringerung, sondern Vergrößerung des Vorhersagebeitrags des Begabungskonzepts unter Kontrolle des potentiellen Mediators Testangst, vgl. $\beta = -.218$, $t = -1.801$, $p = .076$ und $\beta = -.273$, $t = -2.271$, $p < .05$, $N = 67$).

Im Folgenden sollen wieder die nach der Testung angegebenen Einschätzungen der Testanden, in denen sich eventuell treatmentbedingte Veränderungen der Selbsteinschätzung zeigen, genauer betrachtet werden.

Tabelle 6.27 und 6.28 stellen die Zusammenhänge des nach der Testung berichteten situationsspezifischen Begabungskonzepts, der Testangst und der Anstrengungsvermeidung für den Materialbereich Figurenfolgen in den experimentellen Bedingungen gegenüber. Unter einer *intraindividuellen* Perspektive (Versuchsgruppe 1, F- F+) fiel auf, dass die *vor* der figuralen Reasoning-Testung erfassten Selbsteinschätzungen/Erwartungen in keinem signifikanten Zusammenhang zur Testleistung stehen, und zwar unabhängig von der experimentellen Bedingungsvariation Feedback/kein Feedback. Auf die Darstellung der Korrelationskoeffizienten wurde deshalb verzichtet. Die Bewertung des erreichten Leistungsniveaus der Testanden *im Anschluss* an die Testung stimmt ausschließlich in der *Feedback*-Bedingung mit der tatsächlichen Leistung überein (Tabelle 6.27).

Tabelle 6.27

Zusammenhang der situationsspezifischen Selbsteinschätzung in der Feedback- versus feedbackfreien Testbedingung mit der *figuralen* Reasoning-Testleistung in *Versuchsgruppe 1* (F– F+, N =75) – *intraindividuelle* Perspektive

		Test 1 (<i>feedbackfrei</i>)	Test 2 (<i>Feedback</i>)
nach der Testung	Begabungskonzept (Fähigkeit)	.17	.28*
	Begabungskonzept (Lernen)	.00	.39**
	Testangst (Worry)	-.10	-.27*
	Testangst (Emotionality)	-.06	-.19
	Anstrengung	.16	.23*

Anmerkung: Spearmans r_s , * = $p < .05$, ** = $p < .01$

Ein *interindividueller* Gruppenvergleich (Tabelle 6.28) führt zu vergleichbaren Ergebnissen. Insbesondere zum *zweiten* Messzeitpunkt schätzten nur Testanden, die einen Test mit Leistungsfeedback bearbeitet hatten, ihr Leistungsniveau nach der Testung realistisch ein. Es konnten auch hier keine signifikanten Zusammenhänge zwischen vor Testbeginn erfasster Testangst, Anstrengungsvermeidung und Leistung aufgedeckt werden. Auf die Darstellung dieser Korrelationskoeffizienten wurde deshalb ebenfalls verzichtet (Tabelle 6.28).

Die Nullkorrelationen zwischen Begabungskonzept und Testleistung in den *feedbackfreien* Testbedingungen (Tabelle 6.27: Test 1; Tabelle 6.28: Test 2) zeigen, dass Testanden bei Bearbeitung der *figuralen* Reasoning-Aufgaben im Gegensatz zu numerischen Reasoning-Aufgaben ohne Feedback ihre Leistung nicht angemessen bewerteten.

Tabelle 6.28

Zusammenhang der situationsspezifischen Selbsteinschätzung und der *figuralen* Reasoningtest-Leistung in der Feedback- versus feedbackfreien Testbedingung zum ersten und zweiten Messzeitpunkt – *interindividuelle* Perspektive

			Feedback	N_{F+}	kein Feedback	N_{F-}
Test 1	nach der Testung	Begabungskonzept (Fähigkeit)	.34*	35	.22*	107
Test 2	nach der Testung	Begabungskonzept (Fähigkeit)	.28*	75	.07	67
		Begabungskonzept (Lernen)	.39**	75	-.09	67
		Testangst (Worry)	-.27*	75	-.21	67
		Anstrengung	.23*	75	.11	67

Anmerkung: Spearmans r_s , * = $p < .05$, ** = $p < .01$

Wie auch bei Bearbeitung der numerischen Reasoning-Aufgaben gehen nach einer Testung mit Feedback angegebene *Worry-Kognitionen* mit geringen *figuralen* Reaso-

ning-Leistungen einher. Ein hohes Ausmaß an berichteter *Anstrengung* während der Testung zeigt dagegen insbesondere bei Feedback-Angebot hohe Testleistungen an.

Bei Betrachtung der Mediane fällt wie auch bei Analyse der Zahlenfolgen auf, dass ausschließlich in der *Feedback*-Bedingung der Anteil an berichteten Worry-Kognitionen im Testverlauf zunimmt (siehe Tabelle 6.29, auch in Test 2: $M_1 = 3.6$, $M_2 = 4.2$, $Z = -2.227$, $p = .026$, $N = 75$). Dagegen nimmt der Anteil an Emotionality im Testverlauf nur in der feedbackfreien Testbedingung signifikant ab (Tabelle 6.29; auch in Test 2: $M_1 = 3.25$, $M_2 = 1.25$, $Z = -3.88$, $p < .001$, $N = 68$).

Tabelle 6.29

Situationspezifische Testangst (Mediane) vor sowie nach der Reasoning-Testung (Figurenfolgen) in einer feedbackfreien versus Feedback-Bedingung zum *ersten* Testzeitpunkt

	Komponente der Testangst	vor der Testung	nach der Testung	Z	p	N
Feedback	Emotionality	3.4	3.6	-1.008	.313	36
	Worry	2.5	4.6	-3.226**	.001	36
kein Feedback	Emotionality	3.5	2.1	-3.655**	.000	107
	Worry	3.5	3.5	-.897	.37	107

Anmerkung: In die Analysen gehen die Items „I am afraid I may not do as well on this test as I could.“ versus „I am pretty optimistic that I will do as well on this test as I can.“ (Worry) und Item „Just now I am nervous about doing these problems.“ versus „Just now I am relaxed about doing these problems.“ (Emotionality) ein.

6.2.2.4.4.2 Schulbezogene Selbsteinschätzungen

Zwischen den Maßen zur Erfassung der schulbezogenen Selbsteinschätzung und der figuralen Reasoning-Testleistung bestehen, neben einem erwartungswidrigen Befund, keine systematischen Zusammenhänge. Bei Wiederholungsmessung und zwar nur in der feedbackfreien Testbedingung ging lediglich die schulbezogene *Anstrengungsvermeidung* überraschend mit hohen Testleistungen einher. Dieser Befund lässt sich vor allem auf die Kontrollgruppe (F- F-) zurückführen ($r = .40$, $p < .05$, $N = 30$). Die Wiederholungstestung kann eventuell Langweile oder Widerstand der Testanden gegen die Testung ausgelöst haben. Möglicherweise haben sonst bemühte, d.h. nicht-anstrengungsvermeidende Testanden, in der feedbackfreien Wiederholungstestung deshalb wenig Anstrengung gezeigt und infolgedessen geringe Leistungen erzielt.

Tabelle 6.30

Zusammenhang von *schulbezogener* Anstrengungsvermeidung und figuraler Reasoningtest-Leistung in der Feedback- versus feedbackfreien Testbedingung zum ersten und zweiten Messzeitpunkt

		Reasoning-Testleistung			
		Feedback	N_{F+}	kein Feedback	N_{F-}
Anstrengungsvermeidung	Test 1	.09	35	-.06	107
	Test 2	-.15	75	.31*	67

Anmerkung: Anstrengungsvermeidung ging als Faktorwert in die Analyse ein; * = $p < .05$, ** = $p < .01$

Zusammenfassend wird festgehalten: Das *schulbezogene* Begabungskonzept leistet keinen Beitrag zur Vorhersage der unter Feedback-Bedingungen entstandenen numerischen Reasoningtest-Leistung. Eine Analyse der entsprechenden *situationsspezifischen* Maße zeigt aber erwartungskonform, dass die situationspezifisch erfasste Einschätzung der eigenen Leistungsfähigkeit (Begabungskonzept) auch in einer Testbedingung mit Feedback von Relevanz ist und Leistungsvarianz aufklärt. Die feedbackbedingten numerischen Reasoning-Leistungen stimmen im Gegensatz zu den feedbackfreien Leistungen offensichtlich nicht mit dem *schulbezogenen* Fähigkeitsselbstbild überein. Die regressionsanalytische Auswertung im Rahmen des Moderatorenansatzes widerspricht allerdings der Annahme einer stärkeren Determination der feedbackfreien Reasoning-Leistung durch das schulbezogene Begabungskonzept im Vergleich zu einer unter Feedback-Bedingungen zustande gekommenen Testleistung.

Analog stellt auch die *schulbezogene* Test- oder Prüfungsangst keinen Prädiktor für die numerische Reasoning-Testleistung dar, wohl aber die *situationsspezifisch* berichtete Testangst. Die vor Testbeginn erfragte Testangst erklärt immerhin bis zu 11% der Leistungsvarianz in einem numerischen Reasoning-Test. Bei Messwiederholung verliert die Testangst in der feedbackfreien Testbedingung an Bedeutung, nicht jedoch in einer Testbedingung mit Feedback. Der Zusammenhang zwischen Begabungskonzept und Reasoning-Leistung wird durch die Testangst mediert. Unter Feedback bedingt ein niedriges Begabungskonzept Testangst und beeinflusst die Manifestation der Reasoningkapazität negativ (*Mediatorhypothese Testangst*). Die Tatsache, dass dieses Ergebnis nicht für die figurale Reasoning-Leistung gefunden wurde, lässt sich durch die kompensatorische Wirkung des Multiple-Choice-Antwortformats erklären.

Wie auch unter Abschnitt 6.2.2.1 diskutiert, ergeben sich also unterschiedliche Befundmuster für die beiden betrachteten Materialbereiche (figural, numerisch). Weder die schulbezogene oder situationspezifische Testangst noch das schulbezogene oder

situationsspezifische Begabungskonzept konnten als Prädiktoren der *figuralen Reasoning*-Leistung in einer Feedback-Bedingung ausgemacht werden.

Als mögliche Ursachen für die divergenten Befundmuster lassen sich bereits diskutierte, inhaltliche und formale Gründe heranziehen. Aufgrund der geringen Schwierigkeit der figuralen Reasoning-Items und/oder aufgrund des relativ geringen Bekanntheitsgrades der entsprechenden Items und eines damit einhergehenden wenig ausgeprägten Fähigkeits-Selbstkonzepts, wurden die figuralen Testversionen von den Testanden möglicherweise als wenig selbstwertbedrohlich bzw. potentiell angstauslösend wahrgenommen (siehe auch Multiple-Choice-Antwortformat). Folglich finden sich auch keine Zusammenhänge zwischen Begabungskonzept und Testangst. Die mangelnde Reliabilität der Figurenfolgen stellt jedoch nach wie vor eine mögliche Hauptursache für die fehlenden Wechselbeziehungen dar. Diese Erklärung ist auch deshalb nahe liegend, da auffällig wenig signifikante Zusammenhänge zwischen figuraler Leistung und Persönlichkeitsmaßen auftreten.

6.2.3 Zum Informationsgehalt von Richtig/Falsch-Feedback

Zahlenfolgen: Tabelle 6.31 zeigt, dass Testanden, die numerische Reasoning-Aufgaben bearbeitet haben, ihre Leistung bereits zum ersten Messzeitpunkt auch ohne Leistungsfeedback realistisch beurteilen. Richtig/Falsch-Feedback bietet hier für die Testanden offensichtlich *keine* neue Information und ist vielleicht auch deshalb nicht effektiv.

Tabelle 6.31

Sie wissen, was sie tun: Zusammenhang von situationsspezifischer Selbsteinschätzung und numerischer Reasoning-Leistung in der *feedbackfreien* Testbedingung zum *ersten* Messzeitpunkt

	<i>vor der Testung</i>	<i>nach der Testung</i>
Begabungskonzept (Fähigkeit)	.49**	.58**
Begabungskonzept (Lernen)	.35**	.27*
Leistungsziel – Zielerreichung nach der Testung	---	.48**
Lernziel – Zielerreichung nach der Testung	---	.27*

Anmerkung: Versuchsgruppe 1 (F– F+) und Kontrollgruppe (F– F–) zum ersten Messzeitpunkt ($N = 78$); * = $p < .05$, ** = $p < .01$; interindividueller Leistungsvergleich im Sinne der Erreichung eines Leistungsziels: Item „I think I gave more correct answers than most of the other students.“ versus „I think most of the other students gave more correct answers than me.“; intraindividueller Leistungsvergleich im Sinne der Erreichung eines Lernziels: Item „I have learned something interesting.“ versus „I have not learned anything interesting.“

Die Korrelationen von situationsspezifischem und schulbezogenem Begabungskonzept in den experimentellen Bedingungen (Tabelle 6.32) zeigen, dass die Einschät-

zung der eigenen Leistung im Anschluss an die Testung in der Feedback-Bedingung weniger stark mit dem schulbezogenen Begabungskonzept zusammenhängt, als in einer Testbedingung ohne Feedback (einseitige Signifikanzprüfung korrelierter Korrelationskoeffizienten nach Ferguson: $t_{47} = 1.805$, $p = .039$, $N = 50$). Mittels Feedback lässt sich demzufolge die Selbsteinschätzung der Testanden durchaus beeinflussen.

Tabelle 6.32

Zusammenhang von situationsspezifischem und schulbezogenem Begabungskonzept in der Feedback- versus feedbackfreien Testbedingung der *Versuchsgruppe 1* (F- F+, $N = 50$) in der Bedingung Zahlenfolgen

		schulbezogenes Begabungskonzept	
		Test 1 (feedbackfrei)	Test 2 (Feedback)
situationsspezifisches Begabungskonzept	vor der Testung	.50**	.48**
	nach der Testung	.60**	.39**

Anmerkung: situationsspezifisches Begabungskonzept: Item 1 (Fähigkeit: „I feel pretty confident that I shall be able to solve most of the problems.“); schulbezogenes Begabungskonzept: Faktorwerte; * = $p < .05$, ** = $p < .01$

Figurenfolgen: Testanden, die figurale Reasoning-Aufgaben bearbeitet haben, beurteilen ihre Leistung zum ersten Messzeitpunkt ohne Leistungsfeedback weniger realistisch als Testanden, die numerische Reasoning-Aufgaben bearbeitet haben (Tabelle 6.33). Richtig/Falsch-Feedback bietet hier mindestens potentiell neue Information und ist vielleicht auch deshalb für eine Teilgruppe der Testanden – leistungsstarke Testanden – leistungsförderlich (siehe Abschnitt 6.2.1). Ein Grund dafür mag sein, dass es bei Figurenfolgen schwieriger ist *Error-Checking*-Routinen in den Bearbeitungsprozess zu integrieren, wohingegen bei Zahlenfolgen „Kontrollrechnungen“ leicht zu realisieren sind.

Tabelle 6.33

Zusammenhang von situationsspezifischer Selbsteinschätzung und figuraler Reasoning-Leistung in der *feedbackfreien* Testbedingung zum *ersten* Messzeitpunkt

	vor der Testung	nach der Testung
Begabungskonzept (Fähigkeit)	-.03	.22*
Begabungskonzept (Lernen)	-.10	.03
Leistungsziel – Zielerreichung nach der Testung	--	.07
Lernziel – Zielerreichung nach der Testung	--	.17

Anmerkung: Versuchsgruppe 1 (F- F+) und Kontrollgruppe (F- F-) zum ersten Messzeitpunkt ($N = 107$); * = $p < .05$, ** = $p < .01$; interindividueller Leistungsvergleich im Sinne der Erreichung eines Leistungsziels: Item „I think I gave more correct answers than most of the other students.“ versus „I think most of the other students gave more correct answers than me.“; intraindividueller Leistungsvergleich im Sinne der Erreichung eines Lernziels: Item „I have learned something interesting.“ versus „I have not learned anything interesting.“

Möglicherweise besitzen Testanden auch a priori kaum ein Fähigkeits-Selbstkonzept in Bezug auf figurale Reasoning-Aufgaben und schätzen deshalb ihre Leistung wenig realistisch ein. Feedback scheint also vor allem bei neuartigen Aufgaben sinnvoll, wenn davon ausgegangen werden muss, dass Testanden noch kein spezifisches Fähigkeits-Selbstkonzept ausgebildet haben.

Im nächsten Kapitel werden die Befunde der vorliegenden Untersuchung zusammengefasst und im Kontext der Ergebnisse aus der Literaturanalyse diskutiert.

7 Zusammenfassung und Diskussion der Befunde

Ziel der vorliegenden Arbeit war es, die Bedeutung von außerintellektuellen Personmerkmalen bei der Manifestation der intellektuellen Leistungsfähigkeit in Abhängigkeit von bestimmten situativen Merkmalen zu untersuchen. Damit ordnet sich die Arbeit in das Spannungsfeld von differentieller Persönlichkeitsforschung und Leistungs-, speziell Intelligenzdiagnostik ein. Darüber hinaus wurde versucht, die differentielle Fragestellung experimentell (bzw. quasiexperimentell) unter den Bedingungen der Feldforschung zu bearbeiten.

Im Rahmen der Hauptfragestellung sollte geklärt werden, inwiefern außerintellektuelle Personmerkmale, wie z.B. das Fähigkeits-Selbstkonzept einer Person, die Verarbeitung und Nutzung von performanzbezogenem Feedback moderieren. Hintergrund dieser Fragestellung war das Bestreben, die Mechanismen im Leistungsverhalten in Leistungstests mit Feedback, z.B. Lerntests (Guthke & Wiedl, 1996), näher aufzuklären, sowie Schlussfolgerungen für die Gestaltung effektiver Feedback-Strategien in dynamischen Leistungstests abzuleiten.

Zur Bearbeitung dieser Fragestellung wurden an einer nordenglischen Schülerstichprobe ($N = 249$) im Alter von 13 bis 15 Jahren in einem balancierten messwiederholten Design die Leistungen in feedbackfreien Reasoning-Tests in zwei Materialbereichen – figural und numerisch – den Leistungen in parallelen Reasoning-Tests mit Richtig/Falsch-Feedback gegenübergestellt. Zusätzlich wurden persönlichkeitsdiagnostische Maße zu den Konstrukten Zielorientierung, Begabungskonzept sowie Testangst erfasst. Die Persönlichkeitsdiagnostik erfolgte zum einen mit Hilfe von bewährten Fragebogenverfahren. Zum anderen wurden mit dem Ziel, situationsspezifische Indikatoren für die interessierenden Merkmale in die Analysen einbeziehen zu können, entsprechende diagnostische Zugänge entwickelt und psychometrisch evaluiert.

7.1 Generelle Feedback-Hypothese

Die generelle Feedback-Hypothese geht davon aus, dass performanzbezogenes Feedback im Testverlauf im Allgemeinen zu einer Leistungsverbesserung der Testanden führt verglichen mit feedbackfreien Testbedingungen. Entgegen diesen Erwartungen ließ sich kein genereller Richtig/Falsch-Feedback-Effekt auf die Leistung in Tests zum schlussfolgernden Denken ausmachen. Weder unter einer interindividuellen (Zwischen-Gruppen-Vergleich), noch unter einer intraindividuellen (Inner-Gruppen-Vergleich)

Perspektive wurde ein performanzbezogener Treatmenteffekt festgestellt. Das gilt sowohl für die figurale als auch für die numerische Reasoning-Leistung. Vielmehr ergaben sich Hinweise auf einen schwachen domänenspezifischen negativen Richtig/Falsch-Feedback-Effekt. So zeigten Testanden bei der Bearbeitung der figuralen Reasoning-Aufgaben unter Feedback-Bedingungen Leistungseinbußen, verglichen mit einer feedback-freien Testvorgabe. Die scheinbare Inkonsistenz der Befunde lässt sich jedoch nicht domänenspezifisch interpretieren, sondern ist wohl eher durch das in den beiden Reasoning-Tests unterschiedliche Antwortformat (Multiple-Choice bei Figurenfolgen, freies Antwortformat bei Zahlenfolgen) erklärbar (Abschnitt 7.2.2.2).

Dieser Befund steht mit den in der Literatur referierten Befunden zur Effektivität von Feedback-Interventionen im Einklang. So stellten Kluger und DeNisi (1996) in einer Metaanalyse die große Variabilität von leistungsbezogenen Feedback-Effekten heraus und machten erstmals gezielt auf häufig ignorierte schwache sowie negative Feedback-Effekte aufmerksam (38% der 131 einbezogenen Studien; siehe auch Bangert-Drowns et al., 1991; Kluger & DeNisi, 1998). Carlson und Wiedl (1980, 1992) sowie Carlson (1989) konnten bei Verwendung von Richtig/Falsch-Rückmeldungen in dynamischen Tests (*Testing-the-Limits*) ebenfalls keine Leistungsverbesserungen nachweisen. Auch neuere Studien berichten von ausbleibenden Effekten von itemspezifischem Richtig/Falsch-Feedback im Hinblick auf die Leistung in Intelligenztests (Delgado & Prieto, 2003; Rousseau & McKelvie, 2000; Stankov & Crawford, 1997).

Als ein wesentliches Bestimmungsstück einer effektiven Feedback-Strategie wird von einigen Autoren ein hinreichender *Elaborationsgrad* der rückgemeldeten Information ausgemacht (Kluger & DeNisi, 1996; Kulhavy, 1977; Musch, 1999; unter Berücksichtigung von außerintellektuellen Personmerkmalen siehe auch Johnson, Perlow & Pieper, 1993; Narciss, 2001). Vor dem Hintergrund der Befunde von Kluger und DeNisi (1996) ist das in der vorliegenden Untersuchung verwendete Richtig/Falsch-Feedback wenig effektiv, da es keine fehlerspezifischen Hinweise zur Aufgabenlösung enthält und daher a priori „informationsarm“ ist. Eine andere Ursache könnte darin bestehen, dass Richtig/Falsch-Feedback keine „neue“ Information anbietet und somit für den Testanden redundant ist. Tatsächlich schätzten in der vorliegenden Untersuchung Testanden ihr Leistungsniveau bei Bearbeitung der numerischen Reasoning-Aufgaben auch ohne Feedback angemessen ein (siehe dazu Abschnitt 7.2.2.2). Auch Stankov und Crawford (1997) konnten zeigen, dass die Sicherheit, mit der Testanden ihre Itemlösungen in ei-

nem Intelligenztest (*Raven's Progressive Matrices*) angaben, hoch mit ihrer realen Leistung übereinstimmte. Roos, Wise und Plake (1997) gehen davon aus, dass auf Richtig/Falsch-Feedback verzichtet werden kann, da auch in ihren Untersuchungen Testanden bereits unter feedbackfreien Bedingungen in der Lage waren, ihre Leistungen realistisch zu beurteilen. Sie konnten bei Einsatz eines adaptiven Fähigkeitstests ebenfalls keinen generellen Richtig/Falsch-Feedback-Effekt nachweisen.

Die Befunde legen insgesamt nahe, dass itemspezifisches Richtig/Falsch-Feedback in Intelligenztests nicht generell leistungsförderlich ist (*generelle Feedback-Hypothese*). Zum einen ist die vermittelte Information zu unspezifisch und ermöglicht es dem Testanden deshalb nicht in ausreichendem Maße, fehlerhafte Hypothesen zur Problemlösung zu korrigieren. Zum anderen stellt Richtig/Falsch-Feedback für manche Testanden keine „neue“ Information dar (siehe auch Abschnitt 7.3). Testanden werden einem solchen redundanten Feedback entweder nicht ausreichend Aufmerksamkeit widmen (schon Kulhavy, 1977), oder ihre Aufmerksamkeit auf die Erreichung von hierarchisch höher angeordneten, selbstrelevanten Zielen richten, z.B. auf das Ziel, einen guten Eindruck zu hinterlassen (siehe Feedback-Interventions-Theorie, Kluger & DeNisi, 1996). Vor dem Hintergrund konsistent nicht nachweisbarer genereller Effekte sollte untersucht werden, ob systematische differentielle Effekte im Leistungsverhalten bei Vermittlung von Richtig/Falsch-Feedback zu erwarten sind. Dazu waren vor allem nicht-kognitive Personmerkmale in den Analysen zu berücksichtigen.

7.2 Differentielle Feedback-Hypothese

7.2.1 Intellektuelle Leistungsfähigkeit

Die differentielle Feedback-Hypothese bezieht sich auf die Erwartung, dass die Leistung unter Feedback-Bedingungen neben der Fähigkeit, von Feedback zu profitieren – als Indikator für die intellektuelle Leistungsfähigkeit der Testanden – auch auf das Wirken außerintellektueller Personmerkmale zurückzuführen ist bzw. durch diese moderiert wird. Dazu war im Sinne der Kontrolle des Wilderschen Ausgangswertgesetzes (Wilder, 1931) zunächst zu klären, ob eine Leistungsverbesserung bei erfolgreicher Nutzung von Feedback in Abhängigkeit des Ausgangswertes variiert (*Ausgangswerthypothese*). Die Ergebnisse legen nahe, dass die Leistungsveränderung unter Feedback-Bedingungen in *figuralen* Reasoning-Tests bedeutsam durch die feedbackfreie Leistung bestimmt wird. So profitierten insbesondere leistungsstarke Testanden von Richtig/Falsch-Feedback,

während leistungsschwache Testanden sich unter Feedback eher verschlechterten. Dieser zwar schwache Treatmenteffekt ist praktisch bedeutsam. Immerhin lösten leistungsschwache Testanden unter Feedback-Bedingungen bis zu 2 (von 12) Aufgaben weniger als unter feedbackfreien Testbedingungen.

In dynamischen Tests – speziell in Kurzzeit-Lerntests – werden zunächst routinemäßig Richtig/Falsch-Rückmeldungen nach jedem Item gegeben und erst nach wiederholten Falsch-Antworten elaboriertes Feedback, d.h. Hilfen, präsentiert. Das elaborierte Feedback soll insbesondere leistungsschwächeren Testanden zu höheren Leistungen verhelfen. Von leistungsstarken Testanden wird erwartet, dass sie bereits unter informationsarmen Feedback-Bedingungen, also bei Richtig/Falsch-Feedback, Leistungsgewinne zeigen. Eine solche Feedback-Strategie scheint zunächst sinnvoll, wird aber durch die Befunde der vorliegenden Untersuchung in Frage gestellt. Danach profitieren leistungsstarke Testanden zwar von Richtig/Falsch-Feedback, zumindest bei Verwendung figuraler Reasoning-Aufgaben mit Multiple-Choice-Antwortformat. Leistungsschwache Testanden werden jedoch unter Richtig/Falsch-Feedback-Bedingungen benachteiligt und in ihrer Leistungsfähigkeit eventuell unterschätzt. Eine solche Wirkung von Richtig/Falsch-Feedback widerspricht der Testintention des Dynamischen Testens, sollen hier doch „versteckte“ Potentiale vor allem leistungsschwächerer Testanden aufgedeckt werden. Der leistungsoptimierende Effekt einer dynamischen Testprozedur (Beckmann, 2001; Guthke & Wiedl, 1996), wie z.B. in den genannten Kurzzeit-Lerntests, wird demnach vor allem durch das hier im Bedarfsfall angebotene *elaborierte* Feedback erzielt (siehe auch Abschnitt 7.4). Wird dagegen im Testverlauf lediglich Richtig/Falsch-Feedback angeboten, sind nach Ergebnissen der vorliegenden Studie Benachteiligungen leistungsschwacher Testanden zu erwarten. Das unterschiedliche Leistungsverhalten in Abhängigkeit des Ausgangswertes der Testanden erklärt auch den nur schwachen generellen Treatmenteffekt in der vorliegenden Untersuchung.

7.2.2 Außerintellektuelle Personmerkmale

Zur Untersuchung von differentiellen performanzbezogenen Feedback-Effekten wurden zunächst *Schulleistungen* analysiert. Zum einen eignen sich Schulleistungen, da sie Leistungen unter „natürlichen“ Feedback-Bedingungen darstellen (ökologische Validität). Zum anderen wurden die in dieser Arbeit betrachteten persönlichkeits-theoretischen Konzepte häufig im Schulsetting untersucht (z.B. Studien zur Zielorientierung nach Broome, 2001; Dresel, 2001; Elliott & Dweck, 1988; Görn, Vollmeyer & Rheinberg, 2001;

Ziegler, 2001). So sollte zunächst geprüft werden, ob sich die von Dweck und Leggett (1988) sowie Nicholls (1984) postulierten Zusammenhänge von Zielorientierung, Begabungskonzept und Leistung auch in der vorliegenden Stichprobe mittels der hier verwendeten Instrumente nachweisen lassen. Anschließend wurde geprüft, inwiefern das Modell auch auf Testsituationen mit Richtig/Falsch-Feedback übertragen werden kann.

7.2.2.1 Zielorientierung

Das im Laufe der individuellen Lerngeschichte über die eigene Begabung erworbene Wissen – das Begabungskonzept – drückt sich unter anderem im *Vertrauen* in die eigenen Fähigkeiten aus. Erwartungsgemäß bestimmt das schulbezogene Begabungskonzept, bzw. das Vertrauen in die eigenen schulischen und intellektuellen Fähigkeiten, ob sich Leistungsziele eines Schülers in seiner *Schulleistung* widerspiegeln (*Moderatorhypothese Begabungskonzept*). In Übereinstimmung mit den Modellannahmen nach Dweck (2000; Dweck & Leggett, 1988) und Nicholls (1984, Nicholls et al., 1990) zeigten Schüler in der vorliegenden Arbeit die höchsten Schulleistungen, je seltener sie Leistungsziele verfolgten und je eher sie ihren intellektuellen und schulbezogenen Fähigkeiten vertrauten. Dieser „Schulleistungs-Vorteil“ ließ sich nicht mit einer höheren intellektuellen Leistungsfähigkeit dieser Substichprobe erklären. Die nach Dweck (1986) und Nicholls (1984) definierte Risikogruppe – leistungszielorientierte Schüler mit geringem Begabungskonzept – fiel allerdings nicht durch besondere Leistungseinbußen unter Feedback auf. Auch konnte die von Dweck (1986, Dweck & Leggett, 1988) als auch Nicholls (1984) postulierte protektive Wirkung eines hohen Begabungskonzepts bei Leistungszielorientierung in der vorliegenden Stichprobe bei Vorhersage der Schulleistungen nicht festgestellt werden.

Als eine Ursache für Leistungsunterschiede von Schülern mit unterschiedlicher Ausprägung in den Merkmalen Zielorientierung und Begabungskonzept konnte das *Anstrengungsverhalten* ausgemacht werden, und zwar ebenfalls unabhängig von der intellektuellen Leistungsfähigkeit der Schüler. Theoriekonform (Dweck & Leggett, 1988; Nicholls, 1984) beschrieben sich leistungszielorientierte Schüler als anstrengungsvermeidend, je weniger Selbstvertrauen in ihre schulbezogenen Fähigkeiten sie angaben. Diese Gruppe wurde von Dweck (1986) sowie Nicholls (1984) als Risikogruppe identifiziert, die nach Misserfolgs-Feedback eventuell Leistungseinbußen zeigt. Dagegen schützte Vertrauen in die eigenen Fähigkeiten erwartungsgemäß leistungszielorientierte Schüler davor, leistungshemmende Anstrengungsvermeidungstendenzen auszubilden.

Hier wird die von Dweck (1986) sowie Nicholls (1984) postulierte protektive Wirkung eines hohen Begabungskonzepts deutlich. Entscheidend war jedoch, inwiefern Schüler die berichtete Anstrengungsbereitschaft auch im Schulalltag umsetzten. Nur die Teilgruppe der Schüler, die sich als wenig anstrengungsvermeidend beschrieb, verhielt sich auch im Klassenraum – also für den Lehrer beobachtbar – tatsächlich anstrengungsbereit. Nur diese Schüler erzielten höhere Schulleistungen gegenüber gleich leistungsfähigen Mitschülern. Auf diese Weise lässt sich auch das Ausbleiben der erwarteten Schulleistungsunterschiede *leistungszielorientierter* Schüler mit über- und unterdurchschnittlichem Begabungskonzept in der vorliegenden Untersuchung erklären. Schulleistungen ließen sich zwar anhand von motivationalen Variablen wie Anstrengung vorhersagen, jedoch muss vor allem das beobachtete *Anstrengungsverhalten* in der konkreten Leistungssituation berücksichtigt werden, wenn Leistungsvorhersagen getroffen werden sollen.

Insgesamt schützt ein hohes Begabungskonzept einhergehend mit geringer Leistungszielorientierung Schüler davor, Leistungseinbußen in Situationen mit Erfolgs- oder Misserfolgspotential zu zeigen (siehe auch Broome, 1998; Dweck, 1986, 1988, 2000; Nicholls, 1984; Elliott & Dweck, 1988; bei Untersuchungen von Studenten siehe Stiensmeier-Pelster et al., 1996; Spinath & Stiensmeier-Pelster, 2000). In der vorliegenden Untersuchung traf dies insbesondere für männliche Schüler zu. Als eine Ursache für Leistungsunterschiede in Abhängigkeit der motivationalen Orientierung und des Begabungskonzepts konnte zudem die Anstrengungsregulation ausgemacht werden, nicht dagegen subjektive Theorien der Schüler über Ursachen von Erfolg und Misserfolg. Dies steht im Widerspruch zu Befunden nach Ziegler und Schober (1999; aber schon Dweck & Leggett, 1988), die einen dysfunktionalen Attributionsstil von schulischem Erfolg/Misserfolg eines „vulnerablen“ Schüler-Typs mit unter anderem Leistungszielorientierung und geringem Begabungskonzept im schulischen Setting diskutieren.

Auch bei Vorhersage der experimentell manipulierten *Reasoning-Leistung* (Zahlenfolgen) bestimmte das vor Testbeginn berichtete spezifische Fähigkeits-Selbstkonzept der Testanden – analog zu den Ergebnissen im Schulsetting – inwiefern situationsspezifisch berichtete Leistungsziele die Reasoning-Testleistung unter Feedback-Bedingungen determinierten, und zwar unabhängig von der intellektuellen Leistungsfähigkeit der Testanden (*Moderatorhypothese Begabungskonzept*). Das Selbstvertrauen in die eigenen Fähigkeiten stellte sich dabei generell als leistungsförderlich heraus (siehe auch Elliott &

Dweck, 1988). Interessant war jedoch, dass bei wenig Selbstvertrauen und durchschnittlicher intellektueller Leistungsfähigkeit eine *Wettbewerbsorientierung* (Leistungsziele) die Testleistung ausschließlich in der Feedback-Bedingung günstig beeinflusste. Dieses Ergebnis steht im Widerspruch zu einigen in der Literatur referierten Befunden aus mehr anwendungsorientierten Leistungskontexten (z.B. Elliott & Dweck, 1988; Fodor & Carver, 2000; VandeWalle, Cron & Slocum, 2001). Danach sind in Leistungssituationen mit implementierten Feedback-Schleifen insbesondere *Lernziele* – nicht dagegen Leistungsziele – leistungsförderlich. Elliott und Dweck (1988) konnten zum Beispiel zeigen, dass Leistungsziele verfolgende Fünftklässler mit geringem Begabungskonzept auf Misserfolgsfeedback mit Anzeichen gelernter Hilflosigkeit reagierten.

Neuere Untersuchungen zeigen jedoch, dass nur eine Facette der Leistungszielorientierung, nämlich das Ziel, Fehler bzw. Misserfolg und damit einhergehende negative Beurteilungen zu *vermeiden*, Leistungsdefizite bedingt (Smith, Duda, Allen & Hall, 2002). Elliot und Harackiewicz (1996) sowie Elliot und Church (1997) referieren Befunde, wonach Studenten, die das Ziel verfolgten, positive Bewertungen zu erhalten (*Annäherungs-Leistungsziel*), ebenso hohe Leistungen erreichten wie Studenten, deren Ziel in der Entwicklung ihrer (Aufgaben-)Kompetenz bestand (Lernziel). Dagegen erzielten Studenten, die das Ziel verfolgten negative Bewertungen zu vermeiden (*Vermeidungs-Leistungsziel*), die geringsten Leistungen, und berichteten weniger Interesse und Aufgabefreude. Auch VandeWalle et al. (2001) konnten eine leistungsbeeinträchtigende Wirkung von Leistungszielen nach Feedback im universitären Setting (schriftlich, kommentierte Examensnoten) nur dann ausmachen, wenn Studenten Vermeidungs-Leistungsziele verfolgten, also Fehler bzw. Misserfolg zu vermeiden suchten. Bei Annäherungs-Leistungszielen, also dem Bestreben der Studenten, im Examen besser als andere Studenten abzuschneiden, waren keine Leistungsdefizite beobachtbar. Die von Elliot und Harackiewicz (1996) vorgeschlagene Aufteilung der Leistungszielorientierung in eine *Annäherungs-Leistungszielorientierung* mit bewältigendem Verhalten bei Misserfolg und eine *Vermeidungs-Leistungszielorientierung* mit hilflosem Verhalten bei Misserfolg wird auch durch die in dieser Arbeit vorgestellten Befunde unterstützt. Diese legen nahe, dass in Testsituationen mit performanzbezogenem Feedback Testanden benachteiligt sind, die bereits vor Testbeginn ein geringes Begabungskonzept besitzen („*I do not feel confident that I shall be able to solve most of the problems.*“). Trauen sich Testanden die Bewältigung der Testaufgaben eher nicht zu, kann jedoch eine *wettbewerbsorientierte* motivationale

Grundhaltung („*What is important to me is getting a higher score on the test than everyone else.*“) vor Leistungseinbußen nach Richtig/Falsch-Feedback schützen (Annäherungs-Leistungszielorientierung).

Da sich entsprechende Befunde nicht in einer *feedbackfreien* Testbedingung nachweisen ließen, kann bereits jetzt geschlussfolgert werden, dass außerintellektuelle Personmerkmale Leistungen in Tests mit Richtig/Falsch-Feedback – im Gegensatz zu Leistungen in *feedbackfreien* Tests – wesentlich (mit)bestimmen. Differentielle performanzbezogene Effekte konnten hier für die Merkmale Zielorientierung und Begabungskonzept aufgezeigt werden. Die Effekte sind zwar schwach ($\Delta R^2 = .04$), dennoch praktisch bedeutsam. Immerhin wurden mittlere Leistungsdifferenzen bis zu 2.5 (von 12) Aufgaben beobachtet, die nicht auf ein unterschiedliches intellektuelles Leistungsniveau der Testanden zurückgeführt werden können. Auch auf dem Hintergrund der Befunde zur Determination der *Schulleistungen* lässt sich ableiten, dass in Leistungssituationen mit Erfolgs-/und Misserfolgsfeedback, das Begabungskonzept bzw. das Vertrauen in die eigenen Fähigkeiten und die Zielorientierung der Rezipienten leistungsbestimmend sind, und zwar auch unabhängig von ihrer intellektuellen Leistungsfähigkeit. Letzteres steht im Einklang mit Ergebnissen aus Untersuchungen zur Rolle von Zielorientierungen unter Feedback-Bedingungen in mehr anwendungsorientierten Settings, z.B. im Arbeitssetting (Tuckey et al., 2002), und an der Universität (Fodor & Carver, 2000; VandeWalle et al., 2001; VandeWalle & Cummings, 1997).

Leistungsdefizite in *Tests* mit Feedback, die in Abhängigkeit von der situationspezifischen Leistungszielorientierung sowie des situationsspezifischen Begabungskonzepts der Testanden auftraten, ließen sich anhand der vorliegenden Daten nicht mit einem unterschiedlichen Ausmaß an *Anstrengungsbereitschaft* erklären (siehe z.B. VandeWalle et al., 2001). Lernzielorientierte Testanden mit geringem Begabungskonzept gaben zwar vor Testbeginn eine hohe Anstrengungsbereitschaft an, zeigten aber dennoch relativ geringe Reasoning-Leistungen. Dieser Befund lässt sich eventuell auch als eine „saure Trauben-Reaktion“ selbstkonzeptschwacher Testanden erklären, und zwar im Sinne von: „Vertraue ich meinen Fähigkeiten nicht, die folgenden Testaufgaben zu lösen, so setze ich mir auch nicht das Ziel, einen höheren Testscore als meine Mitschüler zu erreichen. Ich bin dagegen motiviert und anstrengungsbereit, im Testverlauf zu ler-

nen, wie solche Testaufgaben zu lösen sind.“³³ In Übereinstimmung mit den referierten Befunden zu Determinanten der Schulleistung, schützte zudem ein hohes situationsspezifisches Begabungskonzept, d.h. Vertrauen in die eigenen Fähigkeiten, bei Wettbewerbsorientierung vor Anstrengungsvermeidung. Die vor Testbeginn angegebene Anstrengungsbereitschaft stand jedoch in keinem Zusammenhang zur Testleistung der Testanden. Eine Fremdeinschätzung der Anstrengung (wie bei Einschätzung der schulbezogenen Anstrengungsvermeidung), die wohlmöglich eine realistischere Abbildung des Anstrengungsverhaltens im Test erlaubt hätte, lag jedoch für die situationsspezifische Anstrengungsbereitschaft nicht vor.

Die referierten Befunde zur Wechselwirkung von Begabungskonzept und Leistungszielorientierung bei Vorhersage der Feedback-Testleistung konnten allerdings bei Analyse der *figuralen Reasoning-Testleistung* nicht repliziert werden. Zur Erklärung des unterschiedlichen Befundmusters in den Materialbereichen sind inhaltliche sowie formale Gründe anzuführen. So kann die geringe Reliabilität der figuralen Reasoning-Tests geringe bis fehlende Korrelationen bewirkt haben (z.B. deutlich geringere Varianzaufklärung im Regressionsmodell zur Vorhersage der figuralen versus der numerischen Reasoning-Leistung). Möglicherweise besitzen Testanden aber zu figuralen Reasoning-Aufgaben ein wenig spezifisches Fähigkeitsselbstbild, da sie im Allgemeinen solcher Art Aufgaben seltener ausgesetzt sind, als Aufgaben, die den Umgang mit Zahlen erfordern. Zudem waren die Figurenfolgen für die hier untersuchte Stichprobe von relativ geringer Schwierigkeit und haben eventuell, auch im Vergleich zu den Zahlenfolgen, wenig Misserfolgserleben induziert. Außerdem kann vermutet werden, dass computergestützte Aufgaben mit Multiple-Choice-Antwortformat (Figurenfolgen) im Gegensatz zu Aufgaben mit freiem Antwortformat (Zahlenfolgen) von vornherein als wenig selbstwertbedrohlich wahrgenommen werden, da bei Misserfolg eine Ratestrategie durch einfaches Anklicken einer Antwortalternative mit weniger psychischem Aufwand verbunden ist, als das Eintippen einer geratenen Antwort bei freiem Antwortformat. Hier muss auch angemerkt werden, dass die Ratemöglichkeit bei Multiple-Choice-Antwortformat die Reliabilität der Tests beeinträchtigt haben kann, da der Testscore hier nicht nur die zu messende Fähigkeit sondern auch das „Rateverhalten“ abbildet. In-

³³ Diese Testanden stimmten den folgenden Items zu: „I do not feel confident that I shall be able to solve most of the problems.“; „What is important to me is learning how to solve the problems.“; „I will feel really successful if I can solve the problems once I have worked really hard.“

samt war das Begabungskonzept der Testanden bei der Bearbeitung der Figurenfolgen in der hier untersuchten Stichprobe von geringer Relevanz und hatte somit auch keine Moderatorfunktion für den Zusammenhang von Zielorientierung und Leistung inne.

Es kann festgehalten werden: Werden vor dem Hintergrund eines geringen Begabungskonzepts *Leistungsziele* statt Lernziele verfolgt, wirkt sich Richtig/Falsch-Feedback positiv auf das Leistungsverhalten in Reasoning-Tests aus. Entgegen den in der Literatur berichteten Befunden (Dweck, 1986, 2000; Nicholls, 1984) und den pädagogischen Erwartungen garantiert eine Lernzielorientierung – also die Motivation, sein Wissen und Fähigkeiten zu erweitern – keinen Vorteil hinsichtlich der produktiven Verarbeitung von Feedback. Vielmehr scheint in Testsituationen mit Feedback eine eher wettbewerbsorientierte Leistungszielorientierung, also die Motivation, positive Bewertungen zu erhalten und deshalb in Leistungssituationen vor allem hohe Fähigkeiten und Wissen zu demonstrieren (*Annäherungs-Leistungsziel*), von Vorteil zu sein, insbesondere dann wenn Testanden wenig auf ihre eigenen Fähigkeiten vertrauen.

Hierbei ist zu bedenken, dass Testsituationen und speziell Intelligenzmessungen, letztere auch aufgrund der anzunehmenden hohen persönlichen Bedeutsamkeit der Eigenschaft Intelligenz, a priori *Wettbewerbscharakter* innehaben, eine Lernzielorientierung also in diesem Kontext wohl eher unangemessen ist. Da Zielorientierungen auch situativ ausgelöst werden können (Dweck & Leggett, 1988; Nicholls, 1984; siehe auch Butler, 1987; Stiensmeier-Pelster et al., 1996; Köller & Schiefele, 2001), ist anzunehmen, dass im vorgestellten Experiment eine Leistungszielorientierung induziert wurde. Zudem fördert auch informationsarmes Feedback, wie das hier verwendete Richtig/Falsch-Feedback, eine Leistungszielorientierung der Testanden (auch als „leistungsorientiertes“ Feedback bezeichnet, siehe Johnson et al., 1993). Aufgrund einer besseren Passung zwischen Personmerkmal und Situationscharakteristik, wie z.B. im Rahmen der Theorie des *Mental Self-Governments* im pädagogischen Kontext (Sternberg, 1988, siehe auch Grigorenko & Sternberg, 1995), oder im Rahmen des *Represser-Sensitizer*-Modells im klinisch-psychologischen Kontext diskutiert (Krohne, 1996), liegt es nahe anzunehmen, dass Personen mit habitueller Leistungszielorientierung oder Personen, die sich in ihrer motivationalen Orientierung an den Leistungszielcharakter der Testsituation anpassen konnten, unabhängig von ihrem Selbstvertrauen in der Lage waren, von Feedback zu profitieren. Diese Interpretation steht auch in Übereinstimmung mit Befunden nach Harackiewicz und Elliot (1993). Sie gehen davon aus, dass Kompetenzinformationen

(Feedback) während der Aufgabenbearbeitung intrinsische Motivation und Kompetenzerleben auslösen bzw. aufrechterhalten, wenn sie mit den Zielorientierungen der Rezipienten kongruent sind (auch Harackiewicz & Sansone, 1991). Leistungszielorientierte Personen präferieren Rückmeldungen, die sie über ihr Leistungsniveau im Vergleich zu einer Norm informieren, während sich lernzielorientierte Personen eher für ihren persönlichen Lernfortschritt interessieren. Ein maximaler Feedback-Gewinn ist also zu erwarten, wenn die Situationscharakteristik nicht im Widerspruch zur Zielorientierung der Testanden steht. Einschränkend ist zu bemerken, dass sich das Leistungsverhalten von Personen mit hohem Begabungskonzept und Leistungszielorientierung unter Feedback nicht änderte.

Das in dieser Studie realisierte Design erlaubt den Vergleich der zielorientierungsbezogenen Befundmuster in verschiedenen Domänen (numerische Reasoning-Leistung und figurale Reasoning-Leistung) aber auch unter verschiedenen modellbezogenen *Operationalisierungen* (Dwecks eindimensionales, bipolares Modell versus Nicholls zweidimensionales Modell der Zielorientierung). Beide Zielkonzeptionen und die damit einhergehende diagnostische Umsetzung, nämlich entweder die vom Testanden geforderte Entscheidung für eine der beiden Zielkategorien (Dweck, 2000) oder die Freiheit des Testanden beide Ziele anzugeben (Duda & Nicholls, 1992), erwiesen sich als gleichermaßen geeignet für eine *situationsspezifische* Leistungsvorhersage. Dies traf allerdings nicht für die Messung der *schulbezogenen* Zielorientierung zu, da die große Mehrheit der Schüler angab, im Schulsetting Lernziele zu verfolgen. Ein Grund dafür mag die soziale Erwünschtheit von Lernzielen im Schulalltag sein. Ein zweidimensionales Operationalisierungsmodell also die getrennte Erfassung der Lern- und Leistungsziele ergab aber, dass Schüler durchaus auch Leistungsziele verfolgen. Die Betrachtung der Leistungsziel-Dimension unabhängig von Lernzielen war somit für die Analyse des Zusammenhangs von Begabungskonzept, Zielorientierung und Leistung weitaus Gewinn bringender. Selbst wenn die Zielorientierung ein bipolares Konstrukt darstellt, ließe die von Nicholls vorgeschlagene zweidimensionale Operationalisierung ein dementsprechendes Antwortverhalten zu. Im Falle einer zu berücksichtigenden Mehrdimensionalität des Konstrukts würde die Dwecksche Operationalisierung zu einer invaliden Abbildung dieses Konstrukts führen. Vor diesem Hintergrund ist der Nichollschen Operationalisierung insgesamt der Vorzug zu geben.

7.2.2.2 Testangst

In der vorliegenden Untersuchung spiegelte sich erwartungskonform die im *Schulalltag* berichtete Prüfungsangst als auch das schulbezogene Fähigkeitsselbstbild in den schulbezogenen Leistungsvariablen wider. Prüfungsangst konnte jedoch nicht eindeutig als eine von mehreren möglichen Ursachen für Schulleistungsdefizite von selbstkonzeptschwachen Schülern ausgemacht werden (*Mediatorhypothese Testangst*). Vertrauen in die eigenen schulischen und intellektuellen Fähigkeiten ging tendenziell mit hohen Leistungen, sowohl feedbackbedingten Schulleistungen als auch feedbackfreien Intelligenztest-Leistungen einher. Prüfungsangst trat dagegen insbesondere mit geringen feedbackbedingten Schulleistungen auf.

Die unter Feedback-Bedingungen erbrachte *Reasoning-Leistung* ließ sich erwartungsgemäß nicht mit dem situationsunspezifischen *schulbezogenen* Fähigkeits-Selbstkonzept der Testanden erklären. Bei Analyse der entsprechenden *situationsspezifischen* Begabungskonzept-Maße stellte sich jedoch heraus, dass das spezifische Begabungskonzept, d.h. das vor Testbeginn angegebene Vertrauen in die eigenen Fähigkeiten die Testaufgaben zu lösen, auch in einer Testsituation mit Feedback wesentlich die Leistung determinierte. Das situationsspezifische Begabungskonzept ist demnach sowohl unter Feedback als auch unter feedbackfreien Testbedingungen von hoher Relevanz für eine erfolgreiche Testbearbeitung.

Hier soll keine unidirektionale Ursache-Wirkungs-Kette postuliert werden. Vielmehr determiniert das Begabungskonzept als Ergebnis früherer Lernerfahrungen wesentlich den Erfolg in der aktuellen Leistungssituation. Die in der vorliegenden Untersuchung vorgenommenen Analysen der Effekte auf das Leistungsverhalten wurden aber teilweise unter Kontrolle der interindividuell variierenden Leistungsvoraussetzungen vorgenommen. Deshalb lässt sich mindestens vermuten, dass das Begabungskonzept eines Schülers auch unabhängig von seiner tatsächlichen Leistungsfähigkeit das Verhalten in unterschiedlichen Leistungssituationen (mit)bestimmt (siehe auch Meyer, 1984).

Wenig Vertrauen in die eigenen Fähigkeiten einhergehend mit Testangst beeinträchtigte die Reasoning-Testleistung, unabhängig von der intellektuellen Leistungsfähigkeit der Testanden. Immerhin elf Prozent der Leistungsvarianz unter Feedback-Bedingungen ließen sich auf die vor Testbeginn berichtete Testangst zurückführen. Zudem konnte Testangst als eine entscheidende Ursache für Leistungsdefizite wenig selbstkon-

zeptstarker Testanden herausgestellt werden, und zwar nur unter Feedback-Bedingungen (*Mediatorhypothese Testangst*). Feedback verhinderte darüber hinaus bei Wiederholungsmessung eine Gewöhnung testängstlicher Testanden an die Testsituation.

Leistungseinbußen in der Feedback-Bedingung ließen sich vor allem auf die *Worry-Komponente* der Testangst zurückführen. In der vorliegenden Untersuchung zeigte sich der Erfolg/Misserfolg unter Feedback-Bedingungen im Ausmaß der im Testverlauf entwickelten Worry-Kognitionen. Insbesondere bei Ersttestung löste Feedback Worry-Prozesse aus und führte vermutlich deshalb zu Leistungseinbußen testängstlicher Testanden. Worry-Kognitionen wirken in einem Feedback-Test leistungshemmend, da sie Ressourcen binden, die dann für die Feedback- und Aufgabenverarbeitung nicht mehr zur Verfügung stehen. Ein hohes Ausmaß an *Emotionality* beeinträchtigte bei Wiederholungstestung ebenfalls die Feedback-Leistung. Testangst vor Testbeginn (Worry-Komponente und *Emotionality-Komponente*) stellte sich zwar auch in einer feedbackfreien Testung als leistungshemmend heraus, jedoch nur im Rahmen der Erstkonfrontation. Werden die *trait-näheren* Maße zur *schulbezogenen* Testangst betrachtet, so ging Testangst wiederum lediglich bei einer Ersttestung und zwar speziell unter Feedback mit geringen Leistungen einher. Diese Befunde stützen die Annahme, dass die Test- oder Prüfungsangst insbesondere unter Feedback-Bedingungen eine wesentliche negative Determinante des Leistungsverhaltens darstellt.

Meijer (1993, 2001) zeigt dagegen, dass Testergebnisse in dynamischen Testverfahren verglichen mit feedbackfreien Statustests weniger durch den Faktor Testangst verfälscht werden (siehe auch Wiedl, Bethge & Bethge, 1982). Er konnte sogar nachweisen, dass mit Hilfe dynamischer Testscores validere Prognosen für testängstliche Testanden gestellt werden als mit traditionellen Intelligenztest-Maßen. Diese mit den Ergebnissen der vorliegenden Studie scheinbar im Widerspruch stehenden Befunde lassen sich durch das im Meijerschen computergestützten Kurzzeit-Lerntest (1993, 2001) implementierte Feedback erklären. So erhielten Testanden nach Falsch-Antworten Denkhilfen, also elaboriertes und somit informatives Feedback. Im Gegensatz zu Richtig/Falsch-Feedback ist elaboriertes Feedback im Test offensichtlich geeignet, Leistungsdefizite testängstlicher Testanden zu kompensieren.

Insgesamt ergaben sich in der vorliegenden Studie Hinweise darauf, dass *testängstliche* Testanden in einem Feedback-Test hinsichtlich ihrer „wahren“ Leistungsfähigkeit unterschätzt werden. Ausgeprägtes Vertrauen in die eigene Leistungsfähigkeit

ist in einer Feedback- als auch feedbackfreien Testbedingung von Vorteil. Unter Feedback-Bedingungen sind Leistungseinbußen *selbstkonzeptschwacher* Testanden vor allem auf deren ausgeprägte Testangst zurückzuführen. Auch *Anstrengungsvermeidung* geht mit eher geringen Leistungen in Tests mit Feedback einher. Letztgenannter Befund wird auch von Nagler (1995) berichtet, wonach Anstrengungsvermeidungsverhalten mit Leistungen in computergestützten Kurzzeit-Lerntests negativ korreliert war.

Wie bereits diskutiert, ergaben sich unterschiedliche Befundmuster für die beiden betrachteten Materialbereiche *figurales Reasoning* und *numerisches Reasoning*. Weder die schulbezogene oder situationsspezifische *Testangst* noch das schulbezogene oder situationsspezifische *Begabungskonzept* stellten hinreichende Prädiktoren für die *figurale Reasoning*-Leistung unter Feedback-Bedingungen dar. Stattdessen fanden sich Hinweise darauf, dass die feedbackfreien *figuralen Paralleltests* zum zweiten Messzeitpunkt mit wenig Motivation bearbeitet wurden. Dies lässt sich möglicherweise auf die relativ geringe Schwierigkeit der *figuralen Reasoning*-Items zurückführen. Neben der geringen Schwierigkeit der *figuralen Reasoning*-Items mag der relativ geringe Vertrautheitsgrad mit entsprechenden Items und ein damit einhergehendes kaum ausgeprägtes spezifisches *Fähigkeits-Selbstkonzept* dazu geführt haben, dass die *figuralen* Aufgaben von den Testanden als nicht selbstwertbedrohlich bzw. potentiell angstausslösend wahrgenommen wurden (siehe auch *Multiple-Choice-Antwortformat*). Folglich führte wenig Selbstvertrauen in die eigenen Fähigkeiten auch nicht vermehrt zu Testangst.

7.3 Zum Informationsgehalt von Feedback

Als eine wesentliche Ursache für fehlende oder schwache Effekte von Feedback-Interventionen diskutieren einige Autoren die Redundanz der Feedback-Informationen für den Rezipienten (siehe Abschnitt 7.1; z.B. Kluger & DeNisi, 1998; Bangert-Drowns et al., 1991; Kulhavy, 1977). Vermittelt Richtig/Falsch-Feedback nun „neue“ Information? In der vorliegenden Untersuchung ließ sich diese Frage nur domänenspezifisch beantworten. So schätzten Testanden ihre *numerische Reasoning*-Leistung auch ohne Feedback, schon in der ersten Testsitzung, und zwar bereits vor Testbeginn realistisch ein. Richtig/Falsch-Feedback im Testverlauf bot hier keinen Zugewinn an Information. Testanden, die die *figuralen Reasoning*-Aufgaben bearbeiteten, stellten jedoch keine realistischen Erwartungen an ihre Test-Leistung. Auch nach der Testung berichtete Leistungs-

einschätzungen stimmten nur bedingt mit der tatsächlichen figuralen Leistung überein. Hier fungierte Richtig/Falsch-Feedback zumindest potentiell als *zusätzliche* Information.

Die beobachteten geringeren Zusammenhänge zwischen schulbezogenem und situationsspezifischem Begabungskonzept in der Feedback-Bedingung beim Zahlenfolgentest zeigen zudem, dass mittels Feedback-Interventionen durchaus Einfluss auf die Selbsteinschätzung der Testanden genommen werden kann. Auch Testanden, die die figuralen Reasoning-Aufgaben bearbeiteten, veränderten ihre Leistungseinschätzungen unter Feedback. Sie bewerteten ihr Leistungsniveau nach der Feedback-Testung realistischer als nach einer feedbackfreien Testung. Stankov und Crawford (1997) sowie Kleitman und Stankov (2001) berichten ebenfalls Befunde, wonach Richtig/Falsch-Feedback in Intelligenztests eine realistischere Einschätzung der eigenen Leistung unterstützt. Bei der Konstruktion von Tests mit Feedback-Angebot sollte deshalb berücksichtigt werden, ob Testanden bereits ein spezifisches Fähigkeits-Selbstkonzept zum verwendeten Aufgabentypus ausgebildet haben. Handelt es sich um unbekannte, im Alltag nicht repräsentierte Anforderungen, wie z.B. das Lösen figuraler Matrizen, kann Richtig/Falsch-Feedback zusätzliche Information bieten und wird insbesondere von leistungsfähigen Testanden im Sinne eines Finetunings ihres Leistungsverhaltens genutzt. Es besteht aber die Gefahr, dass „informationsarmes“ Richtig/Falsch-Feedback unerwünschte differentielle Feedback-Effekte hervorruft. Insbesondere testängstliche Testanden, aber auch Testanden mit wenig Vertrauen in ihre Fähigkeiten werden unter solchen Feedback-Bedingungen behindert, so dass die unter diesen Umständen gezeigte Leistung keinen adäquaten Indikator ihrer tatsächlichen Leistungsfähigkeit darstellt.

Entscheidend ist also, ob Feedback dem Testanden *aufgabenbezogene* Information vermittelt. Richtig/Falsch-Feedback enthält keine *aufgabenbezogene* Information, wenn Testanden mit dem Aufgabentyp bereits vertraut sind und somit ihre Fähigkeiten zur Bewältigung solcherart Aufgaben auch ohne Feedback realistisch einschätzen können. Effektives Feedback sollte deshalb vor allem über geeignete Lösungsstrategien informieren. Feedback, das keine Hinweise zur korrekten Aufgabenlösung enthält, kann dennoch vom Testanden als *personbezogene* Information verarbeitet werden, z.B. löst „Falsch“-Feedback unter Umständen Misserfolgserleben aus. Inwiefern Testanden Feedback vordergründig als *personbezogene* Information wahrnehmen und verarbeiten oder als *aufgabenbezogene* Information interpretieren, sollte durch ausgewählte Personmerkmale bestimmt sein. Richtig/Falsch-Feedback stellt also sehr wohl Information dar,

die jedoch unterschiedlich wahrgenommen, interpretiert und genutzt wird. So kann Richtig/Falsch-Feedback entweder im Sinne der Aufgabe verarbeitet werden oder in Bezug auf die eigene Person. Nur eine aufgabenbezogene Verarbeitung ermöglicht Leistungsgewinne unter Feedback-Bedingungen. Bei einer personbezogenen Verarbeitungsweise wirkt insbesondere „Falsch“-Feedback leistungshinderlich. Wie Feedback interpretiert und genutzt wird, hängt von außerintellektuellen Personmerkmalen, wie z.B. dem Fähigkeits-Selbstkonzept, ab.

So ist bekannt, dass Feedback-Interventionen potentiell selbstwertbedrohlich wirken (z.B. Kluger & DeNisi, 1996; Kluger & Adler, 1993). Bisher weniger beachtete Studien aus den 70er und frühen 80er Jahren (Shrauger & Rosenberg, 1970; Stake, 1982; auch Meyer & Starke, 1981/1982) liefern Hinweise, dass das *Selbstwertgefühl* als ein Moderator für leistungsbezogene Feedback-Effekte gesehen werden kann. Im Rahmen der Stressforschung (Jerusalem, 1991, 1992) unternommene Untersuchungen haben weiterhin gezeigt, dass ein hohes *Selbstkonzept* oder hohe *Selbstwirksamkeit* das Bedrohungsleben in Leistungssituationen mit Feedback verringert und ein geringes Selbstkonzept das Bedrohungsleben verstärkt. Die Generalisierbarkeit dieser Befunde einschränkend muss beachtet werden, dass in diesen Studien nicht-kontingentes (auch *Bogus-*) Feedback eingesetzt wurde. Studien zur Konstruktion adaptiver Tests mit Feedback ergaben, dass die subjektive Wahrnehmung der eigenen Leistung im Testverlauf – das Fähigkeits-Selbstkonzept – das Erleben und Verarbeiten des Tests bestimmt und mit anderen nicht-intellektuellen Reaktionen auf den Test zusammenhängt, wie z.B. Testzufriedenheit und Aufgabenmotivation (Tonidandel, Quinones & Adams, 2002). Dem Selbstkonzept als auch der emotionalen Bewertung des Selbstkonzepts, also dem Selbstwertgefühl, kommt somit eine Schlüsselfunktion in Leistungssituationen mit Feedback zu.

In der vorliegenden Arbeit konnte das Fähigkeits-Selbstkonzept einer Person, bzw. das damit einhergehende Vertrauen in die eigenen Fähigkeiten, als ein leistungsdeterminierendes Personmerkmal in zwei verschiedenen Leistungssituationen mit Feedback ausgemacht werden. Diese sind Testsituationen als auch schulische Leistungssituationen. Die Kompetenzerwartung einer Person in einer spezifischen Leistungssituation entscheidet darüber, ob sie das angebotene Feedback als zusätzliche *aufgabenbezogene* Information wahrnimmt und nutzt. Wenig Vertrauen in die eigenen Fähigkeiten kann eine *personbezogene* Verarbeitung von Feedback begünstigen und somit Testangst

auslösen oder in Abhängigkeit der verfolgten Ziele anstrengungsvermeidendes Verhalten fördern. Ein Feedback-Gewinn ist unter diesen Bedingungen nicht zu erwarten.

Die vorliegende Untersuchung hat im Rahmen einer Intelligenzmessung mit Feedback insgesamt gezeigt, dass die Konfrontation der Testperson mit informationsarmem Feedback zu einem vertrauten Aufgabentyp nicht effektiv ist. Besitzt ein Testand bereits ein ausgeprägtes spezifisches Fähigkeitsselbstbild kann er auch ohne Richtig/Falsch-Feedback seine Leistung angemessen einschätzen. Da Richtig/Falsch-Feedback keine Strategien zur Lösung des Aufgabentyps vermittelt, bietet es für den Rezipienten keine *zusätzliche* Information. Gleichzeitig fördert „Falsch“-Feedback aber potentiell Misserfolgserleben im Testverlauf. Problematisch ist, dass Testanden unter solchen informationsarmen Feedback-Bedingungen Leistungseinbußen aufweisen, die nicht auf mangelnde kognitive Fähigkeiten zurückgeführt werden dürfen, sondern aufgrund außerintellektueller Personmerkmalen zustande kommen. Welche Prozesse können hier angenommen werden?

Nach der Feedback-Interventions-Theorie (Kluger & DeNisi, 1996) lenkt informationsarmes Richtig/Falsch-Feedback die Aufmerksamkeit des Testanden auf das Selbst bzw. auf selbstrelevante Ziele, anstatt auf die Lösung der konkreten Aufgabe. Dieser Wechsel des Aufmerksamkeitsfokus kann unter anderem dazu führen, dass der Testand

- versucht die erlebte Ist-Soll-Diskrepanz zu verringern (z.B. durch Abwertung des angestrebten Ziels) und infolgedessen wenig Anstrengungsbereitschaft zeigt,
- die Aufmerksamkeit auf das Selbst richtet (z.B. vermehrt Worry-Kognitionen ausbildet),
- affektive Reaktionen zeigt (z.B. bei Nichterreicherung eines selbstrelevanten Ziels negative Stimmungslagen und hohe Aktiviertheit ausbildet), und somit
- insgesamt nur eingeschränkt kognitive Ressourcen in die Aufgabenbewältigung investiert.

In der vorliegenden Untersuchung wurden ausgewählte Annahmen der Feedback-Interventions-Theorie spezifiziert und empirisch geprüft. Ein Ausbleiben differentieller Effekte der hier betrachteten außerintellektuellen Personmerkmale auf das Leistungsverhalten nach Feedback stünde im Widerspruch zur Feedback-Interventions-Theorie. Die Ergebnisse zeigen aber theoriekonform, dass außerintellektuelle Personmerkmale Leistungen unter Feedback-Bedingungen im Vergleich zu Leistungen unter

feedbackfreien Bedingungen moderieren. Das (Fähigkeits-) *Selbstkonzept* als auch die *Zielstrukturen*, die eine Person in Leistungssituationen mit Feedback besitzt, bestimmen wesentlich ihren Feedback-Profit, auch unabhängig von ihrer Leistungsfähigkeit. Die Anstrengungsregulation stellt eine Ursache für Leistungsdefizite nach Feedback dar. Die Befunde sind hier allerdings nicht ganz eindeutig. Zudem konnte Testangst, insbesondere die Worry-Komponente der Testangst, als wesentliche Ursache für Leistungseinbußen unter Feedback-Bedingungen, speziell wenig selbstkonzeptstarker Personen, ausgemacht werden. Die Befunde der vorliegenden Untersuchung untermauern die Feedback-Interventions-Theorie nach Kluger und DeNisi (1996) empirisch. Die in der Literatur insbesondere ab den 90er Jahren berichteten und auch in der vorliegenden Studie gefundenen schwachen bis fehlenden oder sogar negativen Feedback-Effekte lassen sich demnach zumindest teilweise auf differentielle Ursachen zurückführen.

In der hier vorgestellten Studie wurde *kontingentes*, d.h. das reale Leistungsniveau der Testanden widerspiegelndes Feedback gegeben. Zahlreiche Studien beschäftigten sich bisher mit der Wirkung von nicht-kontingenter, häufig Misserfolg induzierendem Feedback. Auch hier wurden Feedback-Effekte auf die Leistung, einschließlich damit einhergehender differentieller Effekte untersucht (z.B. so genannte Hilflosgkeitsexperimente, z.B. Brunstein, 1989; Helton, Dember, Warm & Matthews, 1999; Kirchhoff & Stiensmeier-Pelster, 1999; Lauterbach, Stecker & Partl, 1988; Schmalt, 1994; Winefield & Rourke, 1991). Ein Vorteil der Verwendung nicht-kontingenter Feedbacks zur Analyse differentieller Feedback-Effekte besteht in der Möglichkeit, allen Rezipienten ein kontrolliertes Verhältnis von positivem und negativem Feedback anzubieten. Wird wie in der vorliegenden Untersuchung kontingentes Feedback verwendet, so erhalten leistungsstarke Rezipienten vor allem positives Feedback, während leistungsschwache Rezipienten überwiegend negatives Feedback erhalten. Unterschiede im Umgang mit positivem versus negativem Feedback sind anzunehmen und werden auch in der Literatur referiert (z.B. schon Latta, 1978; Shrauger & Rosenberg, 1970). So führen Harris et al. (1993) Befunde auf, wonach Feedback-Effekte auf die Zufriedenheit und das Kompetenzerleben von Studenten von deren Leistungsfähigkeit abhängen. Kontingentes Feedback lässt jedoch eine höhere Glaubwürdigkeit bzw. Akzeptanz des Feedbacks erwarten und repräsentiert somit eine ökologisch validere Testsituation. So beinhalten die in der vorliegenden Studie interessierenden Leistungstests auch ausschließlich leistungs-kon-

tingentes Feedback. Nicht-kontingentes Feedback stellt zusätzliche irreführende, bestenfalls nicht glaubwürdige Information für den Testanden dar. Die Verallgemeinerbarkeit der damit aufgedeckten Befunde ist somit begrenzt. Mit anderen Worten, es muss die Frage gestellt werden, wie relevant die Information darüber sein kann, wie eine erfolgsverwöhnte Person mit einer unerwartet hohen Rate an negativem Feedback umgeht.

Abschließend soll nochmals darauf hingewiesen werden, dass sich verschiedene berichtete Befunde, z.B. der Mediatoreffekt der Variablen Testangst bezogen auf den Zusammenhang von Begabungskonzept und Leistung, mit den traditionell erhobenen *trait-näheren* Persönlichkeitsmaßen nicht nachweisen ließen. Gleiches gilt für den berichteten Moderatoreffekt der Variablen Begabungskonzept in Bezug auf den Zusammenhang von Zielorientierung und Leistung. Hierin zeigt sich der konzeptionelle Vorteil einer *situationsspezifischen* Ausrichtung der Persönlichkeitsdiagnostik. Dieses Vorgehen erlaubt für die konkrete Leistungssituation validere Messungen, welche situationspezifische Effekte aufdecken helfen und eine Fehlinterpretation ausbleibender Effekte verhindern (siehe z.B. die geringe Bedeutung des *trait-näheren* schulbezogenen Begabungskonzepts im Vergleich zum situationspezifischen Begabungskonzept bei Vorhersage der numerischen Reasoning-Leistung in der Feedback-Bedingung im Vergleich zur feedbackfreien Testbedingung).

7.4 Implikationen für das Dynamische Testen

Die Wirkung von (Richtig/Falsch-)Feedback ist unter anderem durch das Selbstwertgefühl bzw. das zugrunde liegende Selbstkonzept einer Person bestimmt. Bei einem niedrigen Begabungskonzept wird der Rezipient Feedback als *Selbstwertbedrohung* wahrnehmen und eventuell nicht als zusätzliches auf die Aufgabe bezogenes Informationsangebot nutzen. Ist der Testand mit dem Aufgabentyp jedoch bereits vertraut, so beinhaltet Richtig/Falsch-Feedback a priori keine *zusätzliche* aufgabenrelevante Information. Dennoch löst „Falsch“-Feedback potentiell Misserfolgserleben aus. Außerintellektuelle Personmerkmale bestimmen demnach sehr wohl die Leistung in Intelligenztests mit Richtig/Falsch-Feedback.

In der vorliegenden Untersuchung konnten *Zielstrukturen* im Zusammenhang mit dem *Begabungskonzept* des Testanden als relevante außerintellektuelle Ursachen für Erfolg und Misserfolg in Leistungssituationen mit Feedback ausgemacht werden. Das trifft speziell für Richtig/Falsch-Feedback in Reasoning-Tests zu, aber auch allgemeiner für

verschiedene Feedback-Arten im Schulsetting. Leistungsdefizite wenig wettbewerbsorientierter Testanden mit geringem Begabungskonzept bzw. Leistungsvorteile wettbewerbsorientierter Schüler bei der Bearbeitung von Tests mit Feedback ließen sich nicht mit einer unterschiedlichen intellektuellen Leistungsfähigkeit erklären. Entscheidend ist jedoch, dass diese Leistungsunterschiede nicht in einer feedbackfreien Test-Bedingung auftraten. Eine solche Gegenüberstellung einer feedback- mit einer feedbackfreien Bedingung bei Analyse des Zusammenhangs von Persönlichkeitsmerkmalen und Leistung wurde meines Erachtens weder in der bisherigen Feedback-Forschung noch in der Forschung zum Dynamischen Testen systematisch unternommen (siehe lediglich Carlson & Wiedl, 1976).

Testangst und Begabungskonzept spielen sowohl in einer feedbackfreien als auch in einer Feedback-Bedingung bei Vorhersage der Leistung eine Rolle (siehe auch Meijer, 2001). Es lässt sich aber genauer zeigen, dass selbstkonzeptschwache Testanden unter anderem deshalb Leistungsdefizite zeigen, weil sie unter Feedback-Bedingungen testängstlich sind. Zum anderen ergab sich, dass das Begabungskonzept der Testanden in Abhängigkeit von motivationalen Zielstrukturen die Testleistungen determinierte. Diese Effekte sind spezifisch für die Testbedingung mit Feedback und treten unter feedbackfreien Bedingungen nicht auf.

Auf Basis dieser Befundlage muss zunächst davon ausgegangen werden, dass der in dynamischen Testverfahren erfasste Feedback-Profit keineswegs nur einen Indikator für die intellektuelle Leistungsfähigkeit des Testanden darstellt. Die hier eingesetzten Tests unterscheiden sich allerdings in den verwendeten Feedback-Strategien von traditionellen dynamischen Tests. Es wurden ausschließlich Rückmeldungen bezogen auf die Lösungsgüte und keine zusätzliche Informationen zu geeigneten Lösungsstrategien, im Sinne von Denkhilfen, gegeben. Insofern deuten die vorliegenden Befunde darauf hin, dass Lerntests unter anderem deshalb valider als herkömmliche Intelligenztests die intellektuelle Leistungsfähigkeit bzw. das intellektuelle Veränderungspotential erfassen (Beckmann, 2001), weil nicht nur Informationen über die Richtigkeit der Itemlösung angeboten werden, sondern vor allem auch fehlerorientierte Denkhilfen, d.h. elaboriertes oder „lernorientiertes“ (Johnson et al., 1993) Feedback. Dieses wird von „lernfähigen“ Testanden genutzt, was sich in Leistungsgewinnen im Testverlauf äußert. Auch innerhalb der Feedback-Forschung konnte gezeigt werden, dass Feedback dann zu Leistungsverbesserungen führt, wenn es mit Informationen zur Lösung der Aufgaben ange-

reichert ist und nicht nur über die Richtigkeit der Lösung informiert (Bangert-Drowns, et al., 1991). Einfaches Richtig/Falsch-Feedback, ohne Hinweise auf die Art des unterlaufenen Fehlers und ohne Aufzeigen von Lösungswegen, scheint dagegen eher ungewollte differentielle Effekte auszulösen und bestimmte Personengruppen zu benachteiligen. Dies sind z.B. Leistungsschwache, aber auch testängstliche Testanden, die in einer feedbackfreien Bedingung bessere Leistungen erreichen als unter Feedback. Wettbewerbsorientierte Testanden zeigen dagegen Leistungsgewinne in Tests mit Feedback.

Mit den hier referierten ersten Befunden zu differentiellen Effekten von außertellektuellen Merkmalen auf die Verarbeitung von Feedback in Reasoning-Tests wird die Validität eines leistungsdiagnostischen Verfahrens mit reinen *Richtig/Falsch-Feedback* in Frage gestellt. Vor dem Hintergrund der bisherigen Analysen kann deshalb angenommen werden, dass bei der Konstruktion von dynamischen Testverfahren besonderes Augenmerk auf eine Implementierung von geeigneten fehlerorientierten Hilfen, d.h. *elaboriertes Feedback*, zu legen ist. Weiterhin sollte überlegt werden, in diesen Testprozeduren gänzlich auf Richtig/Falsch-Feedback zu verzichten.

Richtig/Falsch-Feedback hilft dem Testanden nicht, falsche Hypothesen zum Problemlöseprozess zu verwerfen. Dies trifft insbesondere für leistungsschwache Testanden zu. Richtig/Falsch-Feedback stellt keine „neue“ Information dar, wenn Testanden bereits ein ausgeprägtes Fähigkeits-Selbstkonzept zum Gegenstandsbereich besitzen. Personen mit wenig Vertrauen in ihre Fähigkeiten erleben Richtig/Falsch-Feedback eventuell als selbstwertbedrohlich, und zeigen deshalb Leistungsdefizite unter Feedback. Die Nutzung der Richtig/Falsch-Information kann nur von leistungsstarken oder selbstkonzeptstarken Testanden erwartet werden.

Hierbei ist aber zu bedenken, dass Leistungsdefizite selbstkonzeptschwacher als auch testängstlicher Testanden auch unter feedbackfreien Testbedingungen beobachtet wurden. Eine *kompensatorische* Wirkung von (Richtig/Falsch-)Feedback in Bezug auf ungünstige Leistungsvoraussetzungen, wie z.B. ein geringes Begabungskonzept oder Testangst, konnte nicht festgestellt werden. Richtig/Falsch-Feedback ist also einer kompensatorischen Wirkung von dynamischen Testprozeduren (Abschnitt 3.2.1) nicht zuträglich.

Die in der Literatur referierten Befunde sowie die Ergebnisse der vorliegenden Studie lassen folgende Schlussfolgerungen für die Gestaltung von Feedback-Strategien

in dynamischen Testverfahren zu. Feedback-Interventionen in dynamischen Tests sollten:

- fehlerhafte Problemlösestrategien korrigieren,
- auf explizite Richtig/Falsch-Informationen verzichten,
- die korrekte Itemlösung enthalten,
- eher computergestützt als durch einen Testleiter vorgegeben werden,
- weder Lob noch Tadel enthalten, z.B. „Das hast Du gut gemacht“, sowie
- einen intraindividuellen Vergleich zulassen, d.h. Leistungsveränderungen des Testanden anzeigen, wie beispielsweise das Leistungsniveaus nach jeder 5. Aufgabe rückmelden, bzw.
- keine interindividuellen normbezogenen Vergleiche fördern.

Nur bei Verwendung geeigneter Feedback-Strategien im Testprozess wird eine Unterschätzung der Leistungsfähigkeit von Personengruppen mit ungünstigen Leistungsvoraussetzungen, z.B. wenig selbstkonzeptstarker Testanden oder testängstlicher Testanden, verhindert und eine validere Diagnosestellung ermöglicht.

7.5 Ausblick

Die Befunde legen nahe, dass die Fähigkeit von Richtig/Falsch-Feedback zu profitieren, durch außerintellektuelle Personmerkmale moderiert wird. In Testsituationen mit Feedback erwies sich Vertrauen in die eigenen Fähigkeiten von besonderer Relevanz. Weitere Untersuchungen sollten die zugrunde liegenden Prozesse näher aufklären. Eine Frage könnte sein, inwiefern eine Regulation des Anstrengungsverhaltens zu Leistungseinbußen nach Feedback beiträgt. Ziel sollte dabei sein, Hinweise für die Gestaltung von Testsituationen, aber auch Lernumgebungen mit Feedback zu erarbeiten, die leistungsbeeinträchtigende differentielle Effekte von Feedback weitestgehend ausschließen, mindestens aber kontrollieren oder sogar kompensieren helfen. Da vor allem negatives Leistungsfeedback potentiell als Selbstwertbedrohung wahrgenommen wird, sollte dabei besonderes Augenmerk weiteren „selbstkonzept-nahen“ Konstrukten, wie z.B. dem Selbstwertgefühl oder dem Selbstwirksamkeitserleben zuteil werden.

8 Zusammenfassung

Ziel der vorliegenden Arbeit war es zu klären, inwiefern außerintellektuelle Personmerkmale, wie z.B. das Fähigkeits-Selbstkonzept einer Person, die Verarbeitung und Nutzung von leistungsbezogenem Feedback moderieren. Hintergrund dieser Fragestellung ist das Bestreben, das Leistungsverhalten von Testanden in Leistungstests mit Feedback, z.B. in dynamischen Tests (Guthke & Wiedl, 1996), näher aufzuklären. Im Gegensatz zu traditionellen rein statusorientierten Intelligenztests zeichnen sich dynamische Intelligenztests, z.B. Lerntests, unter anderem durch im Testprozess implementiertes performanzbezogenes Feedback aus. Die Fähigkeit des Testanden von Feedback bei der Lösung nachfolgender Aufgaben zu profitieren, wird als Indikator für seine intellektuelle Leistungsfähigkeit interpretiert. Der in dieser Art von Testprozeduren zu erwartende Einfluss außerintellektueller Personmerkmale auf die Testleistung wurde bisher kaum differentiell untersucht. Mit der hier vorgestellten Studie wird deshalb versucht, einen Beitrag zur Aufklärung differentieller Effekte ausgewählter leistungsrelevanter außerintellektueller Personmerkmale (Zielorientierung, Begabungskonzept, Leistungsangst) auf die Testleistung in Reasoning-Tests mit Feedback zu leisten. Es wurden im Rahmen eines experimentellen Designs 249 Schülern einer englischen Gesamtschule im Alter von 13 bis 15 Jahren Reasoning-Aufgaben aus zwei Materialbereichen (figural, numerisch) computerbasiert vorgegeben. Je nach experimenteller Bedingung erhielten die Schüler Richtig/Falsch-Feedback bzw. kein Feedback zu jeder ihrer Itemantworten. Die Befunde zeigen, dass die Fähigkeit von Feedback zu profitieren durch außerintellektuelle Personmerkmale moderiert ist. Eine wettbewerbsorientierte motivationale Grundhaltung (Leistungszielorientierung) sowie Selbstvertrauen in die eigenen Fähigkeiten unterstützen Leistungsgewinne unter Feedback-Bedingungen. Testangst geht dagegen generell mit Leistungseinbußen einher. Die teilweise hypothesenkonträr ausfallenden Effekte werden vor dem Hintergrund der Forschung zu Feedback-Phänomenen diskutiert.

9 Literatur

- Aguinis, H. (2004). *Regression analysis for categorical moderators*. New York, NY: Guilford Press.
- Aguinis, H., Petersen, S. A., & Pierce, C. A. (1999). Appraisal of the homogeneity of error variance assumption and alternatives to multiple regression for estimating moderating effects of categorical variables. *Organizational Research Methods, 2*, 315-339.
- Amelang, M. & Bartussek, D. (2001). *Differentielle Psychologie und Persönlichkeitsforschung* (5. aktualisierte und erweiterte Aufl.). Stuttgart: Kohlhammer.
- Ames, C., & Ames, R. (1984). Systems of students and teacher motivation: Toward a qualitative definition. *Journal of Educational Psychology, 76*, 535-556.
- Ammons, R. B. (1956). Effects of knowledge of performance: A survey and tentative theoretical formulation. *Journal of General Psychology, 54*, 279-299.
- Annett, J. (1969). *Feedback and human behaviour*. Harmondsworth, England: Penguin Books.
- Asendorpf, J. B. (1999). *Psychologie der Persönlichkeit* (2. überarbeitete und aktualisierte Aufl.). Berlin: Springer.
- Balke, S. & Stiensmeier-Pelster, J. (1995). Die Erfassung der motivationalen Orientierung – eine deutsche Form der Motivational Orientation Scales (MOS-D). *Diagnostica, 41*, 80-94.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.
- Bandura, A. (1991). Social cognitive theory of self regulation. *Organizational Behavior & Human Decision Processes, 50*, 248-287.
- Bangert-Drowns, R. L., Kulik, C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research, 61*, 213-238.
- Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychology research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.
- Bartussek, D. (1970). Eine Methode zur Bestimmung von Moderatoreffekten. *Diagnostica, 16*, 57-76.
- Beckmann, J. F. (2001). *Zur Validierung des Konstrukts des intellektuellen Veränderungspotentials*. Berlin: Logos.
- Beckmann, J. F. & Guthke, J. (1999). *Psychodiagnostik des schlussfolgernden Denkens*. Göttingen: Hogrefe.
- Benjamin, L. T. (1988). A history of teaching machines. *American Psychologist, 43*, 703-712.
- Berg, M. & Schaarschmidt, U. (1984). Überlegungen zu neuen Wegen in der Intelligenzdiagnostik. *Wissenschaftliche Zeitschrift der Humboldt Universität zu Berlin. Mathematisch-naturwissenschaftliche Reihe, 6*, 565-573.
- Bethge, H. J., Carlson, J. S., & Wiedl, K. H. (1982). The effects of dynamic assessment procedures on Raven Matrices performance, visual search behavior, test anxiety and test orientation. *Intelligence, 6*, 89-97.

- Blutke, E. (1979). *Der Zusammenhang zwischen sozial-charakterlichen Eigenschaften, sozialer und intellektueller Lernfähigkeit bei Vorschulkindern*. Unveröffentlichte Diplomarbeit, Universität Leipzig.
- Bortz, J. (1999). *Statistik für Sozialwissenschaftler* (5. vollständig überarbeitete und aktualisierte Aufl.). Heidelberg: Springer.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation* (2. vollständig überarbeitete und aktualisierte Aufl.). Berlin: Springer.
- Bossong, B. (1982). Kognitive Dissonanz, Attribution und Leistung. *Zeitschrift für Sozialpsychologie*, 13, 194-208.
- Briggs, S. R., & Check, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality and Social Psychology*, 54, 106-148.
- Broome, P. (1998). *Implizite Begabungstheorien und erlernte Hilflosigkeit*. Frankfurt: Lang.
- Broome, P. (2001). The gender related influence of implicate self-theories of one's intelligence with regard to academic performance in introductory physics classes. *Psychologische Beiträge*, 43, 100-123.
- Brunstein, J. C. (1989). Handlungsorientierte versus lageorientierte Reaktionen auf versuchsleiter-induzierte Misserfolgsereignisse. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 36, 349-367.
- Brunstein, J. C. (1995). *Motivation nach Misserfolg*. Göttingen: Hogrefe.
- Buchner, A., Faul, F., & Erdfelder, E. (1997). G•Power: A-priori, post-hoc and compromise power analysis for the Macintosh (Version 2.1.2). Trier, Germany: University of Trier.
- Budoff, M. (1970). Learning Potential: Assessing ability to reason in the educable mentally retarded. *Acta Paedopsychiatrica*, 37, 293-309.
- Budoff, M. (1975). *Learning Potential Measurement* (Final Report). Cambridge, MA: Research Institute of Educational Problems.
- Buggle, F. & Baumgärtel, F. (1972). *Hamburger Neurotizismus und Extraversionsskala für Kinder und Jugendliche*. Göttingen: Hogrefe.
- Burkert, D. (1981). *Die Beziehungen zwischen Ergebnissen in Status- und Lerntests (CPM, RKL, MFT) und Persönlichkeitsbesonderheiten (EFB 69) bei Schulanfängern*. Unveröffentlichte Diplomarbeit, Universität Leipzig.
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interests and performance. *Journal of Educational Psychology*, 79, 474-482.
- Campione, J. C., Brown, A. L., Ferrara, R. A., Jones, R. S., & Steinberg, E. (1985). Differences between retarded and non-retarded children in transfer following equivalent learning performance: Breakdowns in flexible use of information. *Intelligence*, 9, 297-315.
- Carbonell, J. R. (1970). AI in CAI: An artificial intelligence approach to computer-assisted instruction. *IEEE Transactions on Man-Machine Systems*, 11, 190-202.
- Carlson, J. S. (1989). Advances in research on intelligence: The dynamic assessment approach. *Mental Retardation & Learning Disability Bulletin*, 17, 1-20.
- Carlson, J. S., & Wiedl, K. H. (1976). *Modes of presentation of the Raven Coloured Matrices Tests: Toward a differential testing approach* (Trierer Psychologische Berichte Nr. 3/7). Trier: Universität Trier.

- Carlson, J. S., & Wiedl, K. H. (1979). Toward a differential testing approach: Testing-the-limits employing the Raven matrices. *Intelligence*, 3, 323-344.
- Carlson, J. S., & Wiedl, K. H. (1980). Applications of a dynamic testing approach in intelligence assessment: Empirical results and theoretical formulations. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 1, 303-318.
- Carlson, J. S., & Wiedl, K. H. (1992). Principles of dynamic assessment: The application of a specific model. *Learning & Individual Differences*, 4, 153-166.
- Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of processing in the Raven's Progressive Matrices Test. *Psychological Review*, 97, 404-431.
- Carver, C. S., & Scheier, M. F. (1981). *Attention and self regulation: A control theory to human behaviour*. New York, NY: Springer.
- Cervone, D., & Wood, R. (1995). Goals, feedback, and the differential influence of self-regulatory processes on cognitively complex performance. *Cognitive Therapy & Research*, 19, 519-545.
- Chen Idson, L., & Higgins, E. T. (2000). How current feedback and chronic effectiveness influence motivation: Everything to gain versus everything to lose. *European Journal of Social Psychology*, 30, 583-592.
- Clauß, G., Conrad, H., Knöchel, W. & Lohse, H. (1974). *Einführung in die Programmierung von Lehr- und Lernprozessen. Eine Anleitung für Lehrende an Hoch- und Fachschulen*. Berlin: Deutscher Verlag der Wissenschaften.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioural sciences* (3 ed.). Mahwah, NJ: Erlbaum.
- Cormier, P., Carlson, J. P., & Das, J. P. (1990). Planning ability and cognitive performance: The compensatory effects of a dynamic assessment approach. *Learning & Individual Differences*, 2, 437-449.
- Cronbach, L. J. (1951). Coefficient Alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- Cronbach, L. J. (1987). Statistical tests for moderator variables: Flaws in analyses recently proposed. *Psychological Bulletin*, 102, 414-417.
- Crowder, N. A. (1959). Automatic tutoring by means of intrinsic programming. In E. Galanter (Ed.), *Automatic teaching: The state of the art* (pp. 109-116). New York, NY: John Wiley and Sons.
- Deci, E. L. (1975). *Intrinsic motivation*. New York, NY: Plenum.
- Deffenbacher, J. L. (1978). Worry, emotionality, and task generated interference in test anxiety: An empirical test of attentional theory. *Journal of Educational Psychology*, 70, 248-254.
- Delgado, A. R., & Prieto, G. (2003). The effect of item feedback on multiple-choice test responses. *British Journal of Psychology*, 94, 73-85.
- Dresel, M. (2001). A longitudinal analysis of Dweck's Motivation-Process-Model in the classroom. *Psychologische Beiträge*, 43, 129-149.
- Duda, J. L., & Nicholls, J. G. (1992). Dimensions of achievement motivation in schoolwork and sport. *Journal of Educational Psychology*, 84, 290-299.
- Dunckel, H. (1986). Handlungstheorie. In G. Rexilius & S. Grubitzsch (Hrsg.), *Psychologie* (S. 534-553). Hamburg: Rohwolt.

- Dweck, C. S. (1986). Motivational processes affecting learning. Special Issue: Psychological science and education. *American Psychologist*, 41, 1040-1048.
- Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Philadelphia, PA: Psychology Press.
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review*, 95, 256-273.
- Elliot, A. J., & Church, M. A. (1997). A hierarchical model of approach and avoidance achievement motivation. *Journal of Personality & Social Psychology*, 72, 218-232.
- Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Abnormal and Social Psychology*, 70, 461-475.
- Elliott, E. S., & Dweck, C. S. (1988). Goals: An approach to motivation and achievement. *Journal of Personality and Social Psychology*, 54, 5-12.
- Elliott, J. G. (2003). Dynamic Assessment in Educational Settings: Realising potential. *Educational Review*, 55, 15-32.
- Elliott, J. G., Hufton, N., Illushin, L., & Lauchlan, F. (2001). Motivation in the junior years: international perspectives on children's attitudes, expectations and behaviour and their relationship to educational achievement. *Oxford Review of Education*, 27, 38-68.
- Embretson, S. E. (1995). Working memory capacity versus general central processes in intelligence. *Intelligence*, 20, 169-189.
- Embretson, S. E. (2000). Multidimensional measurement from dynamic tests: Abstract reasoning under stress. *Multivariate Behavioral Research*, 35, 505-542.
- Ettrich, K. U. & Guthke, J. (1988). Therapieorientierte Psychodiagnostik und Psychodiagnostik intraindividuelle Variabilität. In H. Schröder & J. Guthke (Hrsg.), *Fortschritte der klinischen Persönlichkeitspsychologie und klinischen Psychodiagnostik. Psychotherapie und Grenzgebiete* (Bd. 9, S. 95-105). Leipzig: Barth.
- Ewert, O. & Thomas, J. (1996). Das Verhältnis von Theorie und Praxis in der Instruktionspsychologie. In F. E. Weinert (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich D Praxisgebiete, Serie I Pädagogische Psychologie, Band 2 Psychologie des Lernens und der Instruktion* (S. 89-118). Göttingen: Hogrefe.
- Farr, J. L. (1991). Leistungsfeedback und Arbeitsverhalten. In H. Schuler (Hrsg.), *Beurteilung und Förderung beruflicher Leistung* (S. 57-80). Stuttgart: Verlag für Angewandte Psychologie.
- Fengler, J. (1998). *Feedback geben*. Weinheim: Beltz.
- Feuerstein, R. (1972). Cognitive assessment of the socioculturally deprived child and adolescent. In L. J. Cronbach & P. J. Drenth (Eds.), *Mental tests and cultural adaptation* (pp. 265-275). The Hague, Netherland: Mouton.
- Feuerstein, R., Rand, Y., & Hoffmann, M. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device theory, instruments and techniques*. Baltimore: University Park Press.
- Fischer, P. M. (1985). Wissenserwerb mit interaktiven Feedbacksystemen. In H. Mandl & P. M. Fischer (Hrsg.), *Lernen im Dialog mit dem Computer* (S. 68-82). München: Urban & Schwarzenberg.

- Fischer, P. M., & Mandl, H. (1988). Improvement of the acquisition of knowledge by informing feedback. In H. Mandl & A. Lesgold (Eds.), *Learning issues for intelligent tutoring systems* (pp. 187-242). New York, NY: Springer.
- Fischer, P. M., Mandl, H., Frey, H. D., Jeuck, J., Schröder, O. & Ackermann, K. (1988). *Bericht zum DFG-Projekt "Beeinflussung und Förderung des Wissenserwerbs mit audiovisuellen Medien bei kontingenter Rückmeldung"* (Forschungsbericht Nr. 48): Deutsches Institut für Fernstudien an der Universität Tübingen.
- Fodor, E. M., & Carver, R. A. (2000). Achievement and power motives, performance feedback, and creativity. *Journal of Research in Personality*, 34, 380-396.
- Förster, J., Grant, H., Idson, L.-C., & Higgins, E.-T. (2001). Success/failure feedback, expectancies, and approach/avoidance motivation: How regulatory focus moderates classic relations. *Journal of Experimental Social Psychology*, 37, 253-260.
- Försterling, F. & Schöler, J. (1984). Bedeutsamkeit und wahrgenommene eigene Fähigkeit als Determinanten der Informationssuche. *Archiv für Psychologie*, 136, 333-342.
- Försterling, F. & Schuster, B. (1987). Einige Determinanten von Ausdauer und Leistung. *Archiv für Psychologie*, 139, 37-47.
- Frey, D. & Greif, S. (Hrsg.). (1994). *Sozialpsychologie*. Weinheim: PVU.
- Frey, D., Stahlberg, D. & Fries, A. (1986). Informationssuche von stark und schwach ängstlichen Personen nach dem Erhalt positiver oder negativer selbstrelevanter Rückmeldung. *Journal of Personality*, 54, 694-703.
- Frohriep, K. (1978). Einige Ergebnisse zur psychodiagnostischen Validität eines neu entwickelten Kurzzeit-Lerntests für die Differentialdiagnostik entwicklungsrückständiger Vorschulkinder im Vergleich mit konventionellen Verfahren und Langzeit-Lerntests. In G. Clauß, J. Guthke & G. Lehwald (Hrsg.), *Psychologie und Psychodiagnostik lernaktiven Verhaltens* (S. 67-72). Berlin: Gesellschaft für Psychologie.
- Gamlin, P. J. (1996). Issues in dynamic assessment/instruction. In M. G. Luther & E. Cole (Eds.), *Dynamic assessment for instruction: From theory to application*. (pp. 79-87). North York: Captus Press.
- Gaudry, E., & Spielberger, C. D. (1971). *Anxiety and educational achievement*. Sydney: Wiley.
- Gerstenmaier, J. & Mandl, H. (1995). Wissenserwerb unter konstruktivistischer Perspektive. *Zeitschrift für Pädagogik*, 41, 867-888.
- Görn, A., Vollmeyer, R., & Rheinberg, F. (2001). Effects of motivational orientation and perceived ability on performance. *Psychologische Beiträge*, 43, 153-170.
- Grigorenko, E., & Sternberg, R. (1995). Thinking styles. In D. H. Saklofske & M. Zeidner (Eds.), *International handbook of personality and intelligence. Perspectives on individual differences* (pp. 205-229). New York, NY: Plenum Press.
- Grigorenko, E., & Sternberg, R. (1998). Dynamic testing. *Psychological Bulletin*, 124, 75-111.
- Günther, C. & Günther, R. (1981). Zur Bedingungsanalyse von Intelligenztestleistungen Erwachsener. Eine Untersuchung mit einem Langzeitlerntest. *Zeitschrift für Psychologie*, 189, 407-421.

- Günther, R. & Günther, C. (1982). Geistige Leistungsfähigkeit und Beanspruchungserleben von Herz-Kreislaufkranken unter psychischer Belastung durch einen Lerntest. *Zeitschrift für Psychiatrie, Neurologie und medizinische Psychologie*, 34, 339-346.
- Guthke, J. (1972). *Zur Diagnostik der intellektuellen Lernfähigkeit*. Berlin: DVW.
- Guthke, J. (1977). *Zur Diagnostik der intellektuellen Lernfähigkeit*. Stuttgart: Klett.
- Guthke, J. (1992). Learning tests: The concept, main research findings, problems and trends. *Learning & Individual Differences*, 4, 137-151.
- Guthke, J., Beckmann, J. F., Stein, H., Vahle, H. & Rittner, S. (1995). *Adaptive Computergestützte Intelligenz-Lerntestbatterie (ACIL)*. Mödling: Schuhfried.
- Guthke, J., Beckmann, J. F. & Wiedl, K. H. (2003). Dynamik im dynamischen Testen. *Psychologische Rundschau*, 54, 225-232.
- Guthke, J., & Lehwald, G. (1984). On component analysis of the intellectual learning ability in learning tests. *Zeitschrift für Psychologie*, 192, 3-17.
- Guthke, J. & Wiedl, K. H. (1996). *Dynamisches Testen. Zur Psychodiagnostik der intraindividuellen Variabilität*. Göttingen: Hogrefe.
- Hacker, W. (1978). *Allgemeine Arbeits- und Ingenieurpsychologie*. Bern: Huber.
- Hamers, J. H. M., & Ruijssenaars, A. J. J. M. (1984). *Learning ability and learning test*. Unpublished Dissertation, Katholik Universiteit Nijmegen.
- Hancock, T., Stock, W., & Kulhavy, R. (1992). Predicting feedback effects from response-certitude estimates. *Bulletin of the Psychonomic Society*, 30, 173-176.
- Harackiewicz, J. M., & Elliot, A. J. (1993). Achievement goals and intrinsic motivation. *Journal of Abnormal and Social Psychology*, 65, 904-915.
- Harackiewicz, J. M., & Sansone, C. (1991). Goals and intrinsic motivation: You can get there from here. In M. L. Maehr & P. R. Pintrich (Eds.), *Advances in motivation and achievement: Goals and self regulatory processes* (pp. 21-49). Greenwich, CT: JAI Press.
- Harris, M. D., Tetrick, L. E., & Tiegs, R. B. (1993). Cognitive ability and motivational interventions: Their effects on performance outcomes. *Current Psychology: Developmental, Learning, Personality, Social*, 12, 57-66.
- Haywood, H. C., & Switsky, H. N. (1974). Children's verbal abstracting: Effects of enrichment input, age, and IQ. *American Journal of Mental Deficiency*, 78, 556-565.
- Haywood, H. C., & Tzuriel, D. (Eds.). (1992). *Interactive Assessment*. New York, NY: Springer.
- Heckhausen, H. (1963). *Hoffnung und Furcht in der Leistungsmotivation*. Meisenheim am Glan: Hein.
- Heckhausen, H. (1974). *Leistung und Chancengleichheit*. Göttingen: Hogrefe.
- Heckhausen, H. (1989). *Motivation und Handeln* (2. völlig überarbeitete und ergänzte Aufl.). Berlin: Springer.
- Helton, W. S., Dember, W. N., Warm, J. S., & Matthews, G. (1999). Optimism, pessimism, and false failure feedback: Effects on vigilance performance. *Current Psychology: Developmental, Learning, Personality, Social*, 18, 311-325.
- Hentrich, O. & Reich, O. (1979). *Der Einfluss außerintellektueller Faktoren auf Lerntestergebnisse im 7. Schuljahr und deren korrelative Beziehungen im Kurz- und Langzeit-Lerntest*. Unveröffentlichte Diplomarbeit, Universität Leipzig.

- Herrmann, T. & Stapf, K.-H. (1973). Positive und negative Rückmeldung bei Alternativdiskriminationen: Rückmeldungsproportion und Lernstrategie. *Psychologische Forschung*, 36, 117-143.
- Hessels, M. (2000). The Learning Potential Test for Ethnic Minorities (LEM): A tool for standardized assessment of children in kindergarten and the first years of primary school. In C. Lidz & J. G. Elliott (Eds.), *Dynamic assessment: Prevailing models and applications* (Vol. 6, pp. 109-131). New York, NY: Elsevier Science.
- Heyme, B. & Militzer, H. (1977). *Die Wirkung von Persönlichkeitsdimensionen auf das Verhalten im Lerntest*. Unveröffentlichte Diplomarbeit, Universität Leipzig.
- Hock, M., Hardt, J. & Krohne, H. W. (1990). *Copingdispositionen, Angst und Blickverhalten in einer Leistungssituation nach unterschiedlichen Rückmeldungen* (Mainzer Berichte zur Persönlichkeitsforschung Nr. 32). Mainz: Universität, Psychologisches Institut, Abteilung Persönlichkeitspsychologie.
- Horn, W. (1969). *Prüfsystem für Schul- und Bildungsberatung (Handanweisung)*. Göttingen: Hogrefe.
- Jäger, R. S. (1978). *Differentielle Diagnostizierbarkeit in der psychologischen Diagnostik*. Göttingen: Hogrefe.
- Jerusalem, M. (1991). Allgemeine Selbstwirksamkeit und differentielle Streßprozesse. *Psychologische Beiträge*, 33, 388-406.
- Jerusalem, M. (1992). Die Bedeutung des Selbstkonzepts für Bedrohungserleben und Attributionen in Leistungssituationen. *Unterrichtswissenschaft*, 20, 293-307.
- Johnson, D. S., Perlow, R., & Pieper, K. F. (1993). Differences in task performance as a function of type of feedback: Learning-oriented versus performance-oriented feedback. *Journal of Applied Social Psychology*, 23, 303-320.
- Kagan, J. (1966). Reflection – impulsivity: The generality and dynamics of conceptual tempo. *Journal of Abnormal Psychology*, 71, 17-24.
- Kirchhof, O. (1995a). *Übergeneralisierung und Lageorientierung nach Misserfolg*. Unveröffentlichte Dissertation, Universität Bielefeld.
- Kirchhof, O. (1995b). Zur Bedeutung von Testaufgabenmerkmalen für die Entstehung von Leistungsdefiziten nach Misserfolg. In O. Güntürkün, R. Guski, C. Walter & A. Wohlschläger (Hrsg.), *Experimentelle Psychologie: Beiträge zur 37. Tagung experimentell arbeitender Psychologen* (S. 192): Ruhr Universität Bochum.
- Kirchhof, O. & Stiensmeier-Pelster, J. (1999). Kompensation und Leistung: Die Rolle von Testaufgabenrückmeldungen und Übergeneralisierung. *Empirische Pädagogik*, 13, 57-75.
- Klein, S. (1975). Lernfähigkeitsdiagnostik mit Hilfe von Unterrichtsmaschinen. *Probleme und Ergebnisse der Psychologie*, 51, 55-61.
- Kleitman, S., & Stankov, L. (2001). Ecological and person-oriented aspects of metacognitive processes in test-taking. *Applied Cognitive Psychology*, 15, 321-341.
- Kliegl, R., & Baltes, P. B. (1987). Theory-guided analysis of development and aging mechanisms through testing-the-limits and research on expertise. In C. Schooler & K. W. Schaie (Eds.), *Cognitive functioning and social structures over the life course* (pp. 95-119). Westport, CT: Ablex Publishing.
- Kluger, A. N. (2001). Feedback-expectation discrepancy, arousal and locus of cognition. In M. Erez & U. Kleinbeck (Eds.), *Work motivation in the context of a globalizing economy* (pp. 111-120). Mahwah, NJ: Lawrence Erlbaum Associates.

- Kluger, A. N., & Adler, S. (1993). Person- versus computer-mediated feedback. *Computers in Human Behavior*, 9, 1-16.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284.
- Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current Directions in Psychological Science*, 7, 67-72.
- Kolleker, A. (1999). *Konzeption und Erprobung eines Feedbackinstruments zur Beurteilung von Führungskompetenz im Assessment Center*. Unveröffentlichte Diplomarbeit, Bergische Universität Gesamthochschule Wuppertal.
- Köller, O. & Möller, J. (1995). Zum Einfluss der Instruktion auf die Anzahl und Lokation von Kausalattributionen nach der Bearbeitung eines komplexen Problems. *Empirische Pädagogik*, 9, 401-422.
- Köller, O. & Schiefele, U. (2001). Zielorientierung. In J. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 811-815). Weinheim: Beltz.
- Korsgaard, M. A., & Diddams, M. (1996). The effect of process feedback and task complexity on personal goals, information searching, and performance improvement. *Journal of Applied Social Psychology*, 26, 1889-1911.
- Krohne, H. W. (1996). *Angst und Angstbewältigung*. Stuttgart: Kohlhammer.
- Kröner, S., & Leutner, D. (2001). *Feedback in computer-aided instruction. Effects of response confidence*. Vortrag gehalten auf der 6. Arbeitstagung der Fachgruppe für Persönlichkeitspsychologie, Differentielle Psychologie und Psychologische Diagnostik, Leipzig.
- Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research*, 47, 211-232.
- Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1, 279-308.
- Kyllonen, P. C. (1991). Principles for creating a computerized test battery. *Intelligence*, 15, 1-15.
- Latta, R. M. (1978). Interactive effects of initial achievement orientation and prior success feedback on the mastery of subsequent difficulty and easy tasks. *American Educational Research Journal*, 15, 17-24.
- Lauchlan, F., & Elliott, J. (2001). The psychological assessment of learning potential. *British Journal of Educational Psychology*, 71, 647-665.
- Lauterbach, W., Stecker, K. & Partl, B. (1988). Hilflosigkeit und Reaktanz: Ihre Variation durch Kontrollmotivation, Attribution und Kontrollüberzeugung. *Zeitschrift für Klinische Psychologie. Forschung und Praxis*, 17, 275-291.
- Lawrence, I. (2004). *Introduction to The Fourth Spearman Conference on Diagnostics for Education: Theory, Measurement, and Applications*. Paper presented at the The Fourth Spearman Conference, Philadelphia, PA.
- Leutner, D. (1992). *Adaptive Lehrsysteme*. Weinheim: PVU.
- Leutner, D. (1998). Programmierter und Computerunterstützter Unterricht. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 404-409). Weinheim: Beltz.

- Lidz, C. S. (1990). The Preschool Learning Assessment Device: An approach to the dynamic assessment of young children. *European Journal of Psychology of Education, 5*, 167-175.
- Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports, 20*, 975-978.
- Lienert, G. A., & Raatz, U. (1994). *Testaufbau und Testanalyse*. (5. völlig neu bearbeitete und erweiterte Aufl.). Weinheim: Beltz.
- Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance*. Englewood Cliffs, NJ: Prentice Hall.
- Lugt-Tappeser, H., Schäfer, H., Scheiblechner, H. & Franzen, U. (1994). Das Aufgabenwahlverhalten hoch- und niedrig-ängstlicher Vorschulkinder nach einer Erfolgs- bzw. Misserfolgssituation bei einem Computerspiel. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 26*, 132-151.
- Lussier, C. M. (2002). Does dynamic assessment reduce the influence of stress on memory and reasoning? *Dissertation Abstracts International, 63(5-A)*, 1711, US: University Microfilms International.
- Mandl, H., Fischer, P. M., Frey, H.-D. & Jeuck, J. (1985). Wissensvermittlung durch ein computergestütztes Rückmeldungssystem. In H. Mandl & P. M. Fischer (Hrsg.), *Lernen im Dialog mit dem Computer* (S. 179-190). München: Urban & Schwarzenberg.
- Mandl, H., Gruber, H. & Renkl, A. (1997a). Lernen und Lehren mit dem Computer. In F. E. Weinert & H. Mandl (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich D Praxisgebiete, Serie I Pädagogische Psychologie, Band 4 Psychologie der Erwachsenenbildung* (S. 437-468). Göttingen: Hogrefe.
- Mandl, H., Gruber, H. & Renkl, A. (1997b). Situiertes Lernen in multimedialen Lernumgebungen. In L. J. Issing & P. Klimsa (Hrsg.), *Information und Lernen mit Multimedia* (S. 167-178). Weinheim: Beltz.
- Mandl, H., & Lesgold, A. (Eds.). (1988). *Learning issues for intelligent tutoring systems*. New York, NY: Springer.
- Meijer, J. (1993). Learning Potential, personality characteristics and test performance. In J. H. M. Hamers, K. Sijtsma & A. J. J. M. Ruijsenaars (Eds.), *Learning Potential Assessment: Theoretical, methodological and practical issues* (pp. 341-362). Amsterdam/Lisse: Swets & Zeitlinger.
- Meijer, J. (1996). *Learning Potential and fear of failure*. Amsterdam: Bauer.
- Meijer, J. (2001). Learning potential and anxious tendency: Test anxiety as a bias factor in educational testing. *Anxiety, Stress, & Coping, 14*, 337-362.
- Meijer, J., & Elshout, J. J. (2001). The predictive and discriminant validity of the zone of proximal development. *British Journal of Educational Psychology, 71*, 93-113.
- Meyer, W.-U. (1973). Anstrengungsintention in Abhängigkeit von Begabungseinschätzung und Aufgabenschwierigkeit. *Archiv für Psychologie, 125*, 245-262.
- Meyer, W.-U. (1984). *Das Konzept der eigenen Begabung*. Bern: Huber.
- Meyer, W.-U. & Starke, E. (1981/1982). Das Einholen begabungsrelevanter Informationen in Abhängigkeit vom Konzept eigener Begabung: Eine Feldstudie. *Archiv für Psychologie, 134*, 105-115.

- Michel, L. & Iseler, A. (1968). Beziehungen zwischen klinischen und psychometrischen Methoden der diagnostischen Urteilsbildung. In K. J. Groffmann & K.-H. Wewetzer (Hrsg.), *Person als Prozess* (S. 115-156). Bern: Huber.
- Mielke, R. (1996). Lerntheoretische Persönlichkeitskonstrukte. In M. Amelang (Hrsg.), *Enzyklopädie der Psychologie: Themenbereich C Theorie und Forschung, Serie VIII, Differentielle Psychologie und Persönlichkeitsforschung, Band 3 Temperaments- und Persönlichkeitsunterschiede* (S. 185-222). Göttingen: Hogrefe.
- Mikulincer, M. (1988). Reactance and helplessness following exposure to unsolvable problems: The effect of attributional style. *Journal of Personality and Social Psychology*, 54, 679-686.
- Mikulincer, M. (1994). *Human learned helplessness: A coping perspective*. New York, NY: Plenum Press.
- Miller, G. A., Galanter, S., & Pribram, K. (1960). *Plans and the structure of behavior*. New York, NY: Holt, Rinehart & Winston.
- Mix, R. (1981). Zur relativen Wirksamkeit positiver und negativer Rückmeldung: Automatisierter Kontrollprozess bei Mehrfachwahlaufgaben. *Psychologische Beiträge*, 23, 566-578.
- Morris, L. W., Davis, M. A., & Hutchings, C. A. (1981). Cognitive and emotional components of anxiety: Literature review and revised worry-emotionality scale. *Journal of Educational Psychology*, 73, 541-555.
- Müller-Schwarz, W. (1969). *Einführung in die Regelungstechnik*. München: Siemens AG.
- Mummendey, H.-D. (1999). *Die Fragebogenmethode* (3. Aufl.). Göttingen: Hogrefe.
- Murray, H. (1938). *Explorations in Personality*. New York, NY: Oxford University Press
- Musch, J. (1999). Die Gestaltung von Feedback in computergestützten Lernumgebungen: Modell und Befunde. *Zeitschrift für Pädagogische Psychologie*, 13, 148-160.
- Nagler, B. (1995). *Zum Einsatz von computergestützten Lerntests: Ein Vergleich von Lerntest- und Statustestdaten im Zusammenhang mit Basiskomponenten der Intelligenz und außerintellektuellen Kriterien im stationären Bereich einer Thüringer Jugendhilfeeinrichtung*. Unveröffentlichte Diplomarbeit, Universität Leipzig.
- Naglieri, J. A., & Das, J. P. (1987). Construct and criterion related validity of planning, simultaneous and successive cognitive processing tasks. *Journal of Psychoeducational Assessment*, 5, 353-363.
- Narciss, S. (2001). *Informative feedback as a bridge from instruction to learning in computer-based trainings*. Paper presented at the 9th European Conference for Research on Learning and Instruction, Fribourg, CH.
- National Foundation for Educational Research (Eds.). (1998). *Non-Verbal Reasoning NFER Entrance Test*. Berkshire: NFER-NELSON.
- Necka, E., & Stettner, Z. (2002). No feedback, please! Detrimental effects of feedback on performance in STM tasks. *Studia Psychologiczne*, 40, 31-54.
- Nicholls, J. G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice and performance. *Psychological Review*, 91, 328-346.
- Nicholls, J. G., Cobb, P., Wood, T., Yackel, E., & Patashnick, M. (1990). Assessing students theories of success in mathematics: Individual and classroom differences. *Journal of Research in Mathematics Education*, 21, 109-122.

- Pressey, S. L. (1926). A simple apparatus which gives tests and scores – and teaches. *School and Society*, 23, 373-376.
- Pressey, S. L. (1946). Further attempts to develop a “mechanical teacher.” *American Psychologist*, 1, 262.
- Quast, H. H. (1987). *Kognition und Leistung unter Belastung* (Bd. 43). Düsseldorf: VDI Verlag.
- Raven, J. C. (1956). *Guide to using the Coloured Progressive Matrices*. Dumfries: Grieve & Sons.
- Raven, J. C. (1976). *The Coloured Progressive Matrices*. London, England: Lewis.
- Raven, J. C., Court, J. H., & Raven, J. (1983). *Standard Progressive Matrices*. London, England: Lewis.
- Rawsthorne, L. J., & Elliott, A. J. (1999). Achievement goals and intrinsic motivation: A meta-analytic review. *Personality and Social Psychology Review*, 3, 326-344.
- Resing, W. (2000). Assessing the learning potential for inductive reasoning (LIR) in young children. In C. Lidz & J. G. Elliott (Eds.), *Dynamic assessment: Prevailing models and applications* (Vol. 6, pp. 229-262). New York, NY: Elsevier Science.
- Rheinberg, F. (1980). *Leistungsbewertung und Lernmotivation*. Göttingen: Hogrefe.
- Rheinberg, F. & Weich, K. W. (1988). Wie gefährlich ist Lob. Eine Untersuchung zum "paradoxen Effekt" von Lehrersanktionen. *Zeitschrift für Pädagogische Psychologie*, 2, 227-233.
- Roos, L. L., Wise, S. L., & Plake, B. S. (1997). The role of item feedback in self-adapted testing. *Educational and Psychological Measurement*, 57, 85-98.
- Rousseau, F. L., & McKelvie, S. J. (2000). Effects of bogus feedback on intelligence test performance. *Journal of Psychology*, 134, 5-14.
- Ryder, M. G. (1995). An analysis of the effects of motivational components on performance during a dynamic assessment procedure. *Dissertation Abstracts International Section A: Humanities & Social Sciences*, 55(8-A), 2356.
- Sarason, I. G., & Spielberger, C. D. (1975). *Stress and anxiety* (Vol. 2). Washington, DC: Hemisphere Publishing.
- Saunders, D. R. (1956). Moderator variables in prediction. *Educational and Psychological Measurement*, 16, 209-222.
- Saunders, D. R. (1966). The moderator variable as a useful tool in prediction (paper 1954). In A. Anastasi (Ed.), *Testing problems in perspective: 25th anniversary volume of topical readings from the Invitational Conference on Testing Problems* (pp. 301-306). Washington, DC: American Council on Education.
- Schmalt, H.-D. (1994). Zur Aufhebung und Akzentuierung von Leistungsdefiziten nach nichtkontingentem Misserfolg. *Zeitschrift für Experimentelle und Angewandte Psychologie*, 41, 261-278.
- Schneider, K. (1978). Die Wirkung von Erfolg und Misserfolg auf die Leistung bei einer visuellen Diskriminationsaufgabe und auf physiologische Anstrengungsindikatoren. *Archiv für Psychologie*, 130, 69-88.
- Schwarzer, R. & Jerusalem, M. (1982). Selbstwertdienliche Attributionen nach Leistungsrückmeldungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 14, 47-57.
- Seligman, M. E. P. (1975). *Helplessness*. San Francisco: Freeman.

- Shrauger, J. S., & Rosenberg, S. E. (1970). Self-esteem and the effects of success and failure feedback on performance. *Journal of Personality*, 38, 404-417.
- Slembek, E. & Geißner, H. (Hrsg.). (1998). *Feedback – Das Selbstbild im Spiegel der Fremdbilder*. St. Ingbert: Röhrig.
- Smith, M., Duda, J., Allen, J., & Hall, H. (2002). Contemporary measures of approach and avoidance goal orientations: Similarities and differences. *British Journal of Educational Psychology*, 72, 155-190.
- Snow, R. E. (1989). Aptitude-treatment interaction as a framework for research on individual differences in learning. In P. L. Ackerman & R. J. Sternberg (Eds.), *Learning and individual differences: Advances in theory and research* (pp. 13-59). New York, NY: Henry Holt & Co.
- Spillner, G. & Spillner, K.-H. (1979). *Untersuchungen an Vorschulkindern zur Normierung des RKL (Raven-Kurzzeit-Lerntest) und zur Beeinflussung von Lerntestergebnissen durch außerintellektuelle Faktoren*. Unveröffentlichte Diplomarbeit, Universität Leipzig.
- Spinath, B. & Stiensmeier-Pelster, J. (2000). Zielorientierung und Leistung: Die Rolle des Selbstkonzepts eigener Fähigkeiten. In H. Metz-Goeckel, B. Hannover & S. Leffelsend (Hrsg.), *Selbst, Motivation und Emotion* (S. 45-55). Berlin: Logos Verlag.
- Stake, J. E. (1982). Reactions to positive and negative feedback: Enhancement and consistency effects. *Social Behavior & Personality*, 10, 151-156.
- Stankov, L., & Crawford, J. D. (1997). Self-confidence and performance on tests of cognitive abilities. *Intelligence*, 25, 93-109.
- Stapf, K. H. & Degner, U. (1981). *Experimentelle Untersuchungen verschiedener Rückmeldungsmodalitäten beim Lernen* (Forschungsbericht Nr. 5). Tübingen: Universität, Institut für Psychologie.
- Stapf, K. H., Fischer, P. M. & Degner, U. (1986). Über die informationelle und motivationale Wirkung verschiedener Rückmeldungsmodalitäten beim Lernen. In K. Daumenlang & J. Sauer (Hrsg.), *Aspekte psychologischer Forschung. Festschrift zum 60. Geburtstag von Erwin Roth* (S. 221-236). Göttingen: Hogrefe.
- Sternberg, R. (1988). Mental self-government: A theory of intellectual styles and their development. *Human Development*, 31, 197-221.
- Sternberg, R., & Grigorenko, E. (2002). *Dynamic testing*. Cambridge: Cambridge University Press.
- Stiensmeier-Pelster, J., Balke, S. & Schlangen, B. (1996). Lern- versus Leistungszielorientierung als Bedingung des Lernfortschritts. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 28, 169-187.
- Stiensmeier-Pelster, J. & Schlangen, B. (1996). Erlernte Hilflosigkeit und Leistung. In J. Möller & O. Köller (Hrsg.), *Emotionen, Kognitionen und Schulleistung* (S. 69-90). Weinheim: Beltz.
- Swanson, H. L. (1995). *The Swanson-Cognitive-Processing Test*. Austin, Texas: Pro-Ed.
- Swanson, H. L., & Lussier, C. M. (2001). A selective synthesis of the experimental literature on dynamic assessment. *Review of Educational Research*, 71, 321-363.
- Thompson, T., Webber, K., & Montgomery, I. (2002). Performance and persistence of worriers and non-worriers following success and failure feedback. *Personality and Individual Differences*, 33, 837-848.

- Thorndike, E. L. (1898). *Animal Intelligence: An Experimental Study of the Associative Processes in Animals* (Psychological Review, Monograph Supplements No. 8). New York, NY: Macmillan.
- Thorndike, E. L. (1911). *Animal intelligence*. New York, NY: Macmillan.
- Thorndike, E. L. (1932). *The fundamentals of learning*. New York, NY: Teachers College.
- Thorndike, R. L., & Hagen, E. P. (1993). *Cognitive abilities test*. Itasca, IL: Riverside Publishing.
- Thurstone, L. L. (1939). *Manual of instructions for the primary mental abilities tests*. Washington, DC: American Council of Education.
- Thurstone, T. G., & Thurstone, L. L. (1949). *Mechanical Aptitude II: Description of group tests* (Report No. 54). Chicago, IL: University of Chicago, Psychometric Laboratory.
- Tonidandal, S., Quinones, M. A., & Adams, A. A. (2002). Computer adaptive testing: The impact of test characteristics on perceived performance and test takers' reactions. *Journal of Applied Psychology, 87*, 320-332.
- Tuckey, M., Brewer, N., & Williamson, P. (2002). The influence of motives and goal orientation on feedback seeking. *Journal of Occupational & Organizational Psychology, 75*, 195-216.
- Tzuriel, D. (1999). Parent-child mediated learning interactions as determinants of cognitive modifiability: Recent research and future directions. *Genetic, Social, & General Psychology Monographs, 125*, 109-156.
- Tzuriel, D. (2001). Dynamic Assessment is not dynamic testing. *Issues in education, 2*, 238-249.
- Tzuriel, D., & Alfassi, M. (1994). Cognitive and motivational modifiability as a function of Instrumental Enrichment (IE) Program. *Special Service in the Schools, 8*, 91-128.
- Tzuriel, D., & Klein, P. S. (1985). Analogical thinking modifiability in disadvantaged, regular, special education, and mentally retarded children. *Journal of Abnormal Child Psychology, 13*, 539-552.
- Tzuriel, D., Samuels, M. T., & Feuerstein, R. (1988). Non-intellective factors in dynamic assessment. In R. M. Gupta & P. Coxhead (Eds.), *Cultural diversity and learning efficiency: Recent developments in assessment* (pp. 141-163). New York, NY: St Martin's Press.
- Tzuriel, D., & Shamir, A. (2002). The effects of mediation in computer assisted dynamic assessments. *Journal of Computer Assisted Learning, 18*, 21-32.
- University of Durham (Eds.). (2000). *Middle Years Information System (MidYIS)*. Durham: CEM Centre.
- Utman, C. H. (1997). Performance effects of motivational state: A Meta-Analysis. *Personality and Social Psychology Review, 1*, 170-182.
- Van-Dijk, D., & Kluger, A. N. (2004). Feedback sign effect on motivation: Is it moderated by regulatory focus? *Applied Psychology: An International Review, 53*, 113-135.
- VandeWalle, D., Cron, W. L., & Slocum, J. W. (2001). The role of goal orientation following performance feedback. *Journal of Applied Psychology, 86*, 629-640.
- VandeWalle, D., & Cummings, L. L. (1997). A test of the influence of goal orientation on the feedback seeking process. *Journal of Applied Psychology, 82*, 390-400.
- Vilsmeier, C. (2000). *Feedback geben – mit Sprache handeln*. Berlin: Metropolitan Verlag.

- Volpert, W. (1982). Das Modell der hierarchisch-sequentiellen Handlungsorganisation. In W. Hacker (Hrsg.), *Kognitive und motivationale Aspekte der Handlung* (S. 38-58). Bern: Huber.
- von Hornstein, E. & von Rosenstiel, L. (2000). *Ziele vereinbaren – Leistung bewerten*. München: Wirtschaftsverlag Langen Müller Herbig.
- Webb, J. M., Pridemore, D. R., Kulhavy, W., Stock, W. A., & Henning, J. E. (1997). Remembering responses and cognitive estimates of knowing: The effects of instruction, retrieval sequences, and feedback. *Contemporary Educational Psychology*, 22, 147-164.
- Weiner, B. (1986). *An attributional theory of motivation and emotion*. Berlin: Springer.
- Wiedl, K.-H. (1980). Kompensatorische Intervention im Rahmen intelligenzdiagnostischer Untersuchungen bei kognitiv impulsiven Kindern. *Zeitschrift für klinische Psychologie*, 9, 219-231.
- Wiedl, K.-H. & Bethge, H.-J. (1981). Zur Auswirkung regulationsfördernder Situationsveränderungen auf Intelligenzleistung und Blickverhalten kognitiv impulsiver Kinder. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 13, 127-141.
- Wiedl, K.-H. & Bethge, H.-J. (1983). Die Anpassung der aufgabenbezogenen Betrachtungszeit an variierende Aufgabenschwierigkeiten: Deskriptive und veränderungsbezogene Analysen bei kognitiv impulsiven und reflexiven Kindern. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 4, 67-77.
- Wiedl, K.-H., Bethge, H.-J. & Bethge, H. (1982). Situative Veränderungen von Leistungsangst, Selbstbild und Situationsbewertung bei Anwendung von Lerntestprozeduren. *Psychologie in Erziehung und Unterricht*, 29, 206-211.
- Wiener, N. (1948). *Kybernetik*. Düsseldorf: Econ-Verlag.
- Wiener, N. (1952). *Mensch und Menschmaschine*. Berlin: Alfred Metzner Verlag.
- Wilder, J. (1931). Das Ausgangswertgesetz – Ein unbeachtetes biologisches Gesetz; seine Bedeutung für Forschung und Praxis. *Klinische Wochenschrift*, 10, 1889-1893.
- Winefield, A., & Rourke, J. A. (1991). Personality differences in responding to nonkontingent reinforcement. *Personality and Individual Differences*, 12, 281-289.
- Wygotski, L. S. (1964, russ. 1934). *Denken und Sprechen*. Berlin: Akademie-Verlag.
- Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482.
- Ziegler, A. (2001). Achievement motivation and implicit theories of intelligence. *Psychologische Beiträge*, 43, 1-21.
- Ziegler, A. & Schober, B. (1999). Implizite Theorien über die eigene Intelligenz bei Grundschüler(inne)n und ihr Einfluss auf hilfloses Verhalten in den Fächern Mathematik und Musik. In C. Enders, C. Hanckel & S. Möley (Hrsg.), *Lebensraum – Lebenstraum – Lebenstrauma Schule*. Kongressbericht der 13. Bundeskonferenz 1998 in Halle an der Saale (S. 386-394). Bonn: Deutscher Psychologen Verlag.
- Zirkle, G. A. (1946). Success and failure in serial learning I. The Thorndike effect. *Journal of Experimental Psychology*, 36, 230-236.

10 Anhang

Situationspezifische Items (vor der Testung)	II
Situationspezifische Items (nach der Testung)	III
Items zur trait-nahen Erfassung der Konstrukte	IV

Situationspezifische Items (vor der Testung)

Vor der Testung wurden alle Testanden mündlich standardisiert instruiert (siehe Kapitel 5). Die im Folgenden dargestellte zusätzliche schriftliche Instruktion beinhaltet außerdem die Erklärung des Antwortformats der persönlichkeitsdiagnostischen Items mittels zweier Beispielitems.

Instruktion (Ausschnitt): „*We would like to know how you feel about tackling problems such as those you see on the screen in front of you. All the following questions concern the problems presented on the computer. Please indicate below how much you agree with the following statements! Remember, you need to put one cross on each line showing how much you agree with the statement.*“

1. I feel pretty confident that I shall be able to solve most of the problems. vs. I do *not* feel confident that I shall be able to solve most of the problems.
2. The more problems I tackle, the easier it will be to solve them. vs. I do *not* think, that tackling more problems will help me to solve them.
3. I think I will have done well on this test if I get more correct answers than other students. vs. I do *not* think that I will have done well unless I get more correct answers than other students.
4. Even if I get most of the problems wrong, I will feel that I have done *well* if I learn something interesting. vs. If I get most of the problems wrong, I will feel that I have done *poorly* even if I learn something interesting.
5. What is important to me is getting a *higher score* on the test than everyone else. vs. What is important to me is *learning* how to solve the problems.
6. I will feel really successful if I can solve the problems *without* having to make a lot of *effort*. vs. I will feel really successful if I can solve the problems once I have *worked really hard*.
7. Just now I am *nervous* about doing these problems. vs. Just now I am *relaxed* about doing these problems.
8. I am afraid I *may not* do as well on this test as I could. vs. I am pretty optimistic that I will do as well on this test as I can.

Situationspezifische Items (nach der Testung)

Instruktion: „*Now that you have finished the test, we would like to ask you a few questions. As before, please read each of the statements outlined below carefully and indicate the extent to which you agree with them. Please do this by placing one cross on each line. There are no right or wrong answers. We do ask that you answer honestly. Please respond to every statement. Remember, all your answers will be made anonymous so no-one at home or school will know how you have responded.*“

1. Just now I am *nervous*. vs. Just now I am *relaxed*.
2. I am afraid I *may not* have done as well on this test as I could. vs. I think I have done as well as I could.
3. The more problems I tackled the easier it was to solve them. vs. I do *not* think, that tackling more problems helped me to solve them.
4. I think I have done *well*. vs. I think I have done *poorly*.
5. What was important to me was getting a *higher score* on the test than everyone else. vs. What was important to me was *learning* how to solve the problems.
6. I think I gave more correct answers than most of the other students. vs. I think most of the *other students* gave more correct answers than me.
7. I have learned something interesting. vs. I have *not* learned anything interesting.
8. I worked really hard. vs. I did *not* work really hard.

(nur Feedback-Bedingung:) „*Immediately after answering each item you were able to see whether your answer was correct or wrong. How do you feel about getting this kind of feedback?*“

1. I found it irritating. vs. I did *not* find it irritating.
2. It was important for me to know whether I was correct or incorrect. vs. It was *not* important for me to know whether I was correct or incorrect.
3. It helped me to solve some of the problems that came afterwards. vs. It did *not* help me to solve any problems that came afterwards.

Items zur trait-nahen Erfassung der Konstrukte

Zielorientierung und Anstrengungsvermeidung:

When do you personally feel most successful in your schoolwork?

1. I feel really successful when I know more than other people.
2. I feel really successful when what I learn really makes sense.
3. I feel really successful when I have better test scores than other people.
4. I feel really successful when I work really hard.
5. I feel really successful when something I learn makes me think about things.
6. I feel really successful when I do not have anything tough to do.
7. I feel really successful when others get things wrong and I do not.
8. I feel really successful when I do my very best.
9. I feel really successful when I do not have to try hard.
10. I feel really successful when I am the only one who can answer questions.
11. I feel really successful when I get a new idea about how things work.
12. I feel really successful when others are not as able to do as well as me.
13. I feel really successful when I solve a problem by working hard.
14. I feel really successful when I beat others.
15. I feel really successful when I learn something interesting.
16. I feel really successful when I can do better than my friends.
17. I feel really successful when something I learn makes me want to find out more.
18. I feel really successful when I am the cleverest.

Ursachenerklärung:

What do you think is most likely to help students to do well or succeed in schoolwork?

19. People succeed if they enjoy thinking about their school subjects.
20. People succeed if they work really hard.
21. People succeed if they are cleverer than others.
22. People succeed if they know how to make themselves look better than they are.
23. People succeed if they help each other learn.
24. People succeed if they are better than others in taking tests.
25. People succeed if they know how to impress the teacher.
26. People succeed if they are interested in learning.
27. People succeed if they are attractive and have the right clothes.

28. People succeed if they try to figure things out.
29. People succeed if they are born naturally intelligent.
30. People succeed if they are just lucky.
31. People succeed if they try to understand instead of just memorizing things.
32. People succeed if they know how to cheat.
33. People succeed if they always do their best.

Testangst:

How do you feel, when you have tests at school?

34. I often think how much brighter the other students are than I am.
35. I usually think I will do well.
36. I usually think that others will do better on tests than I will.
37. I am usually nervous when I have a test to do.
38. I am often concerned about the possibility of getting bad marks.
39. I usually feel worried and concerned when I have a test to do.
40. I am often afraid I may not do as well on the test as I could.
41. I am usually relaxed when I have a test to do.
42. I usually feel anxious when I have a test to do.
43. I usually feel satisfied with my performance on tests.
44. I am often afraid that the items in the tests will be too difficult for me.
45. I often feel miserable when I have a test to do.
46. I often feel that if I do badly others will be disappointed in me.
47. I usually do *not* feel very confident about my performance on tests.
48. I often think that I should have studied more for the test.

Begabungskonzept:

What do you think about yourself?

49. If I thought I was not going to do well at a task, I probably would not try hard even though I might learn a lot from it.
50. It is much more important for me to learn things in my classes than it is to get the top marks.

Zielorientierung:

What is more true for you? Please decide, to which of the following statements do you agree?

51. I feel pretty confident about my intellectual ability. vs. I am not very confident about my intellectual ability.
52. When I get new work in school, I am usually confident I will be able to learn it. vs. When I get new work in school, I often think I may not be able to learn it.

Selbständigkeitserklärung

Die vorliegende Arbeit wurde von mir eigenständig und ohne unerlaubte Hilfe angefertigt. Ich habe keine anderen als die im Schriftenverzeichnis angeführten Quellen benutzt und sämtliche Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen wurden, als solche kenntlich gemacht.

Nadin Beckmann

New Haven, CT im November 2004