

Schriftenreihe „Mannheimer sozialwissenschaftliche Abschlussarbeiten“, Nr. 005/07

## DIPLOMARBEIT

**Höheres Interesse, schlechtere Leistung:  
Geschlechtsspezifische Leistungserwartung in der Mathematik  
und ihr Einfluss auf die Testleistung in der PISA-Studie 2003**

**Markus Tausendpfund**

Im Wingert 5  
68519 Viernheim  
☎ 06204/912843

✉ [post@markus-tausendpfund.de](mailto:post@markus-tausendpfund.de)

Diplom-Sozialwissenschaften (10)

Matrikelnummer: 0770981

Tag der Abgabe: 5. Oktober 2005

## **Schriftenreihe: Mannheimer sozialwissenschaftliche Abschlussarbeiten**

### **Vorwort**

Die Fakultät für Sozialwissenschaften der Universität Mannheim bildet in den Fächern Politikwissenschaft, Soziologie, Psychologie und Erziehungswissenschaft qualifizierten Forschungs- und Führungsnachwuchs aus. Viele Abschlussarbeiten der Studierenden zeugen von dem hohen wissenschaftlichen Niveau der Mannheimer Ausbildungsinhalte, die insbesondere gekennzeichnet sind von der empirisch-analytischen Ausrichtung unter Betonung quantitativer Methoden.

Die Ergebnisse und Inhalte vieler dieser Arbeiten sind publikationswürdig. Aus diesem Grund bietet die Fakultät für Sozialwissenschaften ihren besten Absolventen die Möglichkeit, ihre Arbeiten einem breiteren Publikum zu präsentieren und hat hierfür diese Schriftenreihe ins Leben gerufen. Diese Schriftenreihe soll dazu beitragen, die wissenschaftlichen Ergebnisse der besten Abschlussarbeiten dem Fachpublikum zugänglich zu machen. Damit sind sie für weitere Untersuchungen verfügbar und können eventuell eine Grundlage für weitere Forschungen bieten.

In dieser Reihe werden nur Abschlussarbeiten veröffentlicht, die von beiden Gutachtern mit „sehr gut“ bewertet und für veröffentlichungswürdig befunden wurden.

Prof. Dr. Josef Brüderl  
Dekan der Fakultät für Sozialwissenschaften

**„The true nature of the relationship between gender and mathematics is much more complex than most people have been led to believe.“**

Gallagher & Kaufman, 2005, S. 316

# Inhaltsverzeichnis

<b>1. Einleitung</b>	<b>Seite 3</b>
<b>2. Die Situation von Mädchen und Frauen in der Mathematik</b>	<b>Seite 5</b>
<b>3. Gibt es geschlechtsspezifische Leistungsunterschiede in der Mathematik?</b>	<b>Seite 7</b>
<b>4. Erklärungsansätze für Geschlechterdifferenzen in der Mathematik</b>	<b>Seite 14</b>
4.1 Biologische Erklärungsansätze	Seite 15
4.1.1 Genetische Erklärungen	Seite 16
4.1.2 Hormonelle Erklärungen	Seite 17
4.1.3 Neuropsychologische Erklärungen	Seite 19
4.1.4 Ursprung und Konsequenzen der Geschlechtsunterschiede	Seite 21
4.2 Sozialisationstheoretische Erklärungsansätze	Seite 23
4.2.1 Geschlechtsspezifische Sozialisation	Seite 24
4.2.2 Psychosoziale Erklärungsmodelle	Seite 32
4.2.2.1 Autonomes Lernverhalten	Seite 32
4.2.2.2 Wert-Erwartungs-Ansatz	Seite 33
4.3 Psychobiosoziales Modell	Seite 35
4.4 Einfluss von Stereotypen in der Testsituation	Seite 37
4.4.1 Theoretische Annahmen	Seite 38
4.4.2 Empirische Befunde	Seite 41
4.5 Theorie des regulatorischen Fokus	Seite 53
4.5.1 Theoretische Annahmen	Seite 53
4.5.2 Das MERF-Modell	Seite 57
4.5.3 Empirische Befunde	Seite 58
4.6 Kurze Zwischenbetrachtung	Seite 61
<b>5. Die PISA-Studie: Ein Überblick</b>	<b>Seite 63</b>
5.1 Design der Studie	Seite 64
5.2 Messinstrumente	Seite 66
5.3 Hypothesen	Seite 69

## Inhaltsverzeichnis

<b>6. Empirische Analysen</b>	Seite 73
6.1 Deutschland	Seite 73
6.2 Internationaler Vergleich	Seite 85
6.2.1 Ländervergleich mit Deutschland, Österreich und Schweiz	Seite 86
6.2.2 Ländervergleich mit Dänemark, Finnland, Norwegen und Schweden	Seite 88
6.2.3 Ländervergleich mit Spanien, Italien und Portugal	Seite 91
6.2.4 Ländervergleich mit den Niederlanden, Belgien und Luxemburg	Seite 94
6.2.5 Kurze Zwischenbetrachtung	Seite 96
6.3 Länderunterschiede in den geschlechtsspezifischen Leistungsdifferenzen	Seite 97
6.3.1 Vergleich Schweiz und Österreich	Seite 99
6.3.2 Vergleich Dänemark und Norwegen	Seite 100
6.3.3 Vergleich Italien und Spanien	Seite 102
6.3.4 Vergleich Luxemburg und Niederlande/Belgien	Seite 103
6.3.5 Kurze Zwischenbetrachtung	Seite 104
<b>7. Von der Mathematik zu den Naturwissenschaften</b>	<b>Seite 105</b>
7.1 Design und Messinstrumente	Seite 106
7.2 Empirische Analysen	Seite 108
<b>8. Biologie, Sozialisation oder Situation?</b>	<b>Seite 115</b>
<b>9. Geschlechtsspezifische Leistungsunterschiede: Handlungsempfehlungen</b>	<b>Seite 118</b>
<b>10. Resümee</b>	<b>Seite 122</b>
<b>11. Literaturverzeichnis</b>	<b>Seite 124</b>
<b>12. Anhang</b>	
Abbildungsverzeichnis	Seite 134
Tabellenverzeichnis	Seite 135
Hinweis zur DVD	Seite 137
Persönliche Erklärung	Seite 138

## 1. Einleitung

Mathematische Fähigkeiten werden in einer modernen Gesellschaft zur Bewältigung zahlreicher Anforderungen benötigt. Ohne Kenntnisse der Prozentrechnung kann nicht verstanden werden, wie viel Geld bei Sonderangeboten gespart wird. Das Prinzip von Glücksspielen kann nur durchschauen, wer über Grundkenntnisse der Wahrscheinlichkeitsrechnung verfügt. Mathematisches Wissen ist aber auch zur Modellierung von Situationen oder Gesetzmäßigkeiten und zur Vorhersage von Ereignissen notwendig (Stern, 1997). Die Beispiele zeigen: Mathematische Kompetenz kann als mächtiges Werkzeug bei der Lösung unterschiedlichster Probleme dienen und ist für das Verständnis zahlreicher Phänomene in der Welt unentbehrlich.

Die große Bedeutung mathematischer Kompetenz ergibt sich ferner aus einer weit reichenden Mathematisierung vielfältiger Berufs-, Wirtschafts- und Kulturbereiche. Die objektiven Möglichkeiten, einen mathematisch-naturwissenschaftlichen Ausbildungsweg zu wählen, sind heute für Mädchen und Jungen annähernd gleich. Trotz wachsender Berufsorientierung und Berufstätigkeit der Frauen sind die mathematisch-naturwissenschaftlichen Berufe mit ihren guten Aufstiegs- und Einkommenschancen noch immer eine Männerdomäne. Liegt die niedrige Zahl von Mathematikerinnen an einer durchschnittlich geringeren Begabung, an einem natürlichen Vorteil des männlichen Geschlechts bei mathematischen Aufgaben oder an dem Geschlechtsstereotyp, wonach Mädchen mathematisch unbegabt seien, das mathematisch durchaus fähige und interessierte Mädchen während der Schulzeit entmutigt?

Die bisherige Forschung zur Erklärung von geschlechtsspezifischen Leistungsunterschieden in der Mathematik lässt sich grob den Polen Nature und Nurture zuordnen. Auf der Nature-Seite sind Ansätze anzusiedeln, die geschlechtsspezifische Leistungsdifferenzen auf biologische Faktoren zurückführen, während sich auf der Nurture-Seite Erklärungen finden, die diese Unterschiede als Resultat von unterschiedlichen Sozialisationseinflüssen ansehen.

Im Mittelpunkt dieser Arbeit steht ein relativ neuer Ansatz, der die Wirkung von gesellschaftlichen Erwartungen in der Testsituation betont. Claude Steele (1997) geht davon aus, dass Mitglieder von Gruppen, über die ein negatives Stereotyp bezüglich ihrer Kompetenz existiert, allein durch das Wissen um die Existenz dieses Stereotyps in ihren Leistungen beeinträchtigt werden können. Ein immer noch weit verbreitetes Stereotyp spricht Frauen im Vergleich zu Männern geringere mathematische Fähigkeiten zu. Nach Steele werden in Situationen, in denen sich eine Frau der Gefahr ausgesetzt sieht, dieses Stereotyp zu bestätigen, Prozesse ausgelöst, die von der korrekten Bewältigung der Aufgabe ablenken. Dieses Phänomen wird als Stereotype Threat bezeichnet. Stereotype Threat kann als ein besonderer Leistungsdruck verstanden werden, der unter bestimmten Randbedingungen zu einer Reduzierung der Leistungsfähigkeit führt.

In zahlreichen experimentellen Studien konnte der Stereotype-Threat-Effekt bereits nachgewiesen werden. Gegenüber der Stereotype-Threat-Theorie wird jedoch kritisch eingewendet, dass die Forschung unter künstlichen Bedingungen stattfindet, deren Ergebnisse nicht auf das tatsächliche Leben übertragen werden können (Davies & Spencer, 2005).

Diese Arbeit versucht die Kritik bezüglich selektiver Stichproben und künstlichen Testbedingungen zu überwinden. Anstatt die Stereotype-Threat-Theorie einem weiteren experimentellen Test zu unterziehen, wird regressionsanalytisch geprüft, ob sich bei einem repräsentativen Mathematiktest Hinweise finden, die mit der Stereotype-Threat-Theorie kompatibel sind. Für die empirischen Analysen greift diese Arbeit in erster Linie auf das Programme for International Student Assessment 2003 (PISA) zurück, der PISA-Datensatz wurde im Dezember 2004 vom internationalen PISA-Konsortium freigegeben. Als repräsentative Untersuchung ermöglicht die PISA-Studie einerseits eine Überprüfung der Stereotype-Threat-Theorie unter normalen Testbedingungen, andererseits aber auch in verschiedenen Ländern.

Diese Arbeit gliedert sich in zehn Kapitel. Im Anschluss an die Einleitung gibt das zweite Kapitel einen Überblick über die Situation von Frauen in der Mathematik und verwandten Bereichen. In knapper Form soll nicht nur die Genderlücke im tertiären Bildungsbereich skizziert werden, sondern auch die Situation im Schulalltag und der Berufsausbildung. Ausgehend von der Tatsache, dass sich nur wenige Mädchen und junge Frauen für eine Ausbildung im mathematisch-naturwissenschaftlichen Bereich entscheiden, wird im dritten Kapitel der aktuelle Forschungsstand dargelegt, inwieweit tatsächlich Geschlechtsunterschiede in der Mathematikleistung bestehen. Im vierten Kapitel werden zunächst die klassischen Erklärungsansätze erörtert, ehe die theoretischen Annahmen und empirischen Befunde zur Stereotype-Threat-Theorie diskutiert werden. Eine theoretische Weiterentwicklung der Stereotype-Threat-Theorie und eine kurze Zwischenbetrachtung runden dieses Kapitel ab und leiten zum empirischen Teil dieser Arbeit über.

Im fünften Kapitel werden zunächst das Design der PISA-Studie und die dabei eingesetzten Messinstrumente vorgestellt, bevor Hypothesen formuliert werden, die im sechsten Kapitel regressionsanalytisch überprüft werden. Im ersten Schritt werden dazu Regressionsmodelle mit den Daten der deutschen Schüler aus der PISA-Studie berechnet, anschließend folgt mit einer Auswahl europäischer Länder der internationale Vergleich. Zwischen den ausgewählten Ländern zeigen sich beachtliche Unterschiede in dem Ausmaß der geschlechtsspezifischen Leistungsdifferenzen, die im Mittelpunkt des folgenden Unterabschnitts stehen. Das siebte Kapitel ergänzt die Analysen durch eine Untersuchung der Leistungsunterschiede zwischen Mädchen und Jungen in den naturwissenschaftlichen Fächern, ehe das achte Kapitel „Biologie, Sozialisation oder Situation?“ die empirischen Befunde zusammenfasst und bewertet.

Das neunte Kapitel behandelt schließlich die Frage, wie geschlechtsspezifischen Leistungsunterschieden in der pädagogischen Praxis begegnet werden können. Eine kritische Diskussion und ein Ausblick auf weitere Forschung schließen diese Arbeit ab.

## **2. Die Situation von Mädchen und Frauen in der Mathematik**

Da die Schulbildung Auswirkungen auf die Arbeitsmarktbeteiligung, die berufliche Mobilität und die Lebensqualität hat, haben die bildungspolitischen Entscheidungsträger Fragen der Geschlechtergleichstellung Priorität eingeräumt, wobei den Benachteiligungen von Frauen besondere Aufmerksamkeit geschenkt wird (OECD, 2004a).

Die geschlechtsspezifischen Unterschiede in der Bildungsbeteiligung haben sich in den vergangenen Jahrzehnten deutlich verändert und zweifellos konnten geschlechtsspezifische Unterschiede bei den formalen Bildungsabschlüssen abgebaut werden. So ist die Wahrscheinlichkeit, dass junge Frauen einen Hochschulabschluss erwerben, heute wesentlich größer als vor 30 Jahren: In 18 von 29 OECD-Ländern mit vergleichbaren Daten ist die Zahl der Frauen im Alter zwischen 25 und 34 Jahren mit einem tertiären Bildungsabschluss mehr als doppelt so hoch wie die der Frauen im Alter zwischen 55 und 64 Jahren. Die Abschlussquoten von Frauen entsprechen in 21 von 27 OECD-Ländern denjenigen der Männer beziehungsweise übersteigen diese noch. Im Durchschnitt aller OECD-Länder sind 55 Prozent aller Absolventen Frauen (OECD, 2004b). Vor dem Hintergrund dieser Entwicklung sind Frauen vielfach als die Gewinnerinnen der Bildungsexpansion bezeichnet worden (Faulstich-Wieland, 2001). Frauen haben zwar große Fortschritte bei der Verringerung ihres historischen Bildungsnachteils erzielt, aber es gibt weiterhin große Unterschiede zwischen den einzelnen Studienbereichen. In den Geistes- und Erziehungswissenschaften sowie in den Bereichen Kunst, Gesundheit und Soziales sind im OECD-Durchschnitt mehr als zwei Drittel der Hochschulabsolventen Frauen, wohingegen ihr Anteil in der Mathematik und verwandten Bereichen ein Drittel oder weniger ausmacht. So beträgt der Frauenanteil unter den Hochschulabsolventen der Studiengänge Mathematik und Informatik im Durchschnitt der OECD-Länder 30 Prozent, und in Belgien, Deutschland, Island, den Niederlanden, Norwegen, Österreich, der Schweiz, der Slowakischen Republik und Ungarn liegt er zwischen neun und 25 Prozent. In den Bereichen Ingenieurwesen, Fertigung und Bauwesen liegt der OECD-Durchschnitt lediglich bei 23 Prozent (OECD, 2004b).

Die Genderlücke in Mathematik und verwandten Fächern lässt sich jedoch nicht erst im tertiären Bildungsbereich beobachten, sondern bereits im Schulalltag. Bereits in der Sekundarstufe I äußern Mädchen ein deutlich geringeres Interesse an Naturwissenschaften und Technik, wo-

bei Biologie eine Ausnahme bildet (Beerman, Heller & Menacher, 1992). Bei Mädchen sind eher Sprachen und weniger mathematisch-naturwissenschaftliche Fächer beliebt, bei den Jungen zeigt sich das gegenläufige Muster (Keller, 1998).

In der gymnasialen Oberstufe fällt mit der Belegung der Leistungskursfächer häufig eine Vorentscheidung über die Studienfachwahl, weshalb im achten Studierendensurvey<sup>1</sup> auch nach der Wahl der Leistungskurse in der Sekundarstufe II gefragt wurde. Weit mehr Studentinnen wählten ein Fach aus dem sprachlich-literarischen Aufgabenfeld, während das mathematisch-naturwissenschaftliche Aufgabengebiet – mit Ausnahme des Fachs Biologie – von den Studenten bevorzugt wurde. Der Einfluss des Geschlechts auf die Leistungskurswahl ist dabei gegenüber 1990 stabil geblieben (Bargel, Ramm & Multrus, 2004), ähnliche geschlechtsspezifische Muster der Leistungskurswahl lassen sich auch in zahlreichen anderen Ländern finden (Beerman et al., 1992).

Die Unterschiede zwischen den Geschlechtern bei der Entscheidung für bestimmte Schul- und Studienfächer bestehen auch bei der Berufsausbildung. Die große Mehrheit der Mädchen und Frauen strebt eine Ausbildung im sozialen Bereich oder im Dienstleistungssektor an, während das mathematisch-naturwissenschaftliche Spektrum von jungen Frauen weitgehend gemieden wird (Hannover, 1992).

Diese geschlechtsspezifischen Präferenzen für Schul- und Studienfächer sowie für bestimmte Ausbildungsrichtungen bleiben nicht ohne Konsequenzen. Sowohl die Leistung als auch die Motivation und die Einstellung gegenüber Mathematik und verwandten Fächern können den künftigen Bildungs- und Berufsweg, beispielsweise die Karriere- und Verdienstaussichten, erheblich beeinflussen. So gilt die Mathematik längst als kritischer Faktor auf dem Arbeitsmarkt (Chipman, 2005), und Paglin und Rufolo (1990) konnten zeigen, dass Einkommensunterschiede zwischen Männern und Frauen zumindest teilweise auf die unterschiedliche mathematische Kompetenz zurückgeführt werden können.

Durch die fortschreitende technische Entwicklung hat sich die Qualifikationsanforderung auf dem Arbeitsmarkt zudem weiter auf ein mathematisches, naturwissenschaftliches und technisches Profil verschoben (Keller, 1998). Dies kann für Frauen ungünstige Folgen auf dem Arbeitsmarkt haben: Wenn Frauen weniger mathematische, naturwissenschaftliche und technische Kompetenzen erwerben als Männer, bleibt ihnen der Zugang zu den sich ausweitenden mathematisch-naturwissenschaftlich-technisch fundierten Berufsfeldern eher verschlossen. Bereits heute ist das Ausbildungsspektrum von Mädchen sehr viel schmäler als das von Jun-

---

<sup>1</sup> Beim Studierendensurvey handelt es sich um eine Langzeitstudie über Studiensituation und studentische Orientierungen an deutschen Universitäten und Fachhochschulen. Die erste Befragung fand im Wintersemester 1982/83 statt, die weiteren im Abstand von zwei bis drei Jahren. Durchgeführt werden die Befragungen von der Arbeitsgruppe Hochschulforschung an der Universität Konstanz, gefördert wird das Projekt vom Bundesministerium für Bildung und Forschung.

gen: 2002 erlernten in der Bundesrepublik Deutschland rund 53 Prozent der jungen Frauen gerade einmal zehn Berufe (Geißler, 2004), zudem sind Vergütung sowie Beschäftigungs- und Aufstiegschancen in Frauenberufen niedriger (Hannover, 1991).

Eine Genderlücke in Mathematik und verwandten Fächern lässt sich von Schule über Studium und Berufsausbildung bis zur eigentlichen Berufstätigkeit feststellen. Als Ursache für den geringen Anteil von Frauen in mathematischen Studiengängen oder mathematisch orientierten Berufsfeldern wird in der Literatur auf geschlechtsspezifische Begabungsunterschiede verwiesen. Gibt es geschlechtsspezifische Leistungsunterschiede in der Mathematik? Diese Frage steht im Mittelpunkt des folgenden Kapitels.

### **3. Gibt es geschlechtsspezifische Leistungsunterschiede in der Mathematik?**

Die Tatsache, dass sich nur wenige Mädchen und junge Frauen für eine Ausbildung beziehungsweise ein Studium im mathematisch-naturwissenschaftlichen Bereich entscheiden, hat in den vergangenen Jahren unzählige Arbeiten veranlasst, das intellektuelle Leistungsvermögen von Mädchen und Jungen durch verschiedene Testmethoden zu erfassen und zu vergleichen. Die Untersuchungen zielten darauf ab, ob der geringere Anteil von Frauen in mathematisch-naturwissenschaftlichen Berufen durch geschlechtsspezifische intellektuelle Fähigkeiten zu erklären ist (Beerman et al., 1992).

Obwohl sich Mädchen und Jungen in der allgemeinen Intelligenz nicht unterscheiden, zeigen sich Unterschiede in spezifischen geistigen Fähigkeiten (Berk, 2005).<sup>2</sup> In einer umfassenden Zusammenstellung der amerikanischen Veröffentlichungen zu geschlechtsspezifischen Leistungsdifferenzen bis zum Jahre 1974 kamen Eleanor E. Maccoby und Carol N. Jacklin zu folgendem Ergebnis: Mädchen haben höhere verbale Fähigkeiten als Jungen, Jungen verfügen über ein besseres räumliches Vorstellungsvermögen und bessere mathematische Fähigkeiten. Die männliche Überlegenheit bei der räumlichen Wahrnehmungsfähigkeit zeigt sich ab der Pubertät und steigt bis zum Erwachsenenalter auf eine Standardabweichung, während in der Kindheit noch keine geschlechtsspezifischen Unterschiede festzustellen sind. Die Leistungsdifferenzen in den mathematischen Fähigkeiten zugunsten der Jungen offenbaren sich ab einem Alter von zwölf Jahren, wobei die Geschlechtsunterschiede geringer sind als bei der räumlichen Wahrnehmungsfähigkeit (Maccoby & Jacklin, 1974).

---

<sup>2</sup> Nach einer von der internationalen Nachrichtenagentur Agence France-Presse (AFP) veröffentlichten Meldung vom 25. August 2005 wollen britische Wissenschaftler in einer Studie mit 100.000 Testpersonen einen höheren Intelligenzquotienten des männlichen Geschlechts belegt haben. In ihrer Untersuchung, die demnächst im „British Journal of Psychology“ veröffentlicht werden soll, wollen Paul Irwing und Richard Lynn herausgefunden haben, dass der Intelligenzquotient von Männern im Alter über 14 Jahren durchschnittlich fünf Punkte höher ist als bei Frauen („Die Welt“, 27. August 2005).

In den 1980er-Jahren wurden die von Maccoby und Jacklin angeführten Studien von Janet Hyde (1981) mit dem Verfahren der Metaanalyse<sup>3</sup> erneut ausgewertet. Auch ihre Ergebnisse weisen auf Geschlechtsunterschiede in den verbalen und mathematischen Leistungen hin, jedoch geringer als von Maccoby und Jacklin angenommen. Bei den quantitativen Fähigkeiten und dem räumlichen Vorstellungsvermögen ermittelte Hyde einen durchschnittlichen d-Wert<sup>4</sup> von +0.43 beziehungsweise +0.45, während die Mädchen mit einem d-Wert von -0.24 bei den verbalen Testaufgaben besser abschnitten. Sie kommt zu dem Schluss, dass „the present analysis shows that gender is a poor predictor of one’s performance on ability tests in any of these areas. Thus, presumably, gender would also be a poor predictor of performance on jobs requiring these abilities” (Hyde, 1981, S. 897).

Mit dem Erscheinen einer zweiten Metaanalyse von Hyde, Fennema und Lamon (1990) schien es dann zunächst auch, als verlöre das Thema Geschlechterdifferenzen in der Mathematik jegliche Bedeutung (Köller & Klieme, 2000). Unter Berücksichtigung von 100 Studien zu mathematischen Fähigkeiten von über 3,1 Millionen Personen fanden die Autorinnen für allgemeine Stichproben eine mittlere Effektstärke von  $d = -0.05$ , das heißt, Mädchen schnitten etwas besser ab als Jungen. Sowohl für das Rechnen ( $d = -0.14$ ) als auch für das Verständnis für mathematische Konzepte ( $d = -0.03$ ) zeigten sich leichte Vorteile für die Mädchen, während die Jungen beim Problemlösen mit einer durchschnittlichen Effektstärke von  $d = +0.08$  besser abschnitten (Hyde et al., 1990).

Eine Differenzierung der Studien nach Publikationsdatum wies auf eine Verringerung der geschlechtsspezifischen Leistungsunterschiede über die vergangenen Jahrzehnte hin: Für Arbeiten, die 1973 und früher veröffentlicht wurden, ermittelten Hyde et al. eine durchschnittliche Effektstärke von +0.31, für Studien, die 1974 und später veröffentlicht wurden, lag die Effektstärke nur noch bei +0.14. Dies wird von den Autorinnen als ein „substantial trend for smaller gender differences in more recent studies“ (Hyde et al., 1990, S. 149) gewertet.

Eine Aufteilung nach dem Alter der getesteten Personen zeigte jedoch einen klaren Trend zu zunehmender Überlegenheit des männlichen Geschlechts bei steigendem Alter. Einen Über-

---

<sup>3</sup> In Metaanalysen werden mehrere Studien zusammengefasst, die sich auf ein ähnliches Merkmal oder eine ähnliche Population beziehen, indem für jede Studie die Effektgröße  $d$  bezüglich des betrachteten Kriteriums bestimmt wird und die Effektgrößen dann gemittelt werden. Dabei erfolgt entweder eine einfache Mittelung oder ein gewichtetes Mittel, in das die Ergebnisse von Studien mit größerer Stichprobe stärker eingehen als die Ergebnisse von Studien mit kleinerer Stichprobe (Asendorpf, 1999).

<sup>4</sup> In dieser Arbeit weisen positive  $d$ -Werte auf eine Differenz zugunsten der Jungen, ein negativer  $d$ -Wert auf einen Leistungsvorsprung der Mädchen hin. Entsprechend der üblichen Praxis werden Effektstärken ( $d$ ) von weniger als 0.2 als schwacher Effekt, eine von 0.5 als mittlerer Effekt und eine von mehr als 0.8 als großer Effekt angesehen (Cohen, 1988). Diese schematische Einschätzung, unabhängig vom Inhalt und von der Fragestellung, kann jedoch nur als grobe Richtlinie betrachtet werden. Wenn nämlich aufgrund des Ziels der Schule gar keine Unterschiede zwischen den Geschlechtern auftreten sollten, ist bereits eine Effektgröße von 0.2 von praktischer Bedeutung (Keller, 1998).

blick über die Ergebnisse gibt die folgende Tabelle, wobei positive d-Werte wieder auf eine Überlegenheit zugunsten der Männer hinweisen.

Alter	alle Studien	Art der Testaufgaben		
		Rechnen	Konzepte	Problemlösen
5 bis 10 Jahre	-0.06 (67)	-0.20 (30)	-0.02 (33)	0.00 (11)
11 bis 14 Jahre	-0.07 (93)	-0.22 (38)	-0.06 (28)	-0.02 (21)
15 bis 18 Jahre	+0.29 (53)	0.00 (12)	+0.07 (9)	+0.29 (10)
19 bis 25 Jahre	+0.41 (31)	keine Angabe	keine Angabe	+0.32 (15)
ab 26 Jahre	+0.59 (9)	keine Angabe	keine Angabe	keine Angabe

Anmerkung: Keine Angabe, da nur zwei oder weniger Effektgrößen vorlagen. Angegeben sind die mittlere Effektgröße d und in Klammern die Zahl der Studien.

Tabelle 3-1: Geschlechtsunterschiede in mathematischen Fähigkeiten (nach Hyde et al., 1990)

Die mathematischen Leistungsunterschiede treten nach der von Hyde et al. (1990) vorgelegten Metaanalyse ab 15 Jahren auf, vorher sind nur minimale Unterschiede zu verzeichnen, wobei die Mädchen sogar etwas besser abschneiden. Ausgeprägter als in den Untersuchungen der Normalvariation sind jedoch die Geschlechtsunterschiede in den Extrembereichen der mathematischen Begabung. In der Gruppe der Hochbegabten zeigte sich ein d-Wert von +0.54, bei der Gruppe der Frühbegabten eine durchschnittliche Effektstärke von +0.41, während die geschlechtsspezifischen Leistungsunterschiede bei Niedrigbegabten mit einem d-Wert von +0.11 deutlich geringer ausfallen.

Wenn also schon keine beziehungsweise nur geringe Mittelwertunterschiede, dann gibt es vielleicht geschlechtsspezifische Verteilungsunterschiede in der mathematischen Fähigkeit? So lassen sich zumindest die Ergebnisse von Camilla P. Benbow (1988) interpretieren, die über einen Zeitraum von 16 Jahren den Scholastic Aptitude Test for Mathematics (SAT-M) in den Vereinigten Staaten ausgewertet hat. Um mathematische Talente rechtzeitig identifizieren und fördern zu können, sollten Kinder im Alter von zwölf bis 13 Jahren, die in den üblichen mathematischen Schulaufgaben besonders gut abschnitten, den SAT-M absolvieren, der eigentlich für ältere Schüler konzipiert ist. Dabei zeigten sich ausgeprägte Vorteile zugunsten der Jungen, die durchschnittlich 30 Punkte besser abschnitten als die Mädchen. Je höher die erzielte Punktzahl, desto größer der geschlechtsspezifische Unterschied: Bei einer Punktzahl von 500 war das Verhältnis zwischen Jungen und Mädchen 2:1, bei 600 bereits 4:1 und bei 700 sogar 12.9:1.

„These differences are most pronounced at the highest levels of mathematical reasoning, they are stable over time, and they are observed in other countries as well.” (Benbow, 1988, S. 169)

Diese geschlechtsspezifischen Differenzen in den Extremwerten werden häufig durch einen Geschlechterunterschied in den räumlichen Fähigkeiten, insbesondere der mentalen Rotation von Gegenständen, zu erklären versucht. So beziehen sich schwierige Mathematikaufgaben entweder direkt auf geometrische Aufgaben, zu deren Lösung ein Manipulieren von Figuren in der Vorstellung hilfreich ist, oder sie lassen sich geometrisch veranschaulichen und dadurch leichter lösen (Asendorpf, 1999), weshalb sich viele Untersuchungen auch den geschlechtsspezifischen Unterschieden im räumlichen Vorstellungsvermögen widmen. So attestierte bereits Maccoby und Jacklin (1974) den Jungen bessere Wahrnehmungsfähigkeiten als den Mädchen, und auch Hyde (1981) ermittelte eine durchschnittliche Effektstärke von +0.45 zugunsten der Jungen. Es scheint aber, dass „geschlechtsspezifische Differenzen im räumlichen Vorstellungsvermögen bis etwa zum Eintritt in das Pubertätsalter nicht oder nur in verschwindend geringem Umfang in Erscheinung treten“ (Maier, 1996, S. 251). Geschlechtsspezifische Unterschiede in unterschiedlichen Komponenten der Raumvorstellung sind jedoch ab einem Alter von zehn Jahren nachgewiesen. So ermittelte die Metaanalyse von Marcia C. Linn und Anne C. Petersen (1985), die insgesamt 172 Studien zu diesem Thema umfasste, eine durchschnittliche Effektstärke von +0.44 bei der räumlichen Wahrnehmung zugunsten der Jungen. Der Leistungsvorsprung zugunsten der Jungen im Bereich der Veranschaulichung wird auf eine Effektstärke von +0.13 beziffert, während die geschlechtsspezifischen Unterschiede bei der Vorstellungsfähigkeit von Rotationen mit einer durchschnittlichen Effektstärke von +0.73 besonders groß sind.

Besonders ausgeprägt scheint der männliche Leistungsvorsprung bei den anspruchsvollsten Komponenten der Raumvorstellung, der räumlichen Orientierung und der Vorstellungsfähigkeit von Rotationen, zu sein. Bei den Aufgaben zur Veranschaulichung und räumlichen Wahrnehmung sind keine oder nur geringe Unterschiede vorhanden (Maier, 1996). Insbesondere verringert sich der männliche Leistungsvorsprung, wenn der Einfluss der Zeit minimiert wird, die Testpersonen also keinen Zeitdruck verspüren (Maier, 1994). Vor diesem Hintergrund scheinen Feststellungen globalen Charakters, die ohne Differenzierung nach Alter oder Teilkomponenten der Raumvorstellung eine klare Überlegenheit des männlichen Geschlechts attestieren, zumindest problematisch.

Einen aktuellen Überblick über geschlechtsspezifische Leistungsunterschiede in der Mathematik im internationalen Vergleich bieten die Programme for International Assessment (PISA) 2000 und 2003, die unter anderem auch die mathematische Grundbildung von 15-jährigen Schülern erfassen.<sup>5</sup> Entgegen des Befunds unbedeutender Geschlechtsunterschiede

---

<sup>5</sup> Bei den empirischen Analysen wird auf den internationalen PISA-Datensatz 2003 zurückgegriffen. Eine ausführlichere Beschreibung der Schulleistungsstudie folgt im fünften Kapitel dieser Arbeit.

bei Stichproben aus der allgemeinen Bevölkerung in der Metaanalyse von Hyde et al. (1990), zeigten sich bei der PISA-Studie 2000 in 15 von 32 Staaten signifikante Leistungsunterschiede zugunsten der Jungen (OECD, 2001); im Durchschnitt der OECD-Länder lag der Leistungsvorsprung der Jungen in Mathematik bei elf Punkten (Stanat & Kunter, 2001).

Eine differenzierte Betrachtungsweise ermöglicht die PISA-Studie 2003, bei der in den 41 teilnehmenden Staaten<sup>6</sup> rund 250.000 Schüler getestet wurden (Prenzel, Drechsel, Carstensen & Ramm, 2004). Da über die Hälfte der Testzeit auf die Mathematik entfiel, kann die PISA-Studie 2003 wesentlich detailliertere Informationen zu den Mathematikleistungen liefern, als das bei PISA 2000 möglich war. Neben der Gesamtpunktzahl macht PISA 2003 auch Angaben zu den mathematischen Inhaltsbereichen Raum und Form, Veränderung und Beziehung, quantitatives Denken und Unsicherheit. Einen Überblick über die geschlechtsspezifischen Leistungsunterschiede in der PISA-Studie 2003 gibt die Abbildung, wobei signifikante Unterschiede auf dem 95-Prozent-Konfidenzniveau durch dunklere Farben gekennzeichnet sind.

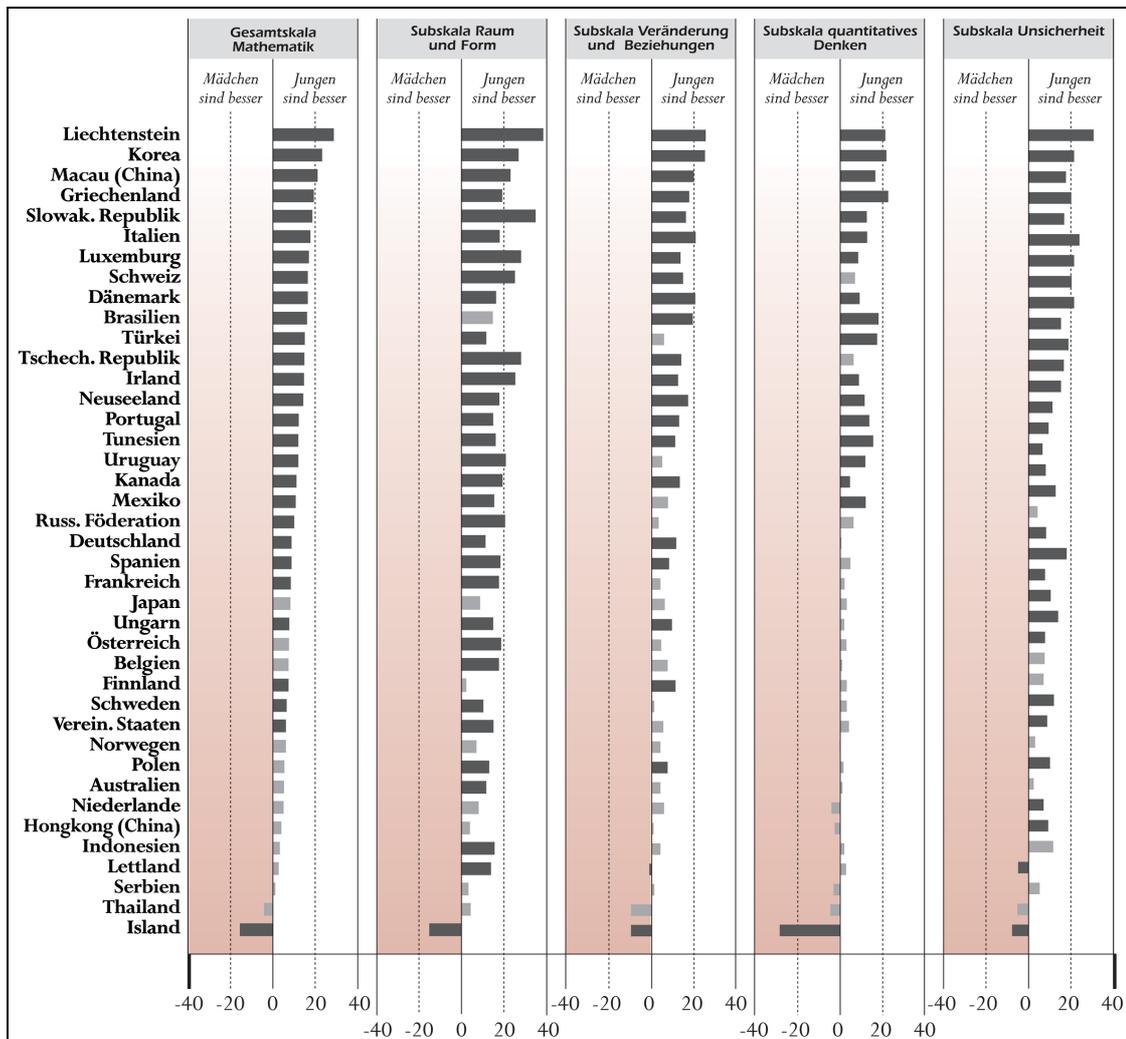


Abbildung 3-1: Geschlechtsspezifische Leistungsunterschiede in der PISA-Studie 2003 in Punkten (nach OECD, 2004a)

<sup>6</sup> Im Vereinigten Königreich war die Beteiligungsquote zu niedrig, um die internationale Vergleichbarkeit zu gewährleisten. Deshalb beziehen sich die folgenden Angaben auf die verbleibenden 40 Länder (OECD, 2004a).

In 27 Ländern zeigen sich bei der Gesamtskala Mathematik Leistungsvorsprünge für die Jungen, in zwölf Ländern ist kein signifikanter Unterschied zu erkennen und in Island zeigt sich ein signifikanter Leistungsvorsprung für die Mädchen. Generell sind die Differenzen zwischen Jungen und Mädchen bei der Mathematikskala Raum und Form am größten. Hier sind mit Ausnahme von Brasilien, Japan, Finnland, Norwegen, der Niederlande, Hongkong, Serbien und Thailand in allen Ländern geschlechtsspezifische Unterschiede zu erkennen. Ebenso gibt es bei der Subskala Unsicherheit bedeutende geschlechtsspezifische Unterschiede, wo in 31 Ländern signifikante Leistungsdifferenzen auftreten. Bei der Skala Veränderung und Beziehungen zeigen sich in 21 Ländern signifikante Leistungsvorsprünge für die Jungen, während bei dem Inhaltsbereich quantitatives Denken Leistungsunterschiede zugunsten der Jungen nur in 17 Ländern zufallskritisch abgesichert sind.

Eine Beurteilung der Größe der geschlechtsspezifischen Leistungsunterschiede ermöglicht wieder die Effektgröße  $d$ . Die Effektstärke beträgt bei der Gesamtskala Mathematik im OECD-Durchschnitt  $+0.11$ , was als kleiner, aber durchaus bedeutsamer Unterschied zwischen Jungen und Mädchen zu bezeichnen ist. In 18 von 40 Ländern liegt der durchschnittliche  $d$ -Wert unter  $0.10$ , in 17 Ländern zwischen  $0.10$  und  $0.20$ , und in fünf Ländern über  $0.20$  (OECD, 2004a).

Bei der Interpretation der Ergebnisse des internationalen Vergleichs ist zu bedenken, dass die Leistungsdifferenzen auch durch geschlechtsspezifische Muster der Bildungsbeteiligung bedingt sein können. Mädchen und Jungen treffen in vielen Ländern eine unterschiedliche Wahl in Bezug auf Schulen, Bildungsgänge und -programme, beispielsweise sind in Deutschland die Mädchen in den leistungsstärkeren Bildungsgängen überrepräsentiert (Stanat & Kunter, 2003). So sind in den meisten Ländern die geschlechtsspezifischen Leistungsunterschiede im mathematischen Bereich innerhalb der Schulen beziehungsweise der Bildungsgänge größer als die insgesamt bestehenden Geschlechterdifferenzen. In Belgien und Deutschland haben die Jungen in der Gesamtpopulation einen Vorsprung von acht beziehungsweise neun Punkten auf der Gesamtskala Mathematik, aber der durchschnittliche Abstand erhöht sich auf 26 beziehungsweise 31 Punkte innerhalb der Schulen. Während die geschlechtsspezifischen Leistungsunterschiede im Globalvergleich also eher gering sind, werden innerhalb der Schulen beziehungsweise Bildungsgängen größere Differenzen beobachtet (OECD, 2004a).

In Übereinstimmung mit den Befunden von Hyde et al. (1990) und Benbow (1988) zeigt sich bei der PISA-Studie 2003, dass die „geschlechtsspezifischen Unterschiede am oberen Ende der Leistungsverteilung in der Regel größer sind“ (OECD, 2004a, S. 108). Im OECD-Durchschnitt erzielen zwar 5,1 Prozent der getesteten Jungen über 668 Punkte, aber nur 2,9 Prozent der Mädchen.

Auch der von Hyde et al. (1990) gefundene Zusammenhang von zunehmender Überlegenheit des männlichen Geschlechts bei steigendem Alter scheint empirisch gesichert. Bei der dritten internationalen Mathematik- und Naturwissenschaftsstudie der IEA (TIMSS) fanden sich bei Schülern der vierten Klasse nur in drei von 16 teilnehmenden OECD-Ländern statistisch signifikante geschlechtsbezogene Unterschiede bei den mathematischen Leistungen, wobei durchgehend die Jungen die besseren Testergebnisse aufwiesen. In der achten Klasse konnte dieselbe Studie bereits in sechs der 16 OECD-Länder signifikante Unterschiede zugunsten der Jungen nachweisen, und für das letzte Schuljahr der Sekundarstufe II fanden sich bis auf zwei Länder große und statistisch abgesicherte Leistungsunterschiede für das männliche Geschlecht (OECD, 2004a).

Insgesamt weisen die vielfältigen Befunde auf eine männliche Überlegenheit in der Mathematik hin, wobei die geschlechtsspezifischen Leistungsunterschiede in der Regel erst ab der Pubertät auftreten und dann insbesondere in der leistungsstarken Gruppe ausgeprägt sind. Die Größe der geschlechtsspezifischen Unterschiede variiert jedoch je nach mathematischem Teilgebiet. So ist beispielsweise der Vorteil der Jungen bei der räumlichen Wahrnehmungsfähigkeit, insbesondere bei der räumlichen Orientierung und der Vorstellungsfähigkeit von Rotationen, besonders stark, während in der Arithmetik oder Algebra deutlich geringere Geschlechterunterschiede zu verzeichnen sind.

Die große Varianz, die bei den geschlechtsspezifischen Unterschieden zwischen den Ländern besteht, beispielsweise zeigten bei der PISA-Studie 2003 die isländischen Mädchen über alle mathematischen Bereiche hinweg bessere Leistungen als die Jungen, kann zumindest als Hinweis gewertet werden, dass die gegenwärtigen Unterschiede in den mathematischen Leistungen von Jungen und Mädchen nicht unabdingbar sind. Darüber hinaus gibt es empirische Evidenz, dass sich die geschlechtsspezifischen Leistungsunterschiede in der Mathematik in den vergangenen Jahrzehnten reduziert haben (Hyde et al., 1990), was wohl weitgehend auf die gestiegene Bildungsbeteiligung der Mädchen in den höheren Bildungsgängen zurückgeführt werden kann.

Dennoch gilt ein (geringer) Leistungsvorsprung der Jungen ab einem Alter von 15 Jahren in der Mathematik als empirisch gesichert. Zur Erklärung von Geschlechterdifferenzen in mathematischen Leistungen wurden eine Vielzahl von Theorien und Hypothesen entwickelt, die im nächsten Abschnitt vorgestellt werden.

#### 4. Erklärungsansätze für Geschlechterdifferenzen in der Mathematik

Die traditionelle Forschung zur Erklärung von geschlechtsspezifischen Leistungsunterschieden in der Mathematik lässt sich grob den Polen Nature und Nurture zuordnen. Auf der Nature-Seite sind Theorien anzusiedeln, die Leistungsdifferenzen zwischen Mädchen und Jungen auf biologische Faktoren zurückführen, während Erklärungsansätze, die der Nurture-Perspektive zuzuordnen sind, die geschlechtsspezifischen Unterschiede in der Mathematik als Ergebnis der Sozialisation betrachten (Stanat & Kunter, 2002). Die beiden Pole Nature und Nurture bestimmen dabei nicht nur die wissenschaftliche Diskussion, sondern zeigen sich auch im Meinungsbild der (Mannheimer) Bürger. Bei einer Telefonumfrage<sup>7</sup> der Universität Mannheim im Frühjahr 2005 zum „Leben in Mannheim“ waren die Befragten aufgefordert, anzugeben, welcher der wichtigste Einflussfaktor bei der Entwicklung der jeweiligen Eigenschaft ist.

	<b>Intelligenz</b>	<b>Persönlichkeit</b>	<b>Aggressivität</b>
Sozialisation <sup>8</sup>	41.51 Prozent	78.29 Prozent	89.10 Prozent
Gene	53.84 Prozent	18.07 Prozent	10.36 Prozent
Gottes Wille	4.65 Prozent	3.64 Prozent	0.54 Prozent
Fallzahl	925	935	927

Tabelle 4-1: Meinungsumfrage zum „Leben in Mannheim“, eigene Berechnungen

Bei kognitiven Fähigkeiten wird eine biologische Ursache von den Bürgern als wichtigster Einflussfaktor eingeschätzt, während die Persönlichkeit und die Aggressivität nach Einschätzung der Mannheimer Bürger stärker von der Sozialisation geprägt werden.

Ob biologische oder sozialisationsbedingte Faktoren für die Ausbildung der Geschlechterdifferenzen in kognitiven Leistungen wichtiger sind, ist Gegenstand einer hitzigen Debatte (Bischof-Köhler, 2002). Die Diskussion ist dabei durch die Grundannahmen geprägt, biologische Faktoren seien überhaupt nicht, sozialisationsbedingte Faktoren hingegen relativ einfach veränderbar. Beide Sichtweisen sind verkürzt. Biologische Faktoren können von der Umwelt überlagert werden, umgekehrt können sozialisationsbedingte Faktoren nicht beliebig beeinflusst werden. Das Hauptproblem biologisch begründeter Geschlechterdifferenzen liegt darin, dass die gesellschaftlich schlechtere Stellung der Frau oft biologisch legitimiert wurde. Ent-

<sup>7</sup> Die Telefonumfrage zum „Leben in Mannheim“ wurde von Studierenden der Universität Mannheim vom 21. Februar bis zum 19. März 2005 durchgeführt und wird vom Lehrstuhl für Methoden der empirischen Sozialforschung und angewandte Soziologie verantwortet.

<sup>8</sup> In der Umfrage hatten die Befragten folgende Antwortmöglichkeiten: Erziehung der Eltern, Gene des Kindes, Freunde des Kindes, äußeres Umfeld und Gottes Wille. Die Antworten zu Erziehung der Eltern, Freunden des Kindes und äußerer Umwelt wurden unter dem Begriff „Sozialisation“ zusammengefasst.

sprechende Argumentationen zu Beginn des 20. Jahrhunderts waren zum Beispiel: Weil Frauen ein geringeres Gehirngewicht haben, seien sie dümmer und damit für verantwortungsvolle gesellschaftliche Positionen ungeeignet (Keller, 1998). Biologische Begründungen für geschlechtsspezifische Unterschiede in kognitiven Leistungen geraten deshalb schnell in den Verdacht, die Benachteiligung der Mädchen und Frauen zu zementieren (Alfermann, 1996). Wenn aber die Geschlechterdifferenzen in mathematischen Leistungen aufgrund eines undifferenzierten Biologieverständnisses erklärt werden, dann besteht die Gefahr, Unterschiede als resistent gegenüber Veränderungen zu betrachten. Geschlechterunterschiede könnten so leichter akzeptiert werden. Mädchen würden in der Schule möglicherweise Gefahr laufen, in der Mathematik und in den Naturwissenschaften weniger gefördert zu werden (Keller, 1998). Die aktuelle Forschung geht grundsätzlich von einer Interaktion von Biologie und Sozialisation aus (Stanat & Kunter, 2002), ein entsprechend integrativer Ansatz wurde von Halpern (2000) vorgelegt.

Neben den biologischen und sozialisationstheoretischen Erklärungen hat sich in den vergangenen Jahren ein neuer Erklärungsansatz etabliert, der dadurch gekennzeichnet ist, dass situative Merkmale der Testdurchführung als Einflussgröße berücksichtigt werden müssen. Dieser Ansatz, die so genannte Stereotype-Threat-Theorie, nimmt an, dass Mitglieder von Gruppen, über die ein negatives Stereotyp in Bezug auf ihre Kompetenz existiert, allein durch das Wissen um die Existenz dieses Vorurteils in ihren Leistungen negativ beeinträchtigt werden können. Danach werden in Situationen, in denen sich die Person der Gefahr ausgesetzt sieht, das Stereotyp zu bestätigen, Prozesse ausgelöst, die zu einer Leistungsminderung führen.

In diesem Kapitel werden zunächst die biologischen und sozialisationstheoretischen Erklärungsansätze erörtert, ehe der integrative Ansatz von Halpern (2000) diskutiert wird. Anschließend stehen die theoretischen Annahmen und empirischen Befunde zum Einfluss von Stereotypen in der Testsituation im Vordergrund, bevor eine theoretische Weiterentwicklung des Stereotype-Threat-Ansatzes vorgestellt wird. Eine kurze Zwischenbetrachtung rundet den Abschnitt ab und leitet zum empirischen Teil dieser Arbeit über.

#### **4.1 Biologische Erklärungsansätze**

Die Theorien, die Geschlechterunterschiede in der Mathematikleistung – insbesondere in der räumlichen Wahrnehmungsfähigkeit – auf biologische Faktoren zurückführen, lassen sich drei Bereichen zuordnen: den genetischen, hormonellen und neuropsychologischen Erklärungsansätzen (Halpern, 2000). Dabei dürfen die Gebiete nicht als isoliert voneinander angesehen werden, vielmehr besteht eine Vielzahl von Wechselbeziehungen zwischen diesen Bereichen (Maier, 1994).

„Like any division in biology, these are not separate systems; genes and hormones presumably operate on behavior through some neurological mechanism, and differences in sex hormones are dependent on genetically coded information.” (Halpern, 2000, S. 137)

Zwischen den einzelnen biologischen Erklärungsansätzen bestehen also interessante und komplizierte, wenn auch vielfach auf hypothetischen Annahmen beruhende Zusammenhänge (Maier, 1996).

#### **4.1.1 Genetische Erklärungen**

Genetische Einflüsse für Geschlechterdifferenzen in der Mathematikleistung beziehungsweise in der räumlichen Wahrnehmungsfähigkeit wurden lange Zeit kontrovers diskutiert. So stellten Bock und Kolakowski (1973) die X-Kopplungshypothese auf, nach der ein rezessives Gen auf dem X-Chromosom mit räumlichen Fähigkeiten verknüpft sei.

Die Chromosomen sind Hauptträger der genetischen Information. 22 der menschlichen Chromosomenpaare sind passende Paare, das 23. Paar repräsentiert die Geschlechtschromosomen. Beim weiblichen Geschlecht liegen zwei X-Chromosomen vor, beim männlichen Geschlecht ein X- und ein Y-Chromosom. Das Y-Chromosom enthält nur wenige funktionsfähige Gene, die für die Festlegung des männlichen Geschlechts verantwortlich sind. Auf dem X-Chromosom gibt es dagegen mehrere hundert Gene. Man bezeichnet sie, da sie auf dem Geschlechtschromosom liegen, als geschlechtsgekoppelt.

Eine hohe räumliche Wahrnehmungsfähigkeit ist nach der X-Kopplungshypothese durch ein geschlechtsgekoppeltes rezessives Gen verursacht. Da Frauen zwei X-Chromosomensätze besitzen, aber nach den Grundlagen dominant-rezessiver Vererbung rezessive Erbanlagen nur dann auftreten, wenn das entsprechende Merkmal in beiden Chromosomensätzen enthalten ist, ist die Wahrscheinlichkeit für eine hohe räumliche Wahrnehmungsfähigkeit bei Frauen geringer als bei Männern (Halpern, 2000). Schließlich unterliegt bei Frauen jedes rezessive Gen auf einem X-Chromosom der Möglichkeit, von einem dominanten Gen auf dem anderen X-Chromosom unterdrückt zu werden (Berk, 2005). Bei Männern stellt sich die Situation anders dar: Sie haben nur einen X-Chromosomensatz und müssen entsprechend das rezessive Gen nur einmal besitzen. Deshalb sei die Wahrscheinlichkeit für ein besseres räumliches Vorstellungsvermögen für Männer größer. Bei Frauen sind jedoch zwei dieser rezessiven Gene nötig, um eine höhere räumliche Wahrnehmungsfähigkeit zu haben (Menacher, 1994).

Um diese These zu verifizieren, untersuchte man unter anderem Zusammenhänge bei räumlichen Wahrnehmungsaufgaben innerhalb einer Familie. Nach der X-Kopplungsthese wären höhere Korrelationen zwischen Müttern und Söhnen als zwischen Vätern und Söhnen zu er-

warten, da das betreffende X-Chromosom von mütterlicher Seite stammt. Frühere Studien wiesen in der Tat auf solche Zusammenhänge hin (Halpern, 2000), jüngere Analysen mit sehr großen Stichproben konnten diese Hypothese jedoch nicht bestätigen (Beerman et al., 1992), so dass aus heutiger Sicht die X-Kopplungsthese als widerlegt angesehen werden kann (Mayer, 1996).

#### **4.1.2 Hormonelle Erklärungen**

Nach Doreen Kimura (1992) spiegeln kognitive Unterschiede zwischen den Geschlechtern hormonelle Einflüsse auf die Gehirnentwicklung wider. Dabei geht es um die Annahme, dass der pränatale Hormonspiegel kognitive Fähigkeiten in geschlechtstypischer Richtung beeinflusst und auch intraindividuelle Leistungsunterschiede bei räumlich-visuellen Aufgaben mit normalen Schwankungen im Hormonhaushalt zusammenhängen (Stanat & Kunter, 2002). Die Mehrzahl der wissenschaftlichen Befunde zeige, dass „der Feinbau des Gehirns bereits so früh von Sexualhormonen beeinflusst wird, daß die Umwelt von Geburt an – und auch schon vorher – bei Mädchen und Jungen auf unterschiedlich verschaltete Gehirne einwirkt“ (Kimura, 1992, S. 104).

Wie aber entstehen solch unterschiedlich verschaltete Gehirne, wenn doch – mit Ausnahme der Geschlechtschromosomen – alle Menschen die gleiche genetische Basis haben? Der menschliche Embryo ist zunächst bisexuell angelegt, das heißt, er besitzt Vorläufer für weibliche (Müller'scher Gang) und männliche Geschlechtsorgane (Wolff'scher Gang). In einer frühen Embryonalphase bestimmen die Wirkungen von Sexualhormonen die Geschlechtsdifferenzierung. Das Geschlecht der inneren Geschlechtsorgane eines Fötus wird dabei durch die Anwesenheit oder das Fehlen von Hormonen bestimmt, die von den Hoden gebildet werden. Wenn die Hormone fehlen, entwickelt sich der Müller'sche Gang, sind die Hormone präsent, entwickelt sich der Wolff'sche Gang. Dabei bilden die Hoden zwei Arten von Hormonen: Testosteron, das wichtigste Androgen des männlichen Geschlechts, stimuliert die Entwicklung des Wolff'schen Gangs und das Anti-Müller'sche Hormon unterdrückt die Entwicklung des Müller'schen Gangs (Carlson, 2004).

Die Geschlechtshormone bewirken jedoch nicht nur die Ausprägung männlicher Geschlechtsorgane, sondern leiten auch Differenzierungen im Gehirn ein (Kimura, 1992). Wenn männliche Hormone durch den Fötus strömen, dann entwickelt sich ein männliches Gehirn; bleibt die Androgenschwemme aus, reift ein weibliches Denkorgan heran. Der englische Psychologe Simon Baron-Cohen differenziert nach S- und E-Gehirnen:

„Männer denken in Systemen, Frauenköpfe dagegen anders. Sie erfassen die Welt mit Hilfe der Empathie, also der Kunst, sich in andere hineinzusetzen. Sie tun dies,

weil ihre Gehirne bereits im Mutterleib unterschiedlich programmiert sind. Das typisch männliche Gehirn bezeichne ich deshalb als S-Gehirn, das typisch weibliche als E-Gehirn.“ (Baron-Cohen, 2003, S. 33)

Während das Gehirn von Frauen demnach von Natur aus auf Einfühlung ausgerichtet ist, ist dem Gehirn von Männern die Gabe angeboren, die Welt systematisch und analysierend begreifen zu wollen (Baron-Cohen, 2004).

Die Einwirkung von Sexualhormonen in einer frühen Lebensphase scheint die Organisation des Gehirns auf irreversible Weise zu beeinflussen, das Verabreichen derselben Hormone in einer späteren Entwicklungsphase hat keinen solchen Effekt. Die Wirkung der Sexualhormone scheint dabei auch das menschliche Verhalten zu betreffen, in dem sich die Geschlechter unterscheiden – die Art des Problemlösens ebenso wie das räumliche Vorstellungsvermögen (Kimura, 1992).

Schon aus ethischen Gründen verbietet sich eine experimentelle Variation des Einflusses von Sexualhormonen auf das menschliche Verhalten. Die Forscher beziehen sich deshalb auf mögliche Parallelen zu anderen Spezies sowie auf auftretende Ausnahmen von der Norm. So ist bei Mädchen, die im Mutterleib oder als Neugeborene einem Übermaß an Androgen ausgesetzt waren, das räumliche Vorstellungsvermögen besser ausgeprägt als bei Mädchen, die sich normal entwickelt haben (Springer & Deutsch, 1998). Je höher der Androgenspiegel, desto besser das räumliche Vorstellungsvermögen? In Untersuchungen fanden sich vielmehr Hinweise auf einen nichtlinearen Zusammenhang.

„Männer mit wenig Testosteron waren ihren Geschlechtsgenossen mit viel Testosteron überlegen, während bei den Frauen mehr Testosteron mit besseren Leistungen korreliert war. Solche Befunde lassen vermuten, daß es sozusagen einen optimalen Androgenspiegel gibt, bei dem das räumliche Vorstellungsvermögen am besten ist; er müßte dann etwa im unteren Teil des für Männer typischen Streubereichs liegen.“ (Kimura, 1992, S. 110)

Insgesamt deuten die Ergebnisse darauf hin, dass sowohl besonders feminine Frauen (geringer Testosteronwert) als auch besonders maskuline Männer (hoher Testosteronwert) in Raumvorstellungstests relativ schlecht abschneiden (Maier, 1996).

Spielt bei dem Erklärungsmodell von Kimura das männliche Sexualhormon Testosteron eine entscheidende Rolle, hat der Mediziner Nyborg eine Theorie entwickelt, die das räumliche Vorstellungsvermögen mit den weiblichen Sexualhormonen in Zusammenhang bringt (Halpern, 2000). Bei Frauen, bei denen aufgrund einer genetischen Anomalie die Eierstöcke Östradiol nur in unzureichender Menge produzieren, stellte er fest, dass sie extrem niedrige

Leistungen auf dem räumlich-visuellen Sektor erbrachten, die sich aber nach einer Östrogenbehandlung drastisch verbesserten. Nach Nyborg produzieren die Eierstöcke Östradiol normalerweise in einer Menge, die über dem Optimum für gute Leistungen liegt. Bei Männern tritt das Hormon nur in geringer Konzentration auf, die eher der Menge entspricht, die eine optimale Entfaltung räumlicher Fähigkeiten begünstigt (Bischof-Köhler, 2002). Sein „General Trait Covariance-Androgen/Estrogen Balance“-Modell postuliert demnach einen umgekehrt u-förmigen Zusammenhang: Sowohl eine zu geringe als auch eine zu hohe Menge von Östradiol hängt negativ mit hohen räumlichen Fähigkeiten zusammen, während ein mittleres Niveau von Östradiol mit guten räumlichen Fähigkeiten korrespondiert (Halpern, 2000).

Nicht nur der pränatale Hormonspiegel kann sich auf das Leistungsvermögen bei kognitiven Aufgaben auswirken, Leistungsunterschiede können darüber hinaus mit normalen Schwankungen im Hormonhaushalt zusammenhängen. Tägliche Schwankungen des Testosteronspiegels haben beispielsweise bei Männern einen Effekt auf räumliche Fähigkeiten. So schneiden Männer bei räumlichen Aufgaben morgens, wenn der Testosteronspiegel am höchsten ist, schlechter ab als später am Tag (Springer & Deutsch, 1998). Kimura berichtet darüber hinaus von jahreszeitlichen Schwankungen der raumbezogenen Fähigkeiten bei Männern. Die Leistungen seien im Frühjahr verbessert, wenn der Testosteronspiegel niedriger ist (Kimura, 1992). Bei Frauen ändert sich das Leistungsvermögen im Laufe des Menstruationszyklus. In der Zyklusmitte, wenn die Östrogene ihr höchstes Niveau erreicht haben, erzielen sie besonders gute Resultate bei sprachlicher Ausdrucksfähigkeit, während das Leistungsniveau bei räumlichen Aufgaben herabgesetzt ist. Während der Zyklusphase, die durch niedrigere Hormonwerte gekennzeichnet ist, zeigen sie dagegen bessere räumliche Leistungen (Springer & Deutsch, 1998).

Eine Erklärung für den Zusammenhang zwischen dem pränatalen Hormonspiegel und geschlechtsspezifischen Leistungsunterschieden im räumlichen Vorstellungsvermögen liefern Norman Geschwind und Albert M. Galaburda (1987), deren Ansatz im nächsten Abschnitt vorgestellt wird.

#### **4.1.3 Neuropsychologische Erklärungen**

Norman Geschwind und Albert M. Galaburda (1987) entwickelten eine Theorie, wonach sich der pränatale und perinatale Einfluss des männlichen Sexualhormons Testosteron verlangsamend auf das Wachstum der linken Gehirnhälfte auswirkt, wodurch sich die rechte Hemisphäre, relativ gesehen, schneller entwickeln kann und infolgedessen beim männlichen Geschlecht eine dominierende Rolle übernimmt (Halpern, Wai & Saw, 2005). Die durch den Androgeneinfluss verursachte Lateralisierung ist dabei nicht nur für Linkshändigkeit oder eine Anfäl-

ligkeit für Immunerkrankungen verantwortlich (Springer & Deutsch, 1998), sondern steigert auch das funktionelle Vermögen der rechten Hemisphäre (Kimura, 1992).

Was ist aber überhaupt mit Lateralisierung gemeint? Die Lateralisierung beschreibt die Spezialisierung der Gehirnhälften. Auch wenn das Gehirn meist als einheitliche Struktur betrachtet wird, ist es dennoch geteilt. Eine tiefe Furche, auch Balken oder Corpus callosum genannt, teilt das Großhirn in zwei Hälften. Trotz der äußerlichen Symmetrie ist es mittlerweile erwiesen, dass das linke und das rechte Gehirn weder in ihren Fähigkeiten noch in ihrer Organisation identisch sind: So steuert die rechte Hemisphäre die Bewegungen der linken Körperseite; die linke Hemisphäre ist für die rechte Körperseite zuständig. Insbesondere die komplexen geistigen Funktionen und Verhaltensweisen des Menschen scheinen zwischen den beiden Gehirnhälften asymmetrisch verteilt zu sein, so auch die Fähigkeit, Sprache zu erzeugen und zu verstehen, oder die Fähigkeit, komplexe räumliche Muster zu verarbeiten (Springer & Deutsch, 1998). Und: Das männliche Gehirn weist eine größere Asymmetrie als das weibliche Denkorgan auf, es ist durch die pränatale Androgenschwemme stärker spezialisiert (Kimura, 1992).

Wenn also die Gehirnhälften von Männern und Frauen unterschiedlich organisiert sind, dann sollte es auch Hinweise geben, dass sich Hirnschädigungen bei beiden Geschlechtern unterschiedlich auswirken. In der Tat hat die Schädigung einer Gehirnhälfte für Männer und Frauen unterschiedliche Folgen. Herbert Lansdell von den National Institutes of Health in den Vereinigten Staaten untersuchte, wie sich bei Patienten beiderlei Geschlechts die Entfernung eines Teils des Temporallappens einer Hemisphäre auswirkte. Er erwartete nach einem Eingriff an der rechten Hirnhälfte ein größeres Leistungsdefizit bei visuell-räumlichen Aufgaben und nach einer Operation an der linken Hemisphäre ein geringeres Leistungsvermögen bei sprachlichen Aufgaben. Die Vorhersagen wurden allerdings nur bei männlichen Patienten bestätigt. Sein Fazit: Die strenge Spezialisierung der Gehirnhälften liegt nur bei Männern vor, während sich die sprachlichen und visuell-räumlichen Leistungen bei Frauen in beiden Hirnhälften überschneiden (Richter, 1996).

Demnach sind bei der Ausübung verbaler Tätigkeiten beim weiblichen Geschlecht beide Hemisphären beteiligt, während bei Jungen und Männern die Spezialisierung der Gehirnhälften ausgeprägter ist, so dass verbale Aktivitäten vor allem in der linken, räumlich-visuelle Aktivitäten vor allem in der rechten Hemisphäre lokalisiert sind (Stanat & Kunter, 2002). Die mit den verschiedensten Untersuchungsmethoden erzielten Ergebnisse weisen alle in dieselbe Richtung: Frauen sind offenbar weniger stark lateralisiert als Männer (Springer & Deutsch, 1998).

Widersprüchlich bleibt vorerst, dass sowohl eine stärkere Lateralisierung räumlicher Fähigkeiten bei Männern als auch gleichzeitig eine eher bilaterale Repräsentation verbaler Fähigkeiten bei Frauen für die jeweils besseren Leistungen verantwortlich sein sollen (Maier, 1994). Möglicherweise bestehen für verschiedene Aufgaben unterschiedliche Beziehungen zwischen Lateralisierung und Leistungsfähigkeit, über den Zusammenhang zwischen Lateralisierung und Fähigkeitsgrad kann derzeit jedoch nur spekuliert werden (Springer & Deutsch, 1998).

#### **4.1.4 Ursprung und Konsequenzen der Geschlechtsunterschiede**

Die Befunde aus verschiedenen Untersuchungen weisen darauf hin, dass die Gehirne von Männern und Frauen bereits von einer sehr frühen Entwicklungsphase an nach unterschiedlichen Prinzipien organisiert sind (Kimura, 1992). Als Erklärung für diese unterschiedliche Entwicklung wird vielfach eine evolutionäre Grundlage für Geschlechtsunterschiede bei kognitiven Fähigkeiten postuliert (Springer & Deutsch, 1998). Nach der Jäger-Sammler-Theorie von Silverman und Eals (1992) weisen Männer überlegene Fähigkeiten bei räumlichen Aufgaben auf, die eine erfolgreiche Jagd vereinfacht haben könnten. So sei das Aufspüren und Erlegen von Tieren mit anderen räumlichen Aufgaben verbunden als das Suchen nach essbaren Pflanzen. Da darüber hinaus die Jagd die Männer oftmals weit vom Heimatlager wegführte, wurden von der Selektion Jäger bevorzugt, die den Rückweg fanden, ohne sich zu verirren (Buss, 2004). Frauen hingegen waren nach dieser Theorie primär Sammlerinnen, suchten Nahrung in der Nähe der Unterkunft und entwickelten eine größere Sensibilität für kleine Veränderungen bei Kindern und in der direkten familiären Umwelt. Aus evolutionärer Perspektive sollten Frauen über ein besseres Gedächtnis für Standorte von Objekten als eine Adaption an das Sammeln verfügen, während der Selektionsdruck bei Männern eher auf Navigationsfähigkeiten lag. Für diesen Ansatz sprechen die besonders bei der räumlichen Orientierung und der Vorstellungsfähigkeit von Rotationen ausgeprägten Geschlechtsunterschiede zugunsten der Männer (Maier, 1996), während Frauen die Männer bei räumlichen Aufgaben, in denen es um ein Gedächtnis für Standorte und Anordnung von Objekten geht, übertreffen (Silverman & Philips, 1998). Auch die Überlegenheit von Frauen bei vielen verbalen Aufgaben wird aus evolutionärer Perspektive verständlich, sofern man der Argumentation folgt, dass Sprache im Zusammenhang mit ihrer Rolle als Nahrungssammlerin oder im Rahmen der Kinderbetreuung zu einem wertvollen Kommunikationswerkzeug wurde (Springer & Deutsch, 1998). Doch hier wendet Doris Bischof-Köhler ein:

„Plausibel ist das nicht. Sprachliche Kommunikation ist eine zentrale evolutionäre Erzungenschaft des Menschen, es leuchtet nicht ein, warum sie für die Männer weniger

bedeutsam sein sollte, vor allem wenn man berücksichtigt, wie gerne sie reden.“ (Bischof-Köhler, 2002, S. 245)

In der Tat stellt gerade die Kommunikation einen Meilenstein der kooperativen Jagd dar, die eine selbstlose Zusammenarbeit und hochflexible Feinabstimmung der Beteiligten voraussetzte. Dadurch gewannen kommunikative Fähigkeiten oder auch die Entwicklung von Sprach- und Artikulationsvermögen hohe adaptive Vorteile. Kommunikation und Sprache wurden sicherlich auch über gewisse Vorformen der Argumentation über gerechte Anteile beim Verteilen der Beute, bei der Mitteilung über die Jagdchancen oder auch durch Berichte über den gerade erlebten Erfolg bei der Rückkehr gefördert (Esser, 1999a). So scheinen zwar die evolutionären Schlussfolgerungen, wonach Männer und Frauen unterschiedliche räumliche Spezialisierungen entwickelt haben, eine, die das effektive Sammeln vereinfacht, und eine, die effektives Jagen begünstigt, einleuchtend (Buss, 2004), ob sich jedoch im Sprachvermögen ein Selektionsvorteil finden lässt, der spezifisch beim weiblichen Geschlecht wirksam wird, bleibt widersprüchlich.

Ohnehin scheinen die biologisch bedingten Vorteile in den spezifischen räumlichen Fähigkeiten nicht sonderlich ausgeprägt. Sie sind gering und können daher leicht durch die individuelle Variabilität oder andere nicht kontrollierte Faktoren überdeckt werden (Springer & Deutsch, 1998). So liegen beispielsweise bei Eskimos keine wesentlichen geschlechtsspezifischen Differenzen in der räumlichen Wahrnehmungsfähigkeit vor. Als mögliche Erklärung wird dabei unter anderem auf die Schwierigkeit verwiesen, sich in der Arktis zu orientieren. Die Raumvorstellung besitzt demnach sowohl für weibliche als auch für männliche Eskimos eine existenzielle Bedeutung (Maier, 1996).

Biologische Erklärungsansätze scheinen nicht ohne die Berücksichtigung von Umweltfaktoren auszukommen. So stellt zwar der pränatale Hormoneinfluss eine wichtige Größe für die Entwicklung des menschlichen Gehirns dar, aber Leistungsunterschiede scheinen auch mit normalen Schwankungen im Hormonhaushalt zusammenzuhängen. Schwankungen im Hormonhaushalt sind aber durch äußere Einflüsse, beispielsweise Stress, Klima, Ernährung oder generell psychische Belastungen, beeinflussbar (Alfermann, 1996).

Darüber hinaus zielen sowohl hormonelle als auch neuropsychologische Theorien weitgehend darauf ab, geschlechtsspezifische Unterschiede in der räumlichen Wahrnehmung zu erklären. Damit ist jedoch noch kein zwingender Zusammenhang zwischen den geschlechtsspezifischen Leistungsunterschieden in der räumlichen Wahrnehmungsfähigkeit und der Mathematikleistung postuliert. So ist die empirische Evidenz der so genannten Spatial Mediation Hypothesis, wonach sich geschlechtsspezifische Unterschiede im mathematischen Leistungsvermögen letztendlich nur auf Differenzen im räumlichen Vorstellungsvermögen zurückfüh-

ren lassen, heftig umstritten. In einer Metaanalyse kann die Spatial Mediation Hypothesis bei keiner der untersuchten mathematischen Anforderungen die Leistungsvorsprünge von Männern vollständig erklären (Klieme, 1986), und gutes räumliches Vorstellungsvermögen bedingt nicht notwendigerweise auch gute Leistungen in Mathematik (Menacher, 1994). Zudem haben sich die Geschlechtsunterschiede zumindest in Teilbereichen der räumlichen Wahrnehmungsfähigkeit verringert (Bischof-Köhler, 2002), und deshalb gebe es keine empirische Evidenz, wonach die Fähigkeit zur räumlichen Vorstellung wesentlich für eine erfolgreiche Beschäftigung mit Mathematik, Technik und Naturwissenschaften sei (Beerman et al., 1992). Entgegen dieser Argumentation werden geschlechtsspezifische Unterschiede im räumlichen Vorstellungsvermögen vielfach als Schlüssel zum besseren Verständnis der Geschlechterdifferenzen in Mathematikleistungen bezeichnet.

„In fact, evidence has begun to accumulate that shows a connection between gender differences in mathematics achievement and gender differences in spatial skills.“ (Nuttall, Casey & Pezaris, 2005, S. 121)

Ungeachtet dieser Widersprüche scheinen biologische Erklärungsansätze nicht ausreichend zu sein, um geschlechtsspezifische Leistungsdifferenzen in der Mathematik vollständig zu erklären (Richter, 1996), so dass ein umfassender Erklärungsansatz neben biologischen auch soziale und situative Faktoren berücksichtigen muss.

#### **4.2 Sozialisationstheoretische Erklärungsansätze**

Im vorigen Abschnitt wurde deutlich, dass biologische Erklärungsmodelle allein nicht ausreichen, um Geschlechterdifferenzen in kognitiven Leistungen zu erklären. Dafür spricht einerseits die Tatsache, dass die geschlechtsspezifischen Leistungsunterschiede in der Mathematik erst im Alter von 15 Jahren auftreten, und andererseits der Befund, dass sich die Leistungsdifferenzen in den vergangenen Jahrzehnten deutlich verringert haben. Die psychosozialen Erklärungsmodelle sehen die Geschlechterdifferenzen in den mathematischen Leistungen deshalb als Ergebnis von Umwelteinflüssen und verweisen auf verschiedene Sozialisationseffekte. Im Sozialisationsprozess werde der Mensch mit Einstellungen und Verhaltensweisen der Familie, seiner Freunde, der Schule und den Medien konfrontiert, und diese Erfahrungen seien für Jungen und Mädchen unterschiedlich (Beerman et al., 1992). Dies – und die Zuordnung bestimmter Tätigkeiten als männliche oder weibliche Domäne – habe zur Folge, dass sich Kinder, Jugendliche und Erwachsene bevorzugt in geschlechtsspezifischen Bereichen engagieren und Lernzuwächse erzielen (Stanat & Kunter, 2003). Ganz allgemein bezeichnet Sozialisation den Prozess der Eingliederung eines Menschen in die Gesellschaft.

„Unter Sozialisation wird jener soziale Prozeß verstanden, über den die Kultur einer Gesellschaft in die Identität der neuen Mitglieder – Kinder oder Fremde – vermittelt wird.“ (Esser, 2001, S. 371)

Gegenüber dem Begriff der Erziehung ist Sozialisation als Oberbegriff aufzufassen. Während Erziehung die gezielte Beeinflussung der Person zum Zweck der Vermittlung von Kenntnissen, Wertorientierungen, Verhaltensweisen und Fertigkeiten bezeichnet (Hillmann, 1994), umschreibt Sozialisation die Gesamtheit aller – gezielten und unbeabsichtigten – Vorgänge, in deren Verlauf die sozialen Regeln einer Gesellschaft vermittelt werden (Esser, 2001).

Während das Geschlecht ein biologisches Faktum ist, stellt die Sozialisation keine unabänderliche Größe dar. So haben die Studien der amerikanischen Ethnologin Margaret Mead gezeigt, dass nicht nur soziale Gewohnheiten, Bräuche und Sitten, sondern auch das Temperament und das geschlechtsspezifische Verhalten jedes einzelnen Menschen von der jeweiligen Kultur geprägt sind. Gemessen an der westlichen Kultur waren bei einem von Margaret Mead untersuchten Stamm, den Tchambuli, die Geschlechterrollen von Mann und Frau nahezu vertauscht (Henecka, 1997). Mit ihren Forschungen wollte Mead den Nachweis erbringen, dass der Mensch außerordentlich anpassungsfähig ist und die Entwicklung der menschlichen Persönlichkeit in starkem Maße von der Eigenart der jeweiligen Kultur abhängt. Insbesondere das Verhalten von Mann und Frau hängt entscheidend von soziokulturell unterschiedlich geprägten Rollenmustern ab (Hillmann, 1994), weshalb sich der nächste Abschnitt mit der geschlechtsspezifischen Sozialisation auseinandersetzt.

#### **4.2.1 Geschlechtsspezifische Sozialisation**

Junge oder Mädchen? Das ist immer noch eine der ersten Fragen nach der Geburt eines Kindes und das biologische Geschlecht ist die vermutlich bedeutsamste und in jeder Kultur vorkommende Kategorie (Lewontin, Rose & Kamin, 1988). Es ist nicht nur äußerlich gut wahrnehmbar und unterscheidbar, sondern es ist auch eine zentrale Kategorie in der Selbst- und Fremdwahrnehmung (Alfermann, 1996). Welche Konsequenzen sind mit der Tatsache verbunden, als Junge oder als Mädchen geboren zu werden? Die Antwort gibt Helga Bilden:

„Schon gegenüber Säuglingen gibt es Erwartungen, Verhaltensinterpretationen und Interaktionsformen, die mit dem angenommenen Geschlecht des Kindes variieren.“ (Bilden, 2002, S. 281)

So schätzten bei einer amerikanischen Untersuchung 24 Stunden nach der Geburt Eltern ihr Kind wesentlich größer ein, wenn es ein Junge war, obwohl sich Mädchen und Jungen weder in ihrer Größe noch in ihrem Gewicht unterschieden (Brehm, Kassin & Fein, 1999). Eltern

verfügen also über ein ausgeprägtes Geschlechtsstereotyp, ein kulturell geprägtes Meinungssystem über Eigenarten und Fähigkeiten der beiden Geschlechter, das ihre Erwartungen prägt und dadurch die Wahrnehmungen beeinflusst (Asendorpf, 1999). In den westlichen Industriegesellschaften gehören zum männlichen Stereotyp beispielsweise die Eigenschaften ehrgeizig, leistungsorientiert, selbstsicher und mutig, zum weiblichen Stereotyp gehören unter anderem die Eigenschaften ängstlich, sozial orientiert, schüchtern, schwach und warmherzig (Trautner, 1994). Aber über diese Geschlechtsstereotype verfügen nicht nur Eltern. In einer anderen Studie waren junge Psychologiestudierende aufgefordert, ihre Eindrücke von einem neun Monate alten Baby zu schildern und dabei körperliche Merkmale, Verhalten und Persönlichkeitseigenschaften zu berücksichtigen. Der Hälfte der Studierenden wurde gesagt, das Baby trage den Namen Keith (männlicher Name), der anderen Gruppe wurde mitgeteilt, das Baby heiße Karen (weiblicher Name). Die Studierenden beurteilten Keith als sportlicher, lauter und aktiver als Karen (Zimbardo & Gerrig, 1999). Die Etikettierung als Junge oder Mädchen führte zu sehr unterschiedlichen Beschreibungen des jeweils identischen Verhaltens. Das Geschlecht löst also bewusst und/oder unbewusst bestimmte Erwartungen, Deutungsmuster und Reaktionsstendenzen aus.

Auch daran, wie Kinder gekleidet werden, wie ihr Zimmer ausgestattet ist, welche Spielsachen oder Bücher sie erhalten, werden geschlechtsspezifische Erwartungen sichtbar. So ist bereits in einem frühen Alter eine geschlechtsspezifische Präferenz beim Spielzeug zu erkennen. Während sich Jungen mehr für Bagger, Autos oder Roboter interessieren, beschäftigen sich Mädchen lieber mit Puppen, Kochen oder Anziehen (Oerter, 1998). Diese Entwicklung wird unter anderem durch das Prinzip der operanten Konditionierung erklärt. Das Verhalten des kleinen Mädchens wird verstärkt, wenn es – zunächst durch einen Zufall – zu einer Puppe greift, das Verhalten des kleinen Jungen beim Griff zum Auto (Richter, 1996). Ebenso denkbar ist der Mechanismus des Lernens am Modell, das durch Beobachtung der bevorzugten Tätigkeiten von Eltern und größeren Geschwistern ausgelöst wird. Eine wichtige Rolle spielt dabei – die Analogie zur operanten Konditionierung ist nicht zu übersehen –, ob das beobachtete Modellverhalten erfolgreich oder erfolglos ist, auf Anerkennung oder Kritik stößt, belohnt oder bestraft wird, wobei insbesondere das erfolgreiche Modellverhalten nachgeahmt wird (Kasten, 2001).

Tatsächlich weisen empirische Ergebnisse darauf hin, dass geschlechtskonformes Verhalten durch die Eltern verstärkt wird und Mädchen und Jungen bis zu einem gewissen Grad auch unterschiedlich behandelt werden. So erhalten Jungen häufiger positive Reaktionen beim Spielen mit maskulinem Spielzeug, während Väter weniger positive Reaktionen zeigen, wenn ihr männlicher Nachwuchs sich mit femininem Spielzeug beschäftigt (Oerter, 1998). Ge-

schlechtstypische Spielzeuge und die teilweise subtile Ermutigung der Spielzeugwahl durch Erwachsene wird vielfach mit dem späterem Sozial- und Leistungsverhalten in Verbindung gebracht, die „spielerische Einübung in und Vorwegnahme der Arbeitsteilung nach Geschlecht scheint eindeutig“ (Bilden, 2002, S. 283). Durch den signifikant häufigeren Umgang mit technischem Spielzeug, Büchern oder Baukästen haben Jungen aber auch einen Erfahrungsvorsprung, der für Mädchen schwierig aufzuholen ist (Menacher, 1994). Nach einer Untersuchung von Cooper und Robinson (1989) spielen die Interessen und Aktivitäten in der frühen Kindheit eine Rolle für die Entwicklung der Fähigkeiten, die für eine gute Leistung in Mathematik und Physik notwendig sind. So berichten junge Frauen, die eine Ausbildung im naturwissenschaftlichen-technischem Bereich begonnen hatten, häufig über männliche Jugendaktivitäten.

Neben der Sozialisation in der Familie hat in den vergangenen Jahrzehnten der Kindergarten als Sozialisationsinstanz zweifellos an Bedeutung gewonnen. Das Wissen um mögliche geschlechtsspezifische Effekte im Kindergarten ist zwar gering (Colberg-Schrader & von Derschau, 2002), aber nach Einschätzung von Bilden wird der Einfluss von Peers (Gleichaltrige) vor allem im Kleinkind- und Kindergartenalter erheblich unterschätzt. Nach ihrer Einschätzung sozialisieren sich Kinder rigoros im Sinne geschlechtstypischen Verhaltens, und durch die zunehmende Bevorzugung gleichgeschlechtlicher Spielkameraden festigen sich unterschiedliche Interessen und Verhaltensstile, die bereits mit fünf Jahren recht ausgeprägt sind (Bilden, 2002). Flankiert wird diese geschlechtsspezifische Sozialisation auch durch die Medien, insbesondere das Fernsehen, die zweifellos Informationen über die Geschlechter vermitteln, die den Erwerb von Stereotypen begünstigen. So vermittelt das Fernsehen „rigid stereotypisierte Geschlechtsrollen, konservativer als die zeitgenössischen Einstellungen“ (Bilden, 2002, S. 288). Ob Medien eine direkte Sozialisationswirkung haben oder lediglich eine Informationsquelle sind, ist umstritten. Unumstritten ist jedoch, dass Medien Modelle liefern, deren Darstellung von Kindern erinnert und gespeichert wird. Geschlechtsstereotype werden demnach nicht nur durch eigene Erfahrungen erworben, sondern zunehmend auch durch Informationen aus den Medien (Alfermann, 1996).

Aufgrund der großen sozialen Bedeutung der Geschlechtszugehörigkeit ist es nicht überraschend, dass Geschlechtsstereotype bereits früh in der kindlichen Entwicklung erworben werden (Trautner, 1994). Bereits bei einjährigen Kindern kann die Fähigkeit nachgewiesen werden, die zwei Geschlechter kognitiv voneinander zu unterscheiden. Bei zweijährigen Kindern lässt sich beobachten, dass sie sich und anderen das Etikett männlich beziehungsweise weiblich zuschreiben, und fünfjährige Kinder verfügen bereits über relativ ausgeprägte Geschlechtsstereotype (Alfermann, 1996).

Vor dem Hintergrund der beschriebenen geschlechtsspezifischen Sozialisationseinflüsse kommen Kinder als Mädchen beziehungsweise als Jungen in die Schule und machen dort entsprechende Erfahrungen.

„Dies wird üblicherweise in der Schule nicht nur nicht thematisiert, sondern sie trägt vielmehr zu einer Verfestigung von geschlechtsspezifischen Unterschieden im Verhalten – einschließlich kognitiver, emotionaler und motivationaler Aspekte – bei.“ (Ulich, 2002, S. 394)

So werden Geschlechtsstereotype durch Lerninhalte des Unterrichtsstoffs verstärkt und die Geschlechtspolarisation der Gesellschaft schlägt sich auch in Schulbüchern nieder (Ziegler, Kuhn & Heller, 1998). Besonders interessant sind jedoch die Lehrer-Schüler-Interaktionen: So widmen Lehrer Jungen deutlich mehr Aufmerksamkeit als Mädchen, sie werden häufiger aufgerufen, gelobt und angesprochen, wobei sich diese Interaktionsmuster insbesondere in den mathematisch-naturwissenschaftlich-technischen Fächern zeigen (Horstkemper, 1992). In der klasseninternen Kommunikation verhalten sich Jungen aktiver, selbstbewusster und abschweifender, Mädchen hingegen eher passiv, hilfsbedürftig und aufgabenorientiert (Ulich, 2002). Zugespitzt formuliert:

„Unter der Decke formaler Gleichheit finden in unseren gegenwärtigen Schulen Sozialisationsprozesse statt, die Mädchen klar benachteiligen.“ (Horstkemper, 1992, S. 93)

Nach Einschätzung von Marianne Horstkemper werden Mädchen nicht nur weniger beachtet, sondern die Stärken und Schwächen werden je nach Geschlechtszugehörigkeit unterschiedlich bewertet. Während Jungen kaum Rückmeldungen erhalten, die Informationen über mangelnde Fähigkeiten transportieren, werden bei Mädchen Misserfolge in einer Weise vermittelt, die sehr leicht mangelnde intellektuelle Fähigkeiten als Ursache nahe legen (Horstkemper, 1992). Konkret: Die Ursachenzuschreibung bezüglich der Leistungen in Mathematik unterscheidet sich je nach Geschlecht. Der Leistungserfolg der Jungen wird auf die mathematische Fähigkeit zurückgeführt, bei den Mädchen wird Anstrengung als Erfolgsursache herausgestellt. Ähnliche Botschaften vermitteln auch die Eltern. So schätzen Mütter die Anstrengung sowie die Schwierigkeit einer Aufgabe für ihre Töchter höher ein als für ihre Söhne, und die mathematischen Erfolge der Töchter werden mit Fleiß, die der Söhne hingegen mit Begabung erklärt (Menacher, 1994).

Diese geschlechtsspezifischen Kausalattributionen von Lehrern und Eltern finden sich auch bei den Betroffenen selbst. Eine Metaanalyse von geschlechtsspezifischen Einstellungen zur Mathematik lässt den Schluss zu, dass Mädchen und Jungen ihre jeweiligen Leistungsergeb-

nisse anderen Ursachen zuschreiben. So schreiben Jungen Erfolg ihrer mathematischen Fähigkeit, Mädchen ihrer Anstrengung beziehungsweise ihrem Glück zu. Misserfolg führen Mädchen im Gegenzug aber nicht auf fehlendes Glück oder fehlende Anstrengung zurück, sondern vielmehr auf eine geringere mathematische Fähigkeit (Hyde, Fennema, Ryan, Frost & Hopp, 1990).

Dieses für Mädchen beziehungsweise Frauen ungünstige Attributionsmuster beeinflusst die weitere Motivation sowie die Interessen und kann unter Umständen in einer Situation der erlernten Hilflosigkeit münden. Die attributionsorientierte Fassung der Theorie der erlernten Hilflosigkeit besagt, dass mit der Zuschreibung eines negativen Erlebnisses auf interne und stabile Faktoren ein Gefühl der Hilflosigkeit verbunden ist, das sich unter Umständen auf künftige Situationen im Mathematikunterricht auswirkt. So sieht der Betroffene möglicherweise keine Möglichkeit, seine Situation zu verändern, was sich in Resignation und schließlich Depression niederschlagen kann (Davison & Neale, 2002). Ganz allgemein ist die Kausalattribution auf Begabung für die Entstehung und Aufrechterhaltung von Motivation wohl günstiger, da Anstrengung zeit- und kraftaufwändig und somit ein langfristiges Engagement weniger wahrscheinlich ist (Menacher, 1994).

Auch Schlagzeilen wie „Do males have a math gene?“, „A new study says males may be naturally abler than females“ oder „Are girls born with less ability?“ (Chipman, 2005) haben negative Auswirkungen auf die elterlichen Einstellungen. So berichtet Meredith Kimball (1989), dass Mütter, die einen medienwirksam aufbereiteten Beitrag von Benbow und Stanley (1980)<sup>9</sup> gelesen hatten, ihre Töchter für weniger mathematisch begabt hielten, weniger Erfolg erwarteten und annahmen, dass sich ihre Töchter mehr anstrengen müssten, um gute Leistungen in Mathematik zu erzielen, als jene Mütter, die diesen Beitrag nicht gelesen hatten. Insofern können Erklärungsansätze, die geschlechtsspezifische Leistungsunterschiede in der Mathematik auf biologische Ursachen zurückführen, als eine sich selbst erfüllende Prophezeiung wirken. Der Begriff der sich selbst erfüllenden Prophezeiung bezeichnet Situationen, in denen ursprünglich falsche Annahmen wahr werden (Esser, 1999b). Für die Mathematik wäre folgende Logik denkbar: Eltern und Lehrer richten an Mädchen und Jungen unterschiedliche Erwartungen, der Erfolg und Misserfolg wird auf verschiedene Ursachen zurückgeführt, und entsprechend interagieren sowohl Eltern als auch Lehrer auf unterschiedliche Weise mit Jungen und Mädchen. So könnten beispielsweise Lehrer den Mädchen weniger Möglichkeiten geben, sich aktiv am Unterricht zu beteiligen, was bei den Mädchen den Eindruck erwecken könnte, der Lehrer hätte an ihrem Lernfortschritt ein geringeres Interesse als an dem der Jun-

---

<sup>9</sup> In dem Beitrag „Sex differences in mathematical ability: Fact or artifact“ von Camilla P. Benbow und Julian C. Stanley (1980), führen die Autoren die geschlechtsspezifischen Leistungsunterschiede in der Mathematik auf die biologisch bedingte männliche Überlegenheit beim räumlichen Vorstellungsvermögen zurück.

gen (Horstkemper, 1992). Die Konsequenz: Das Engagement und das Selbstvertrauen im Mathematikunterricht sinken, was sich negativ auf die Leistungen auswirkt. Schließlich bestätigen die schlechteren Leistungen die Erwartungen der Lehrer und Eltern, und der Kreis der Self-fulfilling Prophecy schließt sich.

Die geschilderten Schüler-Lehrer-Interaktionsmuster treten, wie bereits geschildert, insbesondere in den mathematisch-naturwissenschaftlichen Fächern hervor (Ulich, 2002), Gebieten, die immer noch häufig als männliche Domäne stereotypisiert werden (Beerman et al., 1992). Warum wird Mathematik aber als männliche und nicht etwa als weibliche Domäne stereotypisiert? Als Ursache für die Stereotypisierung von Mathematik als männliche Domäne wird darauf verwiesen, dass Mathematik schon immer als schwierig galt sowie durch ein hohes Maß an Abstraktheit und Formalität gekennzeichnet war. Diese Eigenschaften der Mathematik sind aber nicht mit dem weiblichen Stereotyp der Emotionalität vereinbar, weshalb Mathematik eher den Männern zugeschrieben wurde (Keller, 1998). Neben der informationsverarbeitenden Funktion wird Stereotypen aber auch eine motivationale Funktion zugeschrieben, die vorrangig darin besteht, die „bestehende gesellschaftliche Rang- und Wertordnung zu rechtfertigen und zu perpetuieren“ (Alfermann, 1996, S. 11). So kann das Stereotyp der eher praktisch veranlagten Frau in einer Gesellschaft, in der geistige Arbeit höher bewertet wird, dazu dienen, die gesellschaftliche Unterprivilegierung von Frauen zu rechtfertigen. Das Stereotyp des starken und schwachen Geschlechts war in einer Zeit, in der körperliche Kräfte in der Arbeitswelt noch eine größere Rolle spielten, zugleich auch ein Werturteil über die Bedeutsamkeit und Nützlichkeit der beiden Geschlechter (Alfermann, 1996). Entsprechend ist als Erklärung, warum Mathematik als männliche Domäne stereotypisiert wird, denkbar, dass „in einer Gesellschaft, in der der technische Fortschritt eine bedeutende Rolle spielt und hoch bewertet wird, die Abwesenheit und Unterprivilegierung von Mädchen und Frauen in mathematischen Schul- und Berufsfeldern legitimiert und perpetuiert wird“ (Keller, 1998, S. 34).

Die Stereotypisierung von Mathematik als männliche Domäne bleibt nicht ohne Folgen: Mädchen schreiben in diesem Fall Mathematik weniger dem eigenen Geschlecht zu, während Jungen Mathematik mit dem eigenen Geschlecht verbinden. Nach Kohlberg werden Inhalte, die einem Individuum als identifikationswürdig erscheinen, höher bewertet, schneller gelernt und weniger schnell vergessen. Der Effekt der Stereotypisierung von Mathematik als männliche Domäne auf die Leistung von Mädchen konnte empirisch nachgewiesen werden. Je mehr die Mädchen Mathematik als geschlechtsneutrales Fach betrachteten, desto höher waren auch ihre Leistungen (Keller, 1998), wobei mit steigendem Schuljahr Mathematik und Physik von beiden Geschlechtern stärker als männliche Domäne betrachtet wurde. Durch eine Vielzahl von subtilen Verhaltensweisen können Schüler ihren Mitschülerinnen, die gute Leistungen

erbringen, anzeigen, dass sie diese für weniger weiblich halten. Dies könnte dazu führen, dass Mädchen eine Unsicherheit bezüglich ihrer Rolle als Frau und der Tatsache, erfolgreich in Mathematik zu sein, verspüren (Menacher, 1994).

Darüber hinaus können Lehrkräfte bewusst oder unbewusst Mädchen entmutigen, sich im mathematischen Bereich zu engagieren und/oder eine mathematisch-naturwissenschaftliche Ausbildung anzustreben. Bei einer Untersuchung zu impliziten Theorien von gymnasialen Mathematik- und Physiklehrkräften zu geschlechtsspezifischer Begabung und Motivation zeigte sich, dass 26.5 Prozent der Mathematiklehrer und über 30 Prozent der Physiklehrkräfte in dem von ihnen unterrichteten Fach Jungen für begabter einschätzten als Mädchen. Demnach müssen viele Mädchen im Verlauf ihrer Schullaufbahn damit rechnen, einer Lehrkraft zu begegnen, die ihnen eine geringere Begabung zuschreibt. Noch deutlicher zeigte sich die geschlechtsgebundene Sichtweise der Lehrkräfte bei der Frage, welche Studienfächer sie Jungen und Mädchen empfehlen würden. Als geeignete Studienfächer für Mädchen wurden Grundschulpädagogik, Sprachwissenschaften und Medizin eingeschätzt, für Jungen der mathematisch-naturwissenschaftliche und der technische Bereich (Ziegler et al., 1998).

Die Effekte einer geschlechtsspezifischen Sozialisation haben in Deutschland zu einer Diskussion über den gemeinsamen Unterricht von Mädchen und Jungen (Koedukation) geführt (Beerman et al., 1992), denn ohne Konkurrenzverhältnis zu den Jungen gelingt es Mädchen offenbar besser, mathematisch-naturwissenschaftliche Leistungsstärken aufzubauen (Ulich, 2002). In einer Studie von Baumert (1992) findet die These einer generellen Benachteiligung von Mädchen im koedukativen Unterricht keine empirische Bestätigung. Allerdings wurde unter Koedukationsbedingungen eine stärkere Festlegung der geschlechtsspezifischen Interessenausprägung festgestellt: Bei gemeinsamem Unterricht sinkt das Interesse von Jungen an den sprachlichen Fächern, während das Interesse von Mädchen ansteigt. Im Fach Mathematik steht im koedukativen Unterricht ein deutlicher Interessenverlust der Mädchen einem etwas geringeren Interessenzuwachs der Jungen gegenüber, der auch statistisch abgesichert war. Diese Interessenpolarisierung könnte eine langfristige Auswirkung auf die Fach- und Berufswahlen haben, weshalb Baumert zumindest eine zeitweilige getrennte Unterrichtung von Mädchen und Jungen als ernst zu nehmende Möglichkeit betrachtet, diese geschlechtsspezifischen Diskrepanzen im Fach- und Berufswahlverhalten zu reduzieren.

Dieser Abschnitt sollte einen Überblick über die Konsequenzen einer geschlechtsspezifischen Sozialisation geben. Als Fazit bleibt festzuhalten, dass „die Auswirkungen geschlechtsspezifischer Sozialisationsprozesse nachhaltige Konsequenzen auf die Selbstwahrnehmung, die Einschätzung eigener Leistungsfähigkeit und die affektiven Einstellungen von Mädchen und Frauen bezüglich Mathematik, Naturwissenschaften und Technik haben“ (Menacher, 1994, S. 6). Dieses Fazit lässt sich durch empirische Ergebnisse belegen. Bei der internationalen PISA-

Studie 2003 wurden die Schüler unter anderem gefragt, wie sie ihre Leistungsfähigkeit in Mathematik einschätzen und welche Emotionen sie mit Mathematik verbinden. Die Abbildung enthält eine Zusammenfassung der geschlechtsspezifischen Unterschiede bei ausgewählten Schülermerkmalen. Die Unterschiede werden in Effektstärken ausgedrückt, wobei d-Werte von 0.2 und größer mit dunklerer Farbe gekennzeichnet sind.

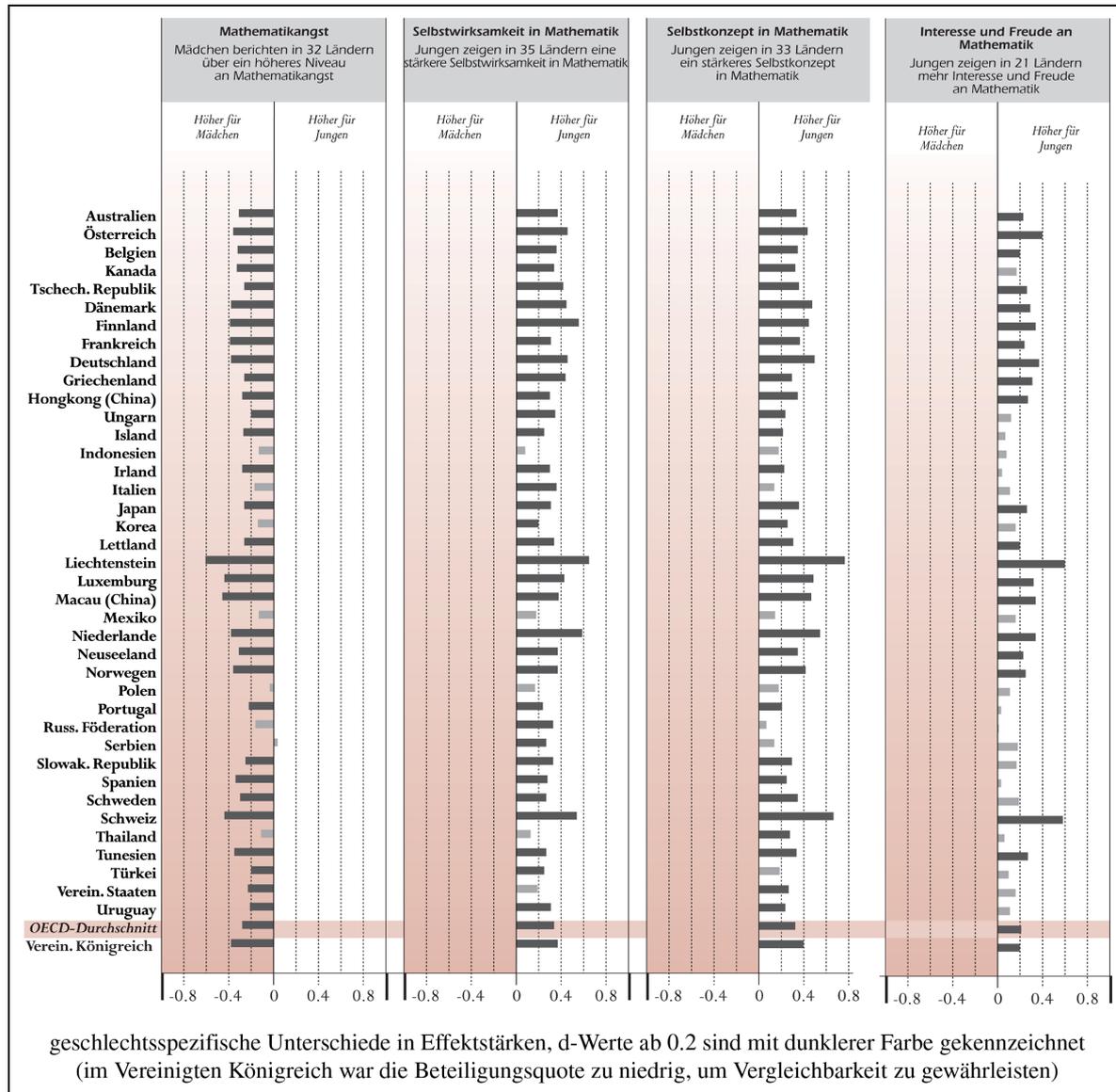


Abbildung 4-1: Geschlechtsspezifische Unterschiede bei Schülermerkmalen in der PISA-Studie 2003 (nach OECD, 2004a)

Bei der PISA-Studie 2003 berichten die Mädchen über mehr Angst, während sich die Jungen durch eine höhere Selbstwirksamkeit, ein höheres mathematisches Selbstkonzept und ein größeres Interesse an der Mathematik auszeichnen. Die Effektstärken betragen bei der Mathematikangst in 32 Staaten 0.20 und größer, bei der Selbstwirksamkeit sind es 35 Staaten, beim mathematischen Selbstkonzept 33 Staaten und beim Mathematikinteresse 21 Staaten. Die geschlechtsspezifischen Unterschiede sind bei den Schülermerkmalen deutlich größer als die Leistungsdifferenzen zwischen Mädchen und Jungen in der Mathematik (vgl. Seite 12).

## 4.2.2 Psychosoziale Erklärungsmodelle

Ausgehend von einer geschlechtsspezifischen Sozialisation wurden zahlreiche Erklärungsmodelle entwickelt, die die geschlechtsspezifischen Leistungsunterschiede in der Mathematik als Ergebnis der verschiedenen Sozialisationseinflüsse erklären. Einen Überblick über die vielfältigen Erklärungsansätze liefern die Monografie von Halpern (2000) und das Werk von Beerman et al. (1992). Einige Mechanismen, die allein oder insbesondere in Kombination einen Einfluss auf die Leistung von Jungen und Mädchen in Mathematik haben können, wurden bereits vorgestellt, beispielsweise das Lernen am Modell oder sich selbst erfüllende Prophezeiungen. An dieser Stelle sollen zwei Erklärungsansätze skizziert werden, die in der Literatur eine besondere Aufmerksamkeit gefunden haben. Dabei handelt es sich um das Modell des autonomen Lernverhaltens von Fennema und Peterson (1985) und das auf dem Wert-Erwartungs-Ansatz beruhende Modell von Eccles (1987).

### 4.2.2.1 Autonomes Lernverhalten

Fennema und Peterson betrachten autonomes Lernverhalten als eine notwendige Bedingung zum Lösen komplexer Mathematikaufgaben. Autonomes Lernverhalten wird dabei als unabhängiges, ausdauerndes, freiwilliges und erfolgreiches Arbeiten an schwierigen Aufgaben bezeichnet. In dem Modell wird das autonome Lernverhalten sowohl von internen als auch von externen Faktoren beeinflusst. Zu den internen Faktoren gehören der Attributionsstil, Vertrauen in die eigene mathematische Fähigkeit, der Wert, der der Mathematik beigemessen wird, sowie die Übereinstimmung der eigenen Geschlechtsrollenidentität mit der Ausführung des autonomen Lernverhaltens. Zu den externen Faktoren gehören die Familie, Peers und die Medien. Von besonderem Interesse sind jedoch das Klassenklima und die Lehrer-Schüler-Interaktionen (Fennema & Peterson, 1985). Durch die internen und externen Variablen wird nach Fennema und Peterson das autonome Lernverhalten derart beeinflusst, dass sich geschlechtsspezifische Leistungsunterschiede in der Mathematik ergeben. Das autonome Lernverhalten wird demnach als eine Mediatorvariable gesehen, über die der Effekt der internen und externen Variablen auf die Entwicklung mathematischer Fähigkeiten vermittelt ist. Die Abbildung zeigt das Modell des autonomen Lernverhaltens.

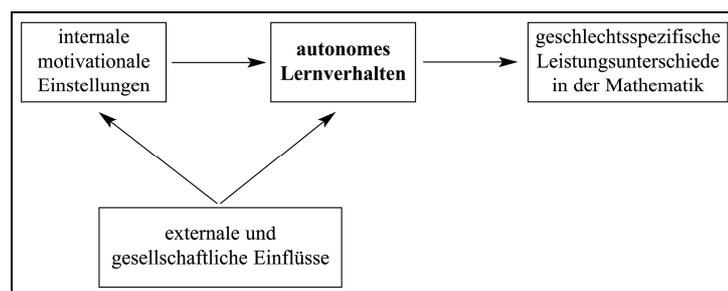


Abbildung 4-2: Modell des autonomen Lernverhaltens (nach Fennema & Peterson, 1985)

Nach Fennema und Peterson entwickelt sich das autonome Lernverhalten über mehrere Jahre, wobei ein sich selbst verstärkender Prozess postuliert wird. So führt ein größeres Engagement bei autonomen Lernaktivitäten wieder zu einer größeren Entwicklung des autonomen Lernverhaltens, was sich schließlich in einer höheren Leistungsfähigkeit bei schwierigen Aufgaben niederschlägt.

Die Leistungsvorteile in der Mathematik zugunsten der Jungen sind demnach das Ergebnis einer geschlechtsspezifischen Entwicklung des autonomen Lernverhaltens. Mädchen nehmen weniger an autonomen Lernaktivitäten teil, weil sie ein geringeres Selbstvertrauen haben, Mathematik eine geringere Bedeutung beimessen und Erfolg eher externen Faktoren zuschreiben, während Misserfolg auf die mangelnde Fähigkeit zurückgeführt wird. Darüber hinaus sind geringere autonome Lernaktivitäten zu erwarten, wenn Mathematik nicht mit der eigenen Geschlechterrolle übereinstimmt. Außerdem wird der Erwerb des autonomen Lernverhaltens bei den Mädchen auch von externen Faktoren negativ beeinflusst. So nehmen Mädchen beispielsweise weniger an autonomen Lernaktivitäten teil, wenn der Lehrer ihnen weniger Aufmerksamkeit schenkt.

Das Modell des autonomen Lernverhaltens verknüpft gleich mehrere Aspekte der geschlechtsspezifischen Sozialisation zu einem größeren Erklärungsansatz. Die vielfach in der Literatur berichteten Unterschiede zwischen Mädchen und Jungen in Selbstvertrauen, Kausalattribution oder auch Einschätzung der Bedeutung der Mathematik werden kombiniert mit externen Einflüssen, beispielsweise dem für Mädchen ungünstigen Lehrer-Schüler-Interaktionsmuster. Fennema und Peterson (1985) haben jedoch weder ihr Modell einer empirischen Prüfung unterzogen, noch die postulierte vermittelte Wirkung des Konstrukts getestet. Unklar bleibt auch, ob die internen und externen Faktoren jeweils den gleichen Einfluss auf das autonome Lernverhalten haben. Nach den Autorinnen entwickelt sich das autonome Lernverhalten über mehrere Jahre, so dass geschlechtsspezifischen Leistungsdifferenzen in der Mathematik allenfalls langfristig begegnet werden kann.

#### **4.2.2.2 Wert-Erwartungs-Ansatz**

Das umfassendste theoretische Modell, um geschlechtsspezifische Unterschiede in der Wahl von Ausbildungsgängen und Berufen vorherzusagen, wurde von Jacquelynne S. Eccles (1987) vorgelegt. Dabei handelt es sich um eine erweiterte Variante des Wert-Erwartungs-Ansatzes, bei dem schulische, berufliche und andere leistungsbezogene Entscheidungen vor allem auf zwei Faktoren zurückgeführt werden: die Erfolgserwartungen und die subjektive Bewertung der Aufgabe. Die Erfolgserwartungen und die subjektive Bewertung der Aufgabe sind dabei multiplikativ verknüpft. Jemand, der Mathematik als nützlich betrachtet, wird sich dennoch

nicht mit der Mathematik beschäftigen, wenn er seine Erfolgsaussichten als gering einschätzt. Die Auseinandersetzung mit der Mathematik unterbleibt jedoch auch, wenn man zwar die Erfolgsaussichten als positiv, die Mathematik jedoch als nicht wichtig beurteilt.

„Similarly, if both expectancy and value are high, their multiplied combination will be high, and if both are low, their multiplied combination will also be low.“ (Halpern, 2000, S. 289)

Der subjektive Wert und die Erfolgserwartungen werden von einer Reihe von Variablen beeinflusst, beispielsweise die vorangegangenen leistungsbezogenen Erfahrungen, die Kausalattribution, die Einstellungen und Verhaltensweisen der Umwelt, das kulturelle Milieu, die subjektive Wahrnehmung der Annahmen, Erwartungen und Einstellungen der Umwelt oder auch die individuellen Ziele und subjektiven Selbstbilder. Das Modell besagt, dass alle diese Faktoren sowohl die Erfolgserwartungen als auch den subjektiven Wert, der der Mathematik beigemessen wird, beeinflussen.

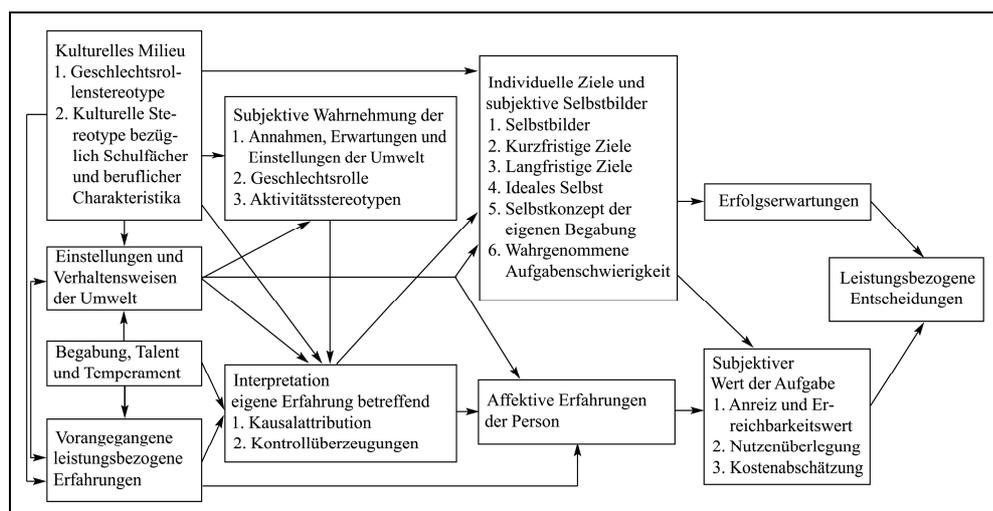


Abbildung 4-3: Modell zur Vorhersage von leistungsbezogenen Entscheidungen (nach Eccles, 1987)

Das Modell von Eccles hat eine Vielzahl von empirischen Arbeiten angeregt und verschiedene Aspekte konnten empirisch bestätigt werden (Eccles, 1994), wobei insbesondere die Einstellungen und Verhaltensweisen der Eltern, die Selbstwahrnehmung der Kinder und die geschlechtstypischen Rollenerwartungen einen Einfluss auf den Wert haben, der der Mathematik beigemessen wird. In einer neueren Studie konnte gezeigt werden, dass „parents appear to provide more math-supportive environments for their sons than for their daughters by purchasing more math/science toys for sons, spending more time on math/science with sons, and holding higher perceptions of their sons’ than daughters’ math abilities as well as gender-typed worldviews about natural talent in math“ (Jacobs, Davis-Kean, Bleeker, Eccles & Malanchuk, 2005, S. 259).

Insgesamt illustriert das Modell von Eccles das komplexe Zusammenspiel vieler psychosozialer Faktoren, worunter allerdings die Übersichtlichkeit etwas leidet. Geschlechtsdifferenzen in der Mathematik sind sowohl bei Eccles als auch schon bei dem Ansatz des autonomen Lernverhaltens von Fennema und Peterson (1985) das Resultat eines langjährigen Prozesses, kurzfristige Faktoren spielen in dem erweiterten Wert-Erwartungs-Modell – wenn überhaupt – nur eine untergeordnete Rolle.

### **4.3 Psychobiosoziales Modell**

Erklärungen, die Leistungsdifferenzen zwischen Jungen und Mädchen in der Mathematik nur auf physiologische Ursachen zurückführen, und Ansätze, die ausschließlich die Bedeutung der Sozialisation betonen, werden häufig als antagonistisch betrachtet (Stanat & Kunter, 2003). Diane F. Halpern (2000) hat ein Modell vorgelegt, das den Gegensatz zwischen Biologie und Sozialisation aufzulösen versucht. Sie versteht ihr psychobiosoziales Modell als eine Alternative zur klassischen Nature-Nurture-Dichotomie.

„The psychobiosocial model is based on the idea that some variables are both biological and social and, therefore, cannot be classified into one of these two dichotomous categories. For example, learning is a socially mediated event with a biological basis. Individuals are predisposed to learn some topics and skills more readily than others.”  
(Halpern & LaMay, 2000, S. 239f)

Grundsätzlich geht das psychobiosoziale Modell von einer Wechselwirkung zwischen biologischen und sozialen Faktoren aus. Forschungsergebnisse zur Spezialisierung der Gehirnhälften (Lateralisierung) und der Einfluss von Geschlechtshormonen sprechen jedoch für unterschiedliche Entwicklungsvoraussetzungen für Männer und Frauen in verschiedenen kognitiven Bereichen (Halpern et al., 2005). Während es den Mädchen im Durchschnitt etwas leichter fällt, ihre Sprachfertigkeiten zu entwickeln, scheinen die Jungen Vorteile beim räumlichen Vorstellungsvermögen zu haben. Diese vermutlich biologisch vorliegenden Tendenzen sind jedoch keineswegs deterministisch.

„However, I do not believe that cognitive abilities are simply determined, nor do I believe that a single answer like ‚It’s all in the hormones’ or ‚It’s all because of mothers’ attitudes’ will ever emerge as the origin of all sex differences in cognitive abilities.”  
(Halpern, 2000, S. 327)

Allerdings wird die biologische Komponente des psychobiosozialen Modells von psychosozialen Faktoren ergänzt. Nach Halpern werden die biologisch bedingten Unterschiede von der

Umwelt wahrgenommen, die dann entsprechende Erwartungen ausbildet, die wiederum den Fähigkeitserwerb in geschlechtstypischer Weise beeinflussen. Die differenziellen Lernerfahrungen haben Rückwirkungen auf biophysiological Funktionen, was den biologisch bedingten Anteil in geschlechtsspezifischen Leistungsunterschieden weiter verstärkt. Wenn beispielsweise Männer biologisch bedingte Vorteile im mathematischen Bereich haben, dann sammeln sie eher als Frauen mathematisch bezogene Erfahrungen, die ihre Fähigkeiten verbessern und sie motivieren, sich auch mit schwierigen mathematischen Problemen auseinander zu setzen. Gesellschaftliche Erwartungen würden diesen Prozess insofern unterstützen, als Männer im Vergleich zu Frauen mehr Ermutigung und Belohnungen erhalten würden. Diese Lernerfahrungen und Umwelteinflüsse können im Gegenzug wieder die Gehirnstruktur beeinflussen, womit die Trennung zwischen Natur und Sozialisation verwischt würde (Halpern & LaMay, 2000). Eine bildhafte Vorstellung von dem kontinuierlichen Zusammenspiel zwischen biologischen und sozialen Faktoren gibt die Abbildung:

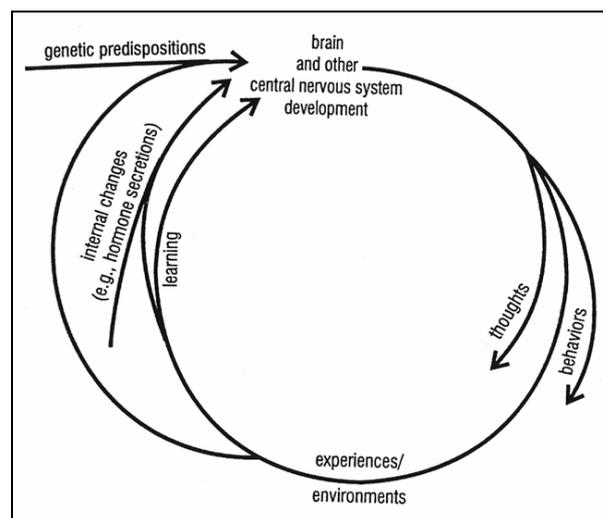


Abbildung 4-4: Psychobiosoziales Modell (nach Halpern, 2000)

Nach Halpern bietet das psychobiosoziale Modell eine bessere Alternative als die Nature-Nurture-Dichotomie oder das Nature-Nurture-Kontinuum. Anstatt anzunehmen, dass biologische und soziale Faktoren jeweils einen unabhängigen Beitrag zur Entwicklung von kognitiven Fähigkeiten leisten, postuliert das psychobiosoziale Modell eine wechselseitige Beziehung, in der Nature und Nurture aufeinander angewiesen sind.

„We believe this sort of model will move the field ahead so unwinnable battles over ‚how much‘ of the variance is attributable to biological or environmental effects are left on the old battlefields in favor of models with multiple and reciprocal influences that can be used to help all students achieve at higher levels in math and other academic areas.“ (Halpern et al., 2005, S. 68)

Das Modell von Halpern vereinigt sowohl biologische als auch soziale Erklärungsansätze, impliziert aber auch, dass geschlechtsspezifische Leistungsunterschiede in kognitiven Fähigkeiten beeinflusst werden können. Schließlich zeigt sich in der TIMSS- und in der PISA-Studie, dass die Größe und auch die Richtung von Leistungsunterschieden zwischen Jungen und Mädchen erheblich in den verschiedenen Staaten variieren. Insofern kann es Bildungssystemen durchaus gelingen, geschlechtsspezifische Benachteiligungen aufzuheben (Stanat & Kunter, 2002).

#### **4.4 Einfluss von Stereotypen in der Testsituation**

Die biologischen und sozialisationstheoretischen Erklärungsansätze sind dadurch gekennzeichnet, dass die Leistungsdifferenzen von Mädchen und Jungen in der Mathematik weitgehend auf dispositionale Eigenschaften der Person zurückgeführt werden. Sowohl die Effekte einer geschlechtsspezifischen Sozialisation als auch vorgeburtliche hormonelle Einflüsse auf die Gehirnstruktur lassen sich jedoch nur sehr langwierig oder eben gar nicht verändern. Demnach lassen sich Geschlechterdifferenzen in der Mathematik – wenn überhaupt – nur schwer beeinflussen und Handlungsempfehlungen müssen langfristig ausgelegt sein.

Eine optimistischere Perspektive bietet ein relativ neuer Erklärungsansatz, der situative Merkmale der Testsituation berücksichtigt. So konnte in einer verblüffenden Serie von Experimenten gezeigt werden, dass die geschlechtsspezifischen Leistungsunterschiede in der Mathematik auf situative Merkmale der Testsituation zurückzuführen sind und auf einem Phänomen beruhen, das als Stereotype Threat bezeichnet wird (Spencer, Steele & Quinn, 1999).

Das Konzept des Stereotype Threat wurde ursprünglich entwickelt, um Leistungsdifferenzen in kognitiven Fähigkeiten zwischen Weißen und Farbigen zu erklären. Wenn sich afroamerikanische Studierende in einer Testsituation befinden, neigen die meisten dazu, darüber besorgt zu sein, dem existierenden negativen kulturellen Stereotyp von intellektueller Minderwertigkeit zu entsprechen. Die daraus resultierende Extralast an Besorgnis kollidiert mit der Fähigkeit der Person, in dieser Situation eine gute Leistung zu erzielen. Die Bedrohung durch das Stereotyp der intellektuellen Minderwertigkeit führt zu einer Leistungsminderung, die in Situationen, in denen das Stereotyp keine Rolle spielt, nicht zu beobachten ist.

Weitere Untersuchungen konnten zeigen, dass das Phänomen der Bedrohung durch Stereotype nicht auf die Rasse beschränkt ist, sondern sich ebenso auf das Geschlecht bezieht. Konkret wird angenommen, dass bei Mädchen beziehungsweise Frauen, die einen Mathematiktest bearbeiten, die Befürchtung auftritt, sie könnten auf der Grundlage des in vielen Gesellschaften verbreiteten Stereotyps „Mädchen beziehungsweise Frauen können keine Mathematik“ beurteilt werden. Diese wahrgenommene Befürchtung führt zu einer Absenkung der Leis-

tungsfähigkeit bei Mädchen beziehungsweise Frauen, die vermieden werden kann, wenn diese wahrgenommene Befürchtung aus der Testsituation entfernt wird. Stereotype Threat kann als ein situationaler Leistungsdruck verstanden werden, der unter bestimmten Randbedingungen zu einer Reduzierung der Leistungsfähigkeit führt (Aronson, Wilson & Akert, 2004).

#### **4.4.1 Theoretische Annahmen**

Claude Steele (1997) argumentiert, dass in Situationen, in denen ein negatives Stereotyp auf bestimmte Gruppen anwendbar ist, dieses Stereotyp die Leistungsfähigkeit von Individuen beeinträchtigen kann. Als „threat in the air“ (Steele, 1997, S. 613) befindet sich das Stereotyp bildlich gesprochen wie eine Bedrohung in der Luft, während sich die Person in der stereotyp-relevanten Situation befindet. Eine allgemeine Definition von Stereotype Threat schlagen Claude M. Steele, Steven J. Spencer und Joshua Aronson (2002) vor:

„We offer the following definition of stereotype threat: When a negative stereotype about a group that one is part of becomes personally relevant, usually as an interpretation of one’s behavior or an experience one is having, stereotype threat is the resulting sense that one can then be judged or treated in terms of the stereotype or that one might do something that would inadvertently conform it.” (Steele et al., 2002, S. 389)

Stereotype Threat entsteht, wenn sich eine Person in einer Situation befindet, in der negative stereotype Erwartungen auf ihr Verhalten anwendbar sind, und ist mit der Befürchtung verbunden, negativ beurteilt zu werden beziehungsweise durch das eigene Verhalten negative stereotype Erwartungen zu bestätigen. Stereotype Threat kann aber auch durch die Anwendbarkeit eines positiven Stereotyps entstehen. Aronson, Lustina, Good, Keough, Steele und Brown (1999) zeigten, dass weiße Männer schlechter bei einem Mathematiktest abschnitten, wenn sie mit asiatischen Männern in einer Testsituation konfrontiert waren – einer Gruppe, die als besonders mathematisch begabt beurteilt wird.

Bezogen auf den Leistungsbereich wird Stereotype Threat als ein situationaler Leistungsdruck verstanden, der entsteht, wenn negative stereotype Erwartungen auf das Verhalten einer Person anwendbar sind, die sowohl durch negative als auch durch positive Stereotype erzeugt werden können. Dabei lassen sich zwei Konsequenzen von Stereotype Threat unterscheiden:

„According to his theory, stereotype threat can hamper achievement in academic domains in two ways. First, the cognitive and emotional reactions to the activation of stereotype threats in an academic setting can directly interfere with performance. Second, if this stereotype threat is chronic in the academic domain, it can cause individu-

als to disidentify from the domain – to dismiss the domain as no longer relevant to their self-esteem and identity.” (Brehm et al., 1999, S. 166)

Für die Entstehung von Stereotype Threat unterscheiden Steele et al. (2002) zwischen situationalen und individuellen Randbedingungen:

„Stereotype threat is a person-situation predicament: a person contending with the possibility of being negatively stereotyped. As such, the strength of the threat, and its influence on behavior, is likely to be moderated by features of both the situation and the person.” (Steele et al., 2002, S. 391)

Zu den situationalen Faktoren zählen die Testschwierigkeit, die Testdiagnostizität und die Relevanz des Stereotyps in der jeweiligen Situation. So muss der Leistungstest einen hohen Schwierigkeitsgrad besitzen, bei leichten Tests ist kein Stereotype-Threat-Effekt zu beobachten. Die Testschwierigkeit kann dabei sowohl durch schwierige Aufgaben als auch durch andere Faktoren, beispielsweise Zeitdruck, verursacht werden. In beiden Fällen ist die Wahrscheinlichkeit höher, dass die Person während der Bearbeitung der Aufgaben Frustration erlebt.

Der Stereotype-Threat-Effekt ist zudem nur nachzuweisen, wenn die Person annimmt, dass der Test die betreffende Fähigkeit zuverlässig misst. So haben Steele und Aronson (1995) in einer Grundlagenarbeit zum Stereotype Threat farbige und weiße Studierende einen verbalen Leistungstest bearbeiten lassen, bei dem sie die Diagnostizität des Tests variierten. Der Hälfte der Versuchspersonen wurde mitgeteilt, dass der Test ihre intellektuellen Fähigkeiten messe, der anderen Gruppe wurde gesagt, der Test ziele darauf ab, die psychologischen Faktoren beim Lösen von verbalen Aufgaben zu identifizieren. Tatsächlich zeigten die Versuchspersonen afroamerikanischer Abstammung in der Diagnostic-Bedingung eine schlechtere Leistung, eine erhöhte Zugänglichkeit stereotyper Wissensinhalte und eine Distanzierung von der Eigengruppe als die Afroamerikaner in der Non-diagnostic-Bedingung.

Ein relevantes Stereotyp muss darüber hinaus auf die Testsituation anwendbar sein. So schneidet beispielsweise eine Frau, die schwierige Mathematikaufgaben bearbeitet, schlechter ab, weil sich für diese Situation das Stereotyp „Frauen können keine Mathematik“ anwenden lässt. Bearbeitet die Frau jedoch einen schwierigen Sprachtest, bleibt eine Absenkung der Leistungsfähigkeit durch Stereotype Threat aus, da das Stereotyp in dieser Situation nicht zutrifft.

Zu den individuellen Randbedingungen des Stereotype-Threat-Effekts zählen die Domain-identifikation, die Gruppenidentifikation und das Stigmaabewusstsein (Stigma Consciousness).

Die stärksten Stereotype-Threat-Effekte sind ironischerweise bei den Personen zu beobachten, „who are highly invested and skilled in the targeted domain“ (Davies & Spencer, 2005, S. 174). Das negative Stereotyp „Frauen können keine Mathematik“ hat also vor allem bei den Frauen eine leistungsreduzierende Wirkung, die sich stark mit Mathematik identifizieren. Dabei müssen die betroffenen Personen gar nicht an das Stereotyp glauben beziehungsweise das Stereotyp internalisiert haben.

„Second, if our reasoning is correct, stereotype threat is likely to have its strongest effects among those who are least likely to internalize or accept the stereotype – those who are heavily invested in the domain.“ (Aronson, Quinn & Spencer, 1998, S. 87)

Als weiterer Moderator des Stereotype-Threat-Effekts wird in der Literatur die Gruppenidentifikation (Gender Identification) genannt. So konnte Schmader (2002) zeigen, dass sich Stereotype Threat stärker bei Personen auswirkt, die sich stark mit ihrer Gruppe identifizieren. Frauen, die ihr Geschlecht als wichtigen Teil ihrer Identität bezeichneten, schnitten in einem schwierigen Mathematiktest schlechter ab als Frauen, die sich weniger stark mit ihrer Geschlechtsgruppe identifizierten.<sup>10</sup> Die Gruppenidentifikation kann möglicherweise aber auch den Stereotype-Threat-Effekt abschwächen. In einer Studie von Oyserman, Harrison und Bybee (2001) zeigten farbige Mädchen, die sich stark mit der Gruppe der African Americans identifizierten, eine höhere Bedeutung der akademischen Selbstwirksamkeit. Als Erklärung für diesen widersprüchlichen Befund verweisen Steele et al. (2002) auf unterschiedliche Erfahrungen mit Stereotypen. So sind African Americans mit einer Vielzahl von negativen Stereotypen konfrontiert. Dies habe möglicherweise zur Entwicklung von individuellen und kollektiven Strategien geführt, die den Stereotype-Threat-Effekt abmildern.

Individuelle Unterschiede in der Stärke des Stereotype-Threat-Effekts sind unter Umständen auch auf das Stigmabewusstsein zurückzuführen. Pinel (1999) entwickelte ein Maß, das die Wahrnehmungen der stereotypen Bedrohungen einer Person in ihrer Umwelt erfasst (Stigma Consciousness). In der Tat vermieden Frauen mit einem hohen Stigmabewusstsein in einem Jeopardy-Spiel gegen Männer eher männliche Antwortkategorien als Frauen mit einem niedrigen Stigmabewusstsein. Bei einem standardisierten Test zeigten Versuchspersonen mit einem hohen Stigmabewusstsein schlechtere Ergebnisse als die Teilnehmer mit einem niedrigen Stigma Consciousness (Brown & Pinel, 2003).

Von den Randbedingungen, die für die Entstehung von Stereotype Threat verantwortlich sind, kann abgeleitet werden, dass die Leistung von weiblichen Testpersonen bei mathematischen

---

<sup>10</sup> Sowohl Brown und Pinel (2003) als auch Cadinu, Maass, Frigerio, Impagliazzo und Latinotti (2003) konnten jedoch keinen Effekt der Gender Identification auf das Testergebnis nachweisen.

Aufgaben ansteigt, wenn die wahrgenommene Befürchtung aus der Testsituation entfernt wird. Dies kann beispielsweise durch eine kurze schriftliche Mitteilung vor der Bearbeitung der Aufgaben geschehen, die den Versuchspersonen glaubwürdig versichert, dass sich bei der Zusammensetzung dieser Aufgaben in der Vergangenheit noch nie Geschlechtsunterschiede gezeigt haben.

Neben dieser als Fairnessbedingung bezeichneten Manipulation wurde in einigen Untersuchungen auch die Diagnostizität des Tests experimentell manipuliert, um Stereotype Threat aufzuheben. So haben in der bereits berichteten Grundlagenstudie Steele und Aronson (1995) einer Hälfte der Versuchspersonen mitgeteilt, der Test messe ihre intellektuellen Fähigkeiten, während der anderen Gruppe gesagt wurde, der Test ziele darauf ab, die psychologischen Faktoren beim Lösen der Aufgaben zu identifizieren. Bei der letztgenannten Bedingung war das betreffende Stereotyp nicht anwendbar und es zeigte sich im Vergleich zur Diagnostic-Bedingung ein deutlicher Leistungsanstieg. Allerdings zeigte sich bei anderen Untersuchungen, dass eine Non-diagnostic-Bedingung nicht per se zu einer Aufhebung von Stereotype Threat führen muss. Bei Davies, Spencer, Quinn und Gerhardstein (2002) schauten sich die Versuchspersonen zunächst mehrere Werbespots an. Die eine Hälfte sah Spots, die Frauen in klassischen Rollenklischees zeigte, während die andere Hälfte mit neutralen Werbefilmen konfrontiert wurde, ehe die Frauen einen Non-diagnostic-Mathematiktest bearbeiteten. Der durch die Non-diagnostic-Bedingung aufgehobene Stereotype-Threat-Effekt zeigte sich nur bei den Frauen, die die neutralen Werbespots sahen. Es wird argumentiert, dass „the negative commercials would activate negative stereotypes about women’s ability, including their math ability, and that once activated, these stereotypes would undermine their math performance and pressure them to disengage from the domain“ (Steele et al., 2002, S. 393).

Weitere Manipulationen beziehen sich auf die Auffälligkeit des Gruppenvergleichs. So haben Aronson et al. (1999) weiße Versuchsteilnehmer mit der angeblichen Überlegenheit asiatischer Personen in einem Mathematiktest konfrontiert. Der leistungsmindernde Effekt wurde in der Kontrollgruppe dadurch verhindert, dass bestimmte Aspekte des Selbstkonzepts aktiviert wurden, auf die negative Erwartungen nicht anwendbar waren.

#### **4.4.2 Empirische Befunde**

Das Phänomen der Bedrohung durch Stereotype ist – wie bereits geschildert – nicht nur auf das Geschlecht beschränkt, sondern „members of any stigmatized group are susceptible to stereotype threat when they find themselves in a situation where negative stereotypes provide a framework for interpreting their behavior“ (Davies & Spencer, 2005, S. 183). So konnte der Stereotype-Threat-Effekt bei Leistungstests zwischen Weißen und Afroamerikanern (Steele &

Aronson, 1995), zwischen jungen und älteren Menschen (Rahhal, Hasher & Colcombe, 2001) oder auch bei Sportaufgaben (Stone, Lynch, Sjomeling & Darley, 1999) nachgewiesen werden. Da diese Arbeit jedoch darauf abzielt, die Leistungsdifferenzen von Mädchen und Jungen beziehungsweise Frauen und Männern in der Mathematik zu erklären, konzentriert sich dieser Abschnitt auf Studien, die den Stereotype-Threat-Effekt im mathematischen Bereich untersuchten.

Die erste Studie, die den Einfluss von Stereotype Threat auf die mathematische Leistung von Frauen gezielt untersucht, haben Steven J. Spencer, Claude M. Steele und Diane M. Quinn (1999) vorgelegt. Im zweiten von drei Experimenten waren 30 Frauen und 24 Männer aufgefordert, einen schwierigen Mathematiktest zu bearbeiten. Abgeleitet von den theoretischen Annahmen wurde argumentiert, dass die Leistung von weiblichen Testpersonen bei mathematischen Aufgaben ansteigt, wenn die wahrgenommene Bedrohung entfernt wird. Dazu haben Spencer et al. (1999) der Hälfte der Versuchspersonen mitgeteilt, dass sich bei den Aufgaben in der Vergangenheit noch keine geschlechtsspezifischen Leistungsunterschiede gezeigt hätten (No-Gender-Difference), während der anderen Hälfte das Gegenteil gesagt wurde (Gender-Difference). Die Ergebnisse der Studie stützen die theoretischen Annahmen: Bei den Frauen war durch den Hinweis auf die Testfairness ein starker Anstieg in der Mathematikleistung zu erkennen. Die Befunde sind in der Abbildung zusammengefasst:

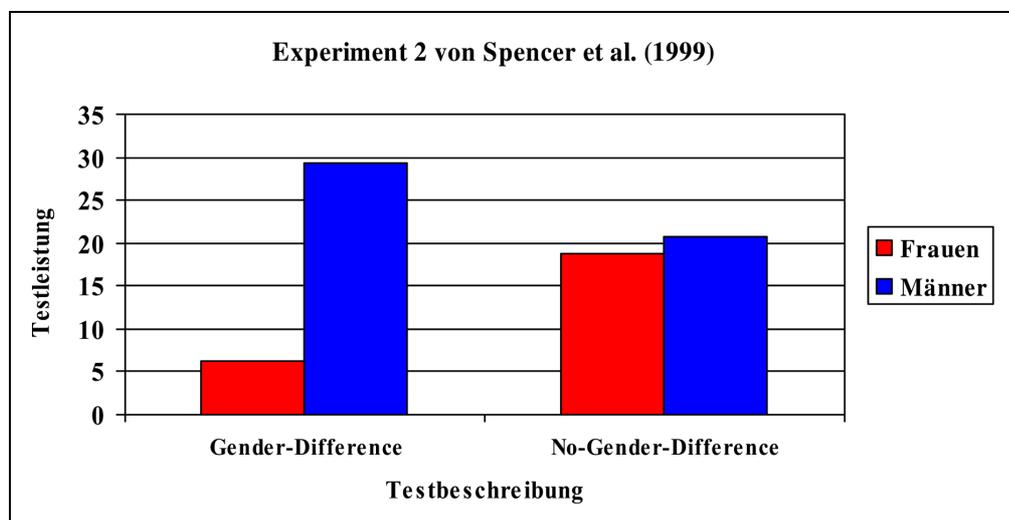


Abbildung 4-5: Testleistung in Abhängigkeit von Testbeschreibung und Geschlecht (nach Spencer et al., 1999)

Ist der Stereotype-Threat-Effekt unabhängig vom Schwierigkeitsgrad der Aufgabe? Diese Frage untersuchten Spencer et al. (1999) im ersten Experiment. 28 Frauen und 28 Männer von der Universität von Michigan waren aufgefordert, einen Mathematiktest zu bearbeiten. Der einen Hälfte der Versuchspersonen wurde ein einfacher Test (GRE general exam), der anderen Hälfte ein anspruchsvoller Test (advanced GRE exam) vorgelegt. Die Instruktionen durch den Versuchsleiter waren dabei für beide Gruppen identisch.

Die Ergebnisse dieses Experiments sind in der Abbildung zusammengefasst:

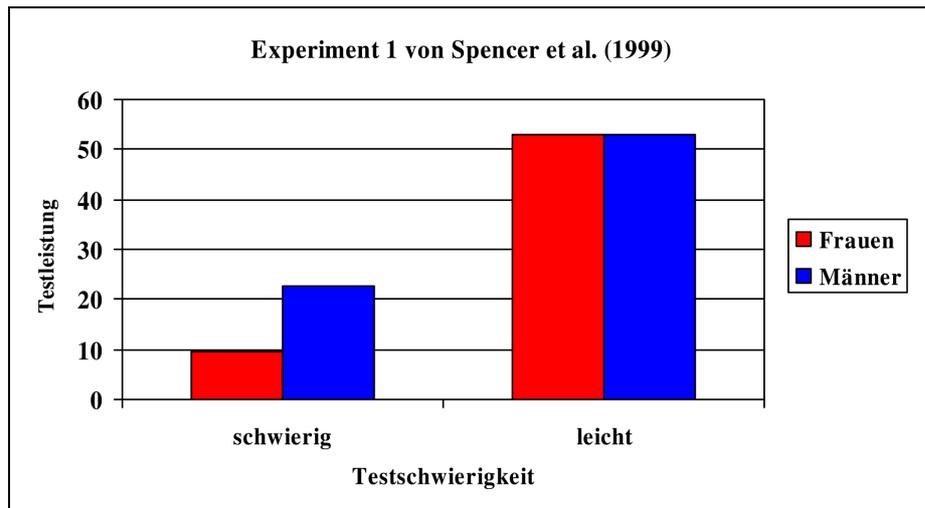


Abbildung 4-6: Testleistung in Abhängigkeit von Testschwierigkeit und Geschlecht (nach Spencer et al., 1999)

Der durch Stereotype Threat verursachte Leistungsabfall von Frauen ist nur bei schwierigen Mathematikaufgaben zu finden, bei leichten Aufgaben unterscheiden sich die Leistungen von Männern und Frauen nicht.

Bereits die Studie von Sharps, Welton und Price (1993) kann als Test der Stereotype-Threat-Theorie angesehen werden, da im zweiten Experiment die Diagnostizität eines Tests der räumlichen Wahrnehmungsfähigkeit manipuliert wurde. Geschlechtsspezifische Leistungsunterschiede traten nur auf, wenn die Aufgaben explizit als ein Test der räumlichen Wahrnehmungsfähigkeit beschrieben wurden, während in der Bedingung ohne Hinweis auf das räumliche Wahrnehmungsvermögen keine Geschlechtsunterschiede zu beobachten waren. Diesen Befund konnten Sharps, Price und Williams (1994) in einer weiteren Untersuchung replizieren. Es zeigte sich, dass „men performed at higher levels than women in the spatial instruction conditions, but no sex differences were observed under nonspatial instructions“ (Sharps et al., 1994, S. 413).

James und Greenberg (1997) haben in ihrer Studie unter anderem die Art der Testauswertung variiert. Der einen Hälfte der weiblichen Versuchspersonen wurde mitgeteilt, die Leistung von Individuen solle verglichen werden, während der anderen Hälfte gesagt wurde, es würden die Leistungsunterschiede zwischen Männern und Frauen verglichen. Dann bearbeiteten die Personen einen Test der räumlichen Wahrnehmungsfähigkeit, bei dem die Frauen besser abschnitten, die annehmen konnten, die Studie zielt darauf ab, individuelle Leistungsunterschiede zu untersuchen.

Bei Brown und Josephs (1999) zeigten die weiblichen Versuchspersonen bessere Leistungen, wenn ihnen zuvor gesagt wurde, der Mathematiktest identifiziere Personen, die über besonders gute mathematische Fähigkeiten verfügen, während sie schlechter abschnitten, wenn ih-

nen mitgeteilt wurde, durch den Test könnten Personen mit schlechten mathematischen Fähigkeiten erkannt werden. Bei den männlichen Probanden zeigte sich das gegenteilige Muster. Die Autoren erklären diesen Befund damit, dass je nach Testbeschreibung ein besonderer Leistungsdruck ausgelöst wird: So stehen Frauen, die einen Test bearbeiten, der Personen identifizieren soll, die in Mathematik besonders schlecht seien, unter dem Druck, die negative Erwartung zu widerlegen. Die männlichen Versuchspersonen seien dagegen aufgefordert, die stereotype Erwartung guter mathematischer Fähigkeiten bei Männern zu belegen. Dieser besondere Leistungsdruck führe dann jeweils zu einer verringerten Leistungsfähigkeit.

Walsh, Hickey und Duffy (1999) konnten belegen, dass Frauen in einem standardisierten Mathematiktest den männlichen Testteilnehmern unterlegen waren, wenn ihnen zuvor gesagt wurde, der Test habe bisher geschlechtsspezifische Leistungsunterschiede nachweisen können. Die Leistungsdifferenzen zwischen Männern und Frauen waren jedoch verschwunden, wenn der Test als nichtsensitiv für Geschlechterdifferenzen präsentiert wurde. Die nicht signifikanten Geschlechtsunterschiede in der No-Threat-Bedingung sind jedoch eine Folge der schlechteren Leistung der Männer und nicht der besseren Testergebnisse der Frauen. Interessanterweise hat bei dieser Untersuchung die Bedrohung durch Stereotype gerade bei den Frauen, die über eine hohe mathematische Fähigkeit verfügen, nicht zu einem Leistungsabfall geführt.

Bereits die Anwesenheit von männlichen Testteilnehmern scheint bei Frauen zu einer verringerten Leistungsfähigkeit bei mathematischen Aufgaben zu führen. Inzlicht und Ben-Zeev (2000) haben einerseits das Geschlecht anderer (vermeintlicher) Testteilnehmer variiert (weiblich/männlich), andererseits unterschiedliche Testaufgaben vorgelegt (mathematische versus sprachliche Aufgaben). Es zeigte sich, dass Frauen beim Bearbeiten von Mathematikaufgaben in Gegenwart männlicher Personen schlechter abschnitten als in Gegenwart weiblicher Testpersonen, während bei sprachlichen Aufgaben die Anwesenheit männlicher/weiblicher Probanden keinen negativen Leistungseinfluss hatte. Das Experiment unterstreicht die Bedeutung der Anwendbarkeit negativer stereotyper Erwartungen – eine schlechtere Leistung der Frauen war nur bei den Mathematikaufgaben zu erwarten, da nur hier das negative Stereotyp „Frauen können keine Mathematik“ anwendbar war.

Stereotype Threat kann auch ganz subtil wirken: Oswald und Harvey (2000/01) haben den weiblichen Probanden zwar mitgeteilt, dass die mathematischen Leistungen zwischen Frauen und Männern gleich seien, sie während des Mathematiktests aber mit einem Cartoon konfrontiert, welches die männliche Überlegenheit in Mathematik deutlich machte. In der Fairnessbedingung ohne Cartoon zeigten die Frauen bessere Leistungen als die Frauen in der Fairnessbedingung mit Cartoon.

Bei einer Untersuchung von Quinn und Spencer (2001) schnitten die männlichen Testteilnehmer unter Stereotype Threat signifikant besser ab als die weiblichen Versuchspersonen, während in der No-Threat-Bedingung (Fairnessbedingung) kein signifikanter Leistungsunterschied bestand (Experiment 2). Die Autoren zeigten zudem, „when stereotype threat was high, women were less able to formulate problem-solving strategies than when stereotype threat was reduced“ (Quinn & Spencer, 2001, S. 55).

In der bereits dargelegten Studie von Davies et al. (2002) sahen die Personen zunächst mehrere Werbefilme, bevor sie einen Non-diagnostic-Mathematiktest bearbeiteten. Dabei sah eine Hälfte der Versuchspersonen Werbespots, die Frauen in klassischen Rollenklischees zeigten, während die andere Hälfte neutrale Werbefilme sahen. Nur Frauen, die die traditionellen Werbefilme sahen, waren den männlichen Versuchsteilnehmern im folgenden Mathematiktest unterlegen, der als Non-diagnostic präsentiert wurde. Die schlechtere Leistung, so die Autoren, sei eine Konsequenz des aktivierten Stereotyps.

Bei Gonzales, Blanton und Williams (2002) zeigten Frauen bei einem Test, der Mathematikaufgaben und Aufgaben zur räumlichen Wahrnehmungsfähigkeit umfasste, unter Stereotype Threat signifikant schlechtere Leistung als Frauen in der Non-diagnostic-Gruppe beziehungsweise Männer in beiden Bedingungen.

Den Einfluss der erhöhten Zugänglichkeit von negativen stereotypen Erwartungen untersuchte eine Studie von Keller (2002), bei der die Hälfte der Versuchspersonen vor dem Bearbeiten eines Mathematiktests den Hinweis erhielt, dass der Test in der Vergangenheit einen Leistungsvorsprung zugunsten der männlichen Personen gezeigt habe. Die weiblichen Versuchspersonen schnitten in der Experimentalbedingung signifikant schlechter ab als die Mädchen, die diesen Hinweis nicht erhielten (Kontrollbedingung). Darüber hinaus zeigte sich in der offensichtlichen Threat-Bedingung ein signifikanter Leistungsunterschied zugunsten der männlichen Testteilnehmer, während in der Kontrollbedingung die Leistungsdifferenz zwischen Mädchen und Jungen nicht signifikant war. Da diese Studie den Stereotype-Threat-Effekt in einer natürlichen Umgebung untersuchte – die Versuchsteilnehmer waren Schüler zweier Gymnasien – scheint der Effekt von Stereotype Threat einerseits nicht nur auf das Labor, andererseits nicht nur auf Personen, die sich durch ein starkes Mathematikinteresse auszeichnen, beschränkt.

In einer Untersuchung von Marx und Roman (2002) konnte der durch Stereotype Threat verursachte Leistungsabfall von Frauen beim Bearbeiten eines Mathematiktests verhindert werden, wenn eine Frau als Versuchsleiterin fungierte. Nach Marx und Roman dient die (mathematisch kompetente) Versuchsleiterin den Frauen als positives Rollenmodell, das vor der Bedrohung des Stereotyps „Frauen können keine Mathematik“ schützt. Die entscheidende Vari-

able ist dabei die angenommene mathematische Kompetenz der Versuchsleiterin. Frauen, die mit einer Versuchsleiterin mit hoher mathematischer Kompetenz konfrontiert wurden, zeigten signifikant bessere Ergebnisse als Frauen, die einer Versuchsleiterin mit vermeintlich niedrigen mathematischen Fähigkeiten ausgesetzt waren. Interessanterweise zeigten die Männer die besten Resultate, wenn sie annehmen mussten, die Versuchsleiterin sei nicht besonders kompetent in Mathematik.

Bei Schmader (2002) zeigte sich bei einem angekündigten Leistungsvergleich zwischen Männern und Frauen, dass „women with higher levels of gender identification performed worse than men, but women with lower levels of gender identification performed equally to men. When gender identity was not linked to test performance, women performed equally to men regardless of the importance they placed on gender identity” (Schmader, 2002, S. 194). Demnach scheint die Identifikation mit der eigenen Geschlechtsgruppe ein möglicher Moderator des Stereotype-Threat-Effekts zu sein.

Bei vielen Untersuchungen wurde den Versuchsteilnehmern gesagt, der Test habe in der Vergangenheit geschlechtsspezifische Leistungsunterschiede gezeigt, während der anderen Hälfte mitgeteilt wurde, bei der Zusammenstellung der Aufgaben seien bisher keine geschlechtsspezifischen Leistungsunterschiede beobachtet worden (Fairnessbedingung). Die erste Variante bezeichnen Smith und White (2002) als „explicitly activated stereotype“, die zweite Variante als „nullified stereotype“. Welche Konsequenzen hat Stereotype Threat, wenn den Personen keine Informationen über bisherige Leistungsunterschiede gegeben werden (implicitly activated stereotype)? Sowohl unter der „explicitly activated stereotype“-Bedingung als auch unter der „implicitly activated stereotype“-Bedingung zeigten Frauen im Vergleich zur Fairnessbedingung (nullified stereotype) eine geringere Mathematikleistung. Die Autoren betonen: „... for the first time, it is shown that both explicitly and implicitly activated stereotypes are equally harmful to performance“ (Smith & White, 2002, S. 184).

Brown und Pinel (2003) konnten belegen, dass individuelle Leistungsunterschiede von Frauen bei einem Mathematiktest unter Stereotype Threat auf ein unterschiedlich ausgeprägtes Stigmabewusstsein (Stigma Consciousness) zurückzuführen sind. Nur die weiblichen Testteilnehmer mit einem hohen Stigma Consciousness zeigten unter Stereotype Threat eine schlechtere Mathematikleistung, während sich Frauen mit einem niedrigen Stigma Consciousness nicht von den Versuchsteilnehmern in der No-Threat-Bedingung unterschieden.

In einer Untersuchung von Cadinu et al. (2003) wurden die weiblichen Testpersonen zufällig einer von drei Bedingungen zugewiesen. In der ersten Bedingung wurde den Teilnehmern gesagt, sie seien in einem bestimmten Mathematiktest besser als die Männer, in der zweiten Bedingung wurde ihnen mitgeteilt, die Leistungen von Männern und Frauen seien gleich, und

in der dritten Bedingung wurde den weiblichen Versuchspersonen schließlich gesagt, sie seien schlechter als die Männer. Anschließend mussten die Personen ihr wahrscheinliches Testergebnis schätzen, ehe sie den Mathematiktest bearbeiteten. Bei den Frauen mit einer geringen Domainidentifikation zeigte sich unabhängig von der Information kein signifikanter Leistungsunterschied, während Frauen, denen Mathematik sehr wichtig war, je nach vorheriger Informationen unterschiedliche Testergebnisse erzielten: Personen, die eine negative Information erhielten, waren signifikant schlechter als Personen, denen eine positive Information mitgeteilt wurde. Dazwischen rangiert die Kontrollgruppe, der zuvor mitgeteilt wurde, dass Frauen und Männer sich nicht signifikant unterscheiden.

Keller und Dauenheimer (2003) haben Stereotype Threat in einer natürlichen Umgebung einem weiteren kritischen Test unterzogen. 74 Schüler zweier Realschulen bearbeiteten einen Mathematiktest, bei dem der einen Hälfte gesagt wurde, der Test habe in der Vergangenheit keine Leistungsunterschiede zwischen Jungen und Mädchen gezeigt, während der anderen Hälfte mitgeteilt wurde, dass in der Vergangenheit die mittlere Punktzahl der Jungen über dem mittleren Testergebnis der Mädchen lag. Die weiblichen Testteilnehmer versuchten in der No-Threat-Bedingung mehr Aufgaben zu lösen als die Mädchen in der Threat-Bedingung. In der Threat-Bedingung zeigten die Mädchen auch signifikant schlechtere Testergebnisse als die Jungen, während in der Fairnessbedingung keine geschlechtsspezifischen Leistungsunterschiede festzustellen waren. Darüber hinaus betonen die Autoren die Relevanz von Stereotype Threat im Unterrichtsalltag:

„Moreover, the fact that we did not select either participants with high domain identification or participants with a special interest or ability in the field of mathematics indicates that stereotype threat seems to affect the performance not only of persons characterized by high domain identification.“ (Keller & Dauenheimer, 2003, S. 378)

Nach den Untersuchungen von Keller (2002) und Keller und Dauenheimer (2003) lassen sich Stereotype-Threat-Effekte auch in nichtselektiven Stichproben nachweisen, wobei mit zunehmender Domainidentifikation wohl ein stärkerer Stereotype-Threat-Effekt verbunden ist. McIntyre, Paulson und Lord (2003) untersuchten den Einfluss von positiver Information auf Stereotype Threat. So zeigten Frauen bei einem Mathematiktest bessere Leistungen, wenn ihnen zuvor gesagt wurde, Frauen seien im Vergleich zu Männern die besseren Probanden. Bei einem zweiten Experiment las die Hälfte der Versuchspersonen Biografien von erfolgreichen Frauen in männlich dominierten Berufsfeldern, während die anderen Personen Abhandlungen über erfolgreiche Firmen zur Lektüre erhielten. Auch hier zeigten die Frauen in der

Successful-Women-Bedingung bessere Leistungen als die Frauen in der Successful-Corporations-Bedingung. Durch das Lesen positiver Information/Präsentation positiver Vorbilder scheint der Stereotype-Threat-Effekt seine leistungshemmende Wirkung zu verlieren.

Bei O'Brien und Crandall (2003) zeigten Frauen bei einem schwierigen Mathematiktest unter Stereotype Threat eine schlechtere Leistung als Frauen, die einen Test bearbeiteten, bei dem die Bedrohung durch Stereotype durch eine Fairnessmanipulation aus der Situation entfernt wurde. Interessanterweise erzielten weibliche Testpersonen unter Stereotype Threat bei einem einfachen Mathematiktest bessere Ergebnisse als Frauen unter der Non-Gender-Bedingung. Nach den Autoren führt Stereotype Threat zu einer stärkeren Erregung (Arousal), die bei leichten Aufgaben zu einer besseren Leistung, bei schwierigen Aufgaben jedoch zu einer schlechteren Leistung führe. Stereotype Threat führt demnach nicht per se zu einer verringerten Leistungsfähigkeit, sondern kann unter bestimmten Randbedingungen, beispielsweise bei Aufgaben mit einem geringen Schwierigkeitsgrad, zu einer Leistungssteigerung führen.

Nach Schmader und Johns (2003) führt Stereotype Threat zu einer verringerten Arbeitsspeicherkapazität (Working Memory Capacity), wobei Personen mit einer geringeren Arbeitsspeicherkapazität bei kognitiven Leistungstests schlechter abschneiden als Personen mit einem hohen Working Memory Capacity. Sie konnten zeigen, dass Frauen unter Stereotype Threat schlechter abschnitten als Frauen in der Kontrollbedingung (Non-diagnostic-Bedingung) und dass der Effekt über die Arbeitsspeicherkapazität vermittelt war.

Ford, Ferguson, Brooks und Hagadone (2004) konnten zunächst den klassischen Stereotype-Threat-Befund belegen: Frauen unter Stereotype Threat schneiden bei einem Mathematiktest schlechter ab als Frauen, bei denen die Bedrohung durch Stereotype durch eine Fairnessmanipulation aus der Situation entfernt wurde. Unter Stereotype Threat zeigen Frauen, die eine positive Grundstimmung haben (sense of humor), jedoch eine bessere Leistung und berichten über weniger Angst, als Frauen mit einem niedrigeren Level von sense of humor. Die Mediationsanalysen unterstützen die These, dass „coping sense of humor would buffer women from the effects of stereotype threat by predisposing them to experience less performance-impairing anxiety“ (Ford et al., 2004, S. 650).

Fazit: In zahlreichen Studien wurde gezeigt, dass durch die Aufhebung von Stereotype Threat die Testleistung von Frauen bei Mathematikaufgaben beziehungsweise Aufgaben der räumlichen Wahrnehmung signifikant verbessert wurde. Die geschlechtsspezifischen Leistungsunterschiede wurden dabei durch eine einfache Manipulation eliminiert. Einen Überblick über die berichteten Arbeiten gibt die folgende Tabelle:

<b>Autor</b>	<b>Testpersonen</b>	<b>Art der Aufgabe</b>	<b>Manipulation</b>
Sharps et al., 1993	Studierende	räumliches Vorstellungsvermögen	Testdiagnostizität
Sharps et al., 1994	Studierende	räumliches Vorstellungsvermögen	Testdiagnostizität
James/Greenberg, 1997	Studentinnen	räumliches Vorstellungsvermögen	Gruppenvergleich versus individuelle Leistung
Brown/Josephs, 1999	Studierende	Mathematiktest	Testdiagnostizität
Spencer et al., 1999	Studierende	Mathematiktest	Testfairness
Walsh et al., 1999 (Experiment 2)	Studierende	Mathematiktest	Gender-Difference und Testdiagnostizität
Inzlicht/Ben-Zeev, 2000 (Experiment 1)	Studierende	Mathematiktest/ Sprachtest	Salienz des Gruppenvergleichs
Oswald/Harvey, 2000/01	Studentinnen	Mathematiktest	Testfairness/Cartoon
Quinn/Spencer, 2001	Studierende	Mathematiktest	Testfairness
Davies et al., 2002	Studierende	Mathematiktest	Werbespots
Gonzales et al., 2002	Studierende	Mathetest/räumliches Vorstellungsvermögen	Testdiagnostizität
Keller, 2002	Schüler (m/w)	Mathematiktest	Testunfairness
Marx/Roman, 2002	Studierende	Mathematiktest	Kompetentes Rollenmodell
Schmader, 2002	Studierende	Mathematiktest	Gruppenvergleich versus individuelle Leistung
Smith/White, 2002 (Experiment 1)	Studentinnen	Mathematiktest	Testfairness
Brown/Pinel, 2003	Studentinnen	Mathematiktest	Testfairness
Cadinu et al., 2003	Studentinnen	Mathematiktest	Leistungsdifferenzen zwischen Gruppen
Keller/Dauenheimer, 2003	Schüler (m/w)	Mathematiktest	Testfairness
McIntry et al., 2003	Studierende	Mathematiktest	positives Rollenmodell
O'Brien/Crandall, 2003	Studierende	Mathematiktest	Testfairness
Schmader/Johns, 2003	Studierende	Mathematiktest	Testdiagnostizität
Ford et al., 2004 Experiment 2	Studentinnen	Mathematiktest	Testfairness

Tabelle 4-2: Experimentelle Arbeiten zum Stereotype-Threat-Effekt bei mathematischen Leistungstests (in chronologischer Reihenfolge)

Stereotype Threat ist, wie bereits erwähnt, nicht nur auf das Geschlecht und Mathematikaufgaben beschränkt, sondern bezieht sich auf alle Situationen, in denen negative stereotype Erwartungen angewendet werden können.

„Using these methods (or somewhat modified versions) for manipulating stereotype threat, a growing number of studies have demonstrated that (cognitive) abilities of dif-

ferent kinds (mathematical reasoning, verbal abilities, spatial abilities, affective sensitivity, problem solving, memory, logical reasoning, creative reasoning, negotiation skills, and golf ball putting) are strongly affected when the applicability of negative stereotypic expectations in the testing situation is manipulated.” (Keller, 2004, S. 14)

Die Leistungsreduzierung durch Stereotype Threat gilt als empirisch gesichert, unklar sind jedoch die vermittelten Prozesse zwischen Stereotype Threat und der niedrigen Testleistung. In so genannten Mediationsanalysen wurden zahlreiche potenzielle Mediatoren des Stereotype-Threat-Effekts untersucht, die sich weit gehend in drei Kategorien einordnen lassen: emotionale, kognitive und motivationale Variablen. Eine Übersicht über die potenziellen Mediatoren, die in der empirischen Forschung besondere Aufmerksamkeit gefunden haben, gibt die folgende Tabelle (für eine ausführliche Übersicht: Keller, 2004).

<b>kognitive Variablen</b>	<b>emotionale Variablen</b>	<b>motivationale Variablen</b>
Stereotype-Aktivierung	Angst	Anstrengung
Stereotype-Verdrängung	physiologische Erregung (Arousal)	Selbstwirksamkeit
Arbeitsspeicherkapazität	Niedergeschlagenheit	Evaluationsbefürchtung
Lösungsstrategien finden		Leistungserwartung

Tabelle 4-3: Übersicht über potenzielle Mediatoren des Stereotype-Threat-Effekts

Bereits in der klassischen Studie von Spencer et al. (1999) versuchten die Autoren Mediatoren zu identifizieren, über die der Stereotype-Threat-Effekt vermittelt wird. Weder Evaluationsbefürchtung noch Selbstwirksamkeit zeigten dabei signifikante Effekte, lediglich „anxiety is still a plausible mediator of the stereotype threat effects seen in these studies“ (Spencer et al., 1999, S. 21). Vor dem Hintergrund dieser Ergebnisse wurde Angst in der empirischen Forschung besondere Aufmerksamkeit geschenkt, doch die Befunde sind widersprüchlich: In einigen Studien zeigte sich ein stark vermittelnder Effekt der Angst, während in anderen Studien eine solche Beziehung nicht nachgewiesen werden konnte (Davies & Spencer, 2005). So konnte Osborne (2001) bei einer nationalen repräsentativen Stichprobe von High School Seniors (Abiturienten) zeigen, dass Angst zumindest teilweise die geschlechtsspezifischen Leistungsdifferenzen in Mathematik erklären kann, und auch Schmader und Johns (2003) fanden empirische Evidenz für die vermittelnde Rolle von Angst. Trotz dieser Befunde konnte der vermittelnde Einfluss von Angst in anderen Studien, beispielsweise Stone et al. (1999), nicht bestätigt werden. Dafür kann es verschiedene Ursachen geben:

„The timing of measurement might have missed actual increases in anxiety; study participants may have been unaware of their increased anxiety, and anxiety may mediate stereotype threat effects only under specific conditions.” (Steele et al., 2002, S. 400)

So untersuchte beispielsweise die Studie von Stone et al. (1999) den Stereotype-Threat-Effekt im sportlichen Bereich und nicht bei Mathematikaufgaben. Eine Erklärung für die widersprüchlichen Befunde liegt möglicherweise auch in der Messung des Angstniveaus. Viele Studien nutzen dafür „self-reported measures of anxiety. Research has confirmed that people may not always be willing or able to accurately report their own level of anxiety” (Davies & Spencer, 2005, S. 181). Bei einer Untersuchung von Blascovich, Spencer, Quinn und Steele (2001), bei der Weiße und Farbige einen verbalen Fähigkeitstest absolvierten, wurde der Blutdruck der Versuchspersonen kontinuierlich überwacht. Die Weißen waren den Farbigen überlegen, wenn der Test als diagnostisch präsentiert wurde, während sich in der Fairnessbedingung kein Leistungsunterschied zwischen den Gruppen zeigte. Interessanterweise berichteten die Farbigen unter der Threat-Bedingung nicht über mehr Angst, dafür hatten sie aber einen deutlich erhöhten Blutdruck. Bosson, Haymovitz und Pinel (2004) konnten zeigen, dass homosexuelle Männer unter Stereotype Threat ein stärkeres Ausmaß von nichtverbaler Angst aufwiesen, aber nicht über mehr Angst berichteten als homosexuelle Männer, die nicht durch Stereotype Threat bedroht waren. Diese Ergebnisse können zumindest dahin gehend interpretiert werden, dass die widersprüchlichen Befunde bei der Untersuchung von Angst als potenziellem Mediator des Stereotype-Threat-Effekts auf den Zeitpunkt der Messung und die Sensitivität des Messinstruments zurückzuführen sind (Davies & Spencer, 2005).

Keller und Dauenheimer (2003) konnten Niedergeschlagenheit als einen potenziellen Mediator identifizieren, der ebenso wie Angst der Gruppe der emotionalen Variablen zuzuordnen ist. Unter Kontrolle von Niedergeschlagenheit zeigte sich kein signifikanter Stereotype-Threat-Effekt mehr.

Bei den kognitiven Variablen wurden Stereotype-Aktivierung (stereotype activation), Stereotype-Verdrängung (stereotype suppression), Arbeitsspeicherkapazität (working memory capacity) und die Fähigkeit, geeignete Lösungsstrategien zu formulieren, als potenzielle Mediatoren genauer untersucht. Davies et al. (2002) fanden Evidenz für stereotype activation, Schmader und Johns (2003) für working memory capacity, und schließlich konnten Quinn und Spencer (2001) zeigen, dass bei Frauen unter Stereotype Threat die Fähigkeit herabgesetzt ist, geeignete Lösungsstrategien zu formulieren.

Bei den motivationalen Variablen, zu denen unter anderem Anstrengung, Selbstwirksamkeit, Evaluationsbefürchtung und Leistungserwartung gehören, konnten Spencer et al. (1999) weder für Evaluationsbefürchtung noch für Selbstwirksamkeit einen Effekt nachweisen, auch

scheint Stereotype Threat nicht die Anstrengung der Personen zu reduzieren (Steele et al., 2002); empirische Evidenz fand jedoch die Rolle von Leistungserwartungen (Cadinu et al., 2003).

Die bisherigen Mediationsanalysen haben zweifellos das Verständnis der Bedrohung durch Stereotype erweitert. Ein klares Bild, welche Prozesse dem Stereotype-Threat-Effekt zugrunde liegen, haben sie bisher nicht geliefert. Zweifellos ist Stereotype Threat eine komplexe Interaktion zwischen Person und Situation und „it is almost unimaginable that an experience as variable as this could be mediated consistently through a single psychological process within the person – always through anxiety, performance expectations, or stereotype activation, for example” (Steele et al., 2002, S. 406f). Noch deutlicher formuliert:

„Stereotype threat, like many social-psychological effects, may have multiple mediational paths, and the mediational path that emerges may depend on the specific situation or social identity being investigated.” (Davies, Spencer & Steele, 2005, S. 278)

Neben der Frage, über welche Prozesse der Effekt des Stereotype Threat vermittelt ist, wird in der aktuellen Literatur diskutiert, ob Stereotype Threat die geschlechtsspezifischen Leistungsunterschiede von Jungen und Mädchen beziehungsweise Frauen und Männern in der Mathematik erklären und eine Antwort auf die Frage geben kann, warum sich nur wenige Mädchen und junge Frauen für eine Ausbildung, ein Studium oder einem Beruf im mathematisch-naturwissenschaftlichen Bereich entscheiden.

Für die erste Frage gibt es klare empirische Evidenz: Unter Stereotype Threat zeigen Frauen eine schlechtere Mathematikleistung als Männer, während bei einer Entfernung der stereotypen Bedrohung aus der Situation kein geschlechtsspezifischer Leistungsunterschied zu erkennen ist. Die gesellschaftlich omnipräsente Konfrontation mit dem Stereotyp „Mädchen können keine Mathematik“ bleibt nicht ohne Konsequenzen. Neben dem unmittelbaren (negativen) Einfluss auf die Leistung kann Stereotype Threat langfristig zur Vermeidung oder sogar Deidentifikation mit dem mathematischen Bereich führen. Diese langfristige Strategie erscheint dabei als eine plausible Erklärung für einen männlich dominierten mathematischen und naturwissenschaftlichen Arbeitsmarkt.

„Disidentification is a defensive strategy of eliminating a domain as a long-term basis of self-evaluation. Obviously, this permanent strategy can lead to a systematic gender gap in math-related achievement and aspirations as women divest themselves from targeted domains. Some members of our society suggest the resulting gender gap accurately reflects sex-based abilities, but we believe this gender gap is simply the by-

product of a defensive effort by women to cope with a self-evaluative threat imposed on them by their own culture.” (Davies & Spencer, 2005, S. 183)

Die Genderlücke in Mathematik und verwandten Fächern als ein Nebenprodukt von Stereotype Threat? Stereotype Threat bezeichnet die Befürchtung, die negative stereotype Erwartung über die eigene Gruppe durch Versagen bei schwierigen Aufgaben zu bestätigen. Die wahrgenommene Befürchtung reduziert die Leistungsfähigkeit und die Person kann ihr Potential in einer gegebenen Leistungssituation nicht ausschöpfen. Langfristig könnte dies dazu führen, dass die Person den Leistungsbereich von ihrem Selbstwertgefühl entkoppelt, um den erhöhten Druck zu vermeiden, was wiederum die Motivation reduziert, sich in diesem Bereich zu engagieren. Folgt man dieser Argumentation, dann ist die Genderlücke in der Mathematik und verwandten Bereichen ein Resultat des in vielen Gesellschaften noch immer weit verbreiteten Stereotyps „Frauen können keine Mathematik“.

Ausgehend von der Frage, über welche Prozesse die Leistungsminderung von Stereotype Threat erfolgt, wird in der Literatur ein aktuelles Modell diskutiert, das die Integration der Stereotype-Threat-Theorie in die von E. Tory Higgins entwickelte Theorie des regulatorischen Fokus ermöglicht. Um die Stereotype-Threat-Theorie in einen größeren theoretischen Rahmen einordnen zu können, wird im nächsten Abschnitt die Theorie des regulatorischen Fokus kurz vorgestellt.

#### **4.5 Theorie des regulatorischen Fokus**

Die Theorie des regulatorischen Fokus von E. Tory Higgins (1997) hat seine Wurzeln im hedonistischen Prinzip, wonach der Mensch bestrebt ist, angenehme Zustände zu erreichen und unangenehme Zustände zu vermeiden. Um einen gewünschten Endzustand zu erreichen und einen unangenehmen Zustand zu vermeiden, schlagen Menschen unterschiedliche Wege ein, die Higgins in seiner Theorie als Promotion Focus und als Prevention Focus charakterisiert.

Die Theorie des regulatorischen Fokus bietet den Rahmen für die Einordnung des Stereotype-Threat-Effekts in einen breiteren theoretischen Kontext. In einem ersten Schritt wird zunächst die Theorie des regulatorischen Fokus kurz zusammengefasst, ehe ein Modell präsentiert wird, das Stereotype Threat und die Theorie des regulatorischen Fokus miteinander verknüpft.

##### **4.5.1 Theoretische Annahmen**

Higgins unterscheidet in seiner Theorie des regulatorischen Fokus zwei Möglichkeiten, angenehme Zustände zu erreichen und unangenehme Zustände zu vermeiden: Selbstregulation mit

Promotion Focus und Selbstregulation mit Prevention Focus. Diese zwei Formen der Selbstregulation werden jeweils durch bestimmte Variablen aktiviert.

„It is postulated that nurturance needs, ideals as relevant standards, and gain/non-gain situations induce self-regulation with a promotion focus, whereas security needs, oughts as relevant standards, and non-loss/loss situations elicit self-regulation with a prevention focus.“ (Keller, 2004, S. 39)

Higgins ordnet jedem regulatorischen Fokus spezifische Reaktionen zu. So soll der Promotion Focus mit einer erhöhten Sensibilität für die Anwesenheit oder Abwesenheit eines positiven Ergebnisses verknüpft sein und als spezifische Vorgehensweise eine Strategie der Annäherung auslösen, bei der Treffer sichergestellt und Auslassungsfehler vermieden werden. Das Vorgehen in einem Promotion Focus ist von Eifer geprägt. Schließlich soll sich eine gelungene Selbstregulation im Promotion Focus in Gefühlen der Fröhlichkeit, eine misslungene Selbstregulation in Emotionen der Niedergeschlagenheit widerspiegeln. Die Abbildung zeigt die Input- und Outputvariablen des Promotion Focus.

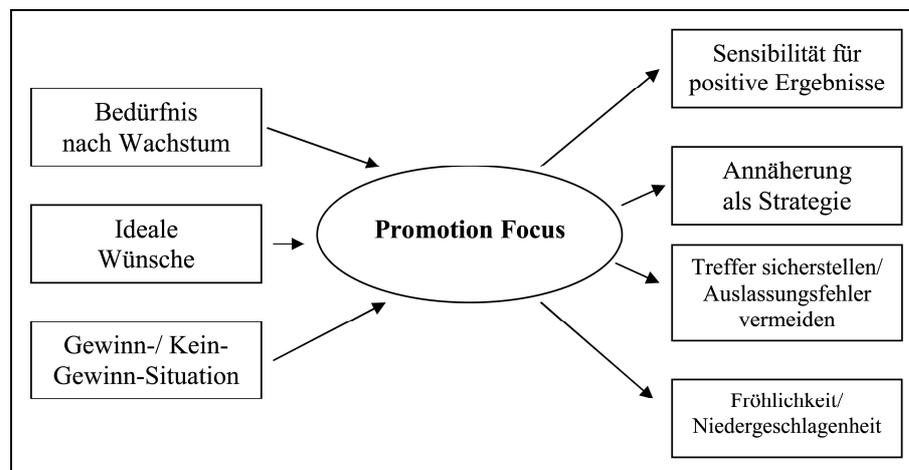


Abbildung 4-7: Promotion Focus (nach Higgins, 1998)

Der Prevention Focus ist dagegen durch eine besondere Sensibilität für das Eintreten oder Ausbleiben von negativen Ergebnissen gekennzeichnet. Die Annäherung zu einem gewünschten Endzustand erfolgt im Prevention Focus durch eine Strategie der Vermeidung, bei der Vorsicht und Sorgfalt dominiert. Im Prevention Focus liegt die Aufmerksamkeit darauf, Fehler zu vermeiden und falsche Alternativen zurückzuweisen. Eine gelungene Selbstregulation ist im Prevention Focus mit Gefühlen der Beruhigung verbunden, während eine missglückte Selbstregulation Aufgeregtheit beziehungsweise Nervosität zur Folge hat. Die Abbildung auf der nächsten Seite zeigt die Input- und Outputvariablen des Prevention Focus.

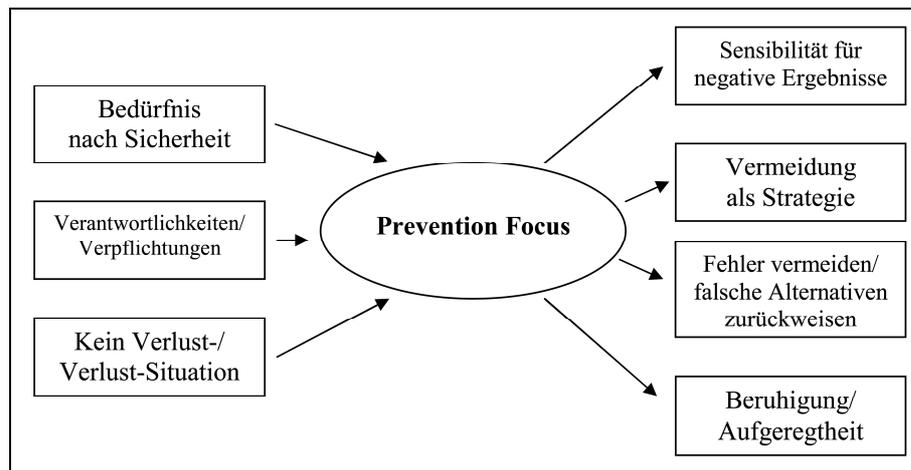


Abbildung 4-8: Prevention Focus (nach Higgins, 1998)

Die Selbstregulation im Promotion Focus oder Prevention Focus ist also immer mit kognitiven Prozessen (Sensibilität für positive oder negative Ergebnisse), motivationalen Prozessen (Strategien der Annäherung oder Vermeidung) und emotionalen Prozessen (Fröhlichkeit und Niedergeschlagenheit sowie Beruhigung und Aufgeregtheit) verbunden. Zahlreiche Studien stützen das Modell des regulatorischen Fokus (für einen Überblick: Higgins, 1998).

Der aktivierte regulatorische Fokus wirkt sich auch auf die Leistung einer Person aus, wobei keiner der beiden Fokusse generell eine bessere Leistung erbringt, da jeder Fokus mit unterschiedlichen Vor- und Nachteilen verbunden ist. Bei Aufgaben, die Genauigkeit erfordern, sind Personen im Prevention Focus besser, da diese sorgfältiger agieren, während bei Aufgaben, die Schnelligkeit benötigen, Personen im Promotion Focus besser abschneiden, da sich diese Personen durch einen besonderen Eifer auszeichnen. Auch zeigen Personen im Promotion Focus eine höhere Ausdauer, eine eventuell unlösbare Aufgabe zu lösen, während Personen im Prevention Focus diese Aufgabe eher auf sich beruhen lassen, um nicht zu viel Zeit zu verlieren. Wann welcher Fokus erfolgreicher ist, hängt entscheidend von den Kontextbedingungen ab (Werth, Denzler & Förster, 2002).

Die Zugänglichkeit zu einem der beiden regulatorischen Fokusse kann von Situation zu Situation variieren. Nach Higgins hat jeder Mensch aber eine Prädisposition für einen der beiden Fokusse, um erwünschte Endzustände zu erreichen beziehungsweise unerwünschte Endzustände zu vermeiden.

„The strength or accessibility of a regulatory focus, like any other kind of procedural knowledge, can vary chronically or momentarily.” (Higgins, 1998, S. 20)

Eine chronische Veranlagung zu einem regulatorischen Fokus erklären Higgins und Silberman (1998) durch unterschiedliche Erziehungsstile. Arbeiten Eltern vorwiegend mit Belohnung oder Belohnungsbezug, lernt das Kind das Eintreten positiver Ereignisse als angenehm

und deren Ausbleiben als unangenehm kennen. Zu diesem Annäherungsfokus passt es auch, wenn Eltern großen Wert auf die Verwirklichung der Potenziale ihrer Kinder legen. Ist dagegen Bestrafung das vorwiegende Mittel bei der Kindererziehung, lernt das Kind die Bestrafung als unangenehm und das Ausbleiben der Bestrafung als angenehm zu empfinden, was einem Vermeidungsfokus entspricht. Gefördert wird die Entstehung eines chronischen Prevention Focus, wenn das Einhalten von Verhaltensregeln und Normen eine große Rolle in der Erziehung spielt (Werth et al., 2002).

Zwar neigt der Mensch tendenziell eher einem der beiden Fokusse zu. Situationen, die zu einer erhöhten Zugänglichkeit zu einem Promotion Focus oder zu einem Prevention Focus führen, lassen sich aber auch ganz gezielt herstellen. So kann eine situative Zugänglichkeit zu einem regulatorischen Fokus beispielsweise durch die Aufgabenformulierung nahe gelegt werden. Bei einer Studie wurde die Salienz von Gewinnen (Promotion Focus) beziehungsweise die Salienz von Verlusten (Prevention Focus) experimentell manipuliert. Der einen Hälfte der Versuchspersonen wurde mitgeteilt, sie erhielten fünf Dollar für die Teilnahme an einer Untersuchung und, falls sie über dem 70. Perzentil abschneiden, würden sie zusätzlich einen Dollar erhalten, ansonsten bliebe es bei fünf Dollar. Durch die Instruktion sollte eine Gewinn-Situation beziehungsweise Keine-Gewinn-Situation geschaffen werden, die eine erhöhte Zugänglichkeit zu einem Promotion Focus auslösen sollte. Der anderen Hälfte wurde gesagt, sie erhielten sechs Dollar, wenn das Ergebnis über dem 70. Perzentil liegt; wenn sie jedoch unter dem 70. Perzentil liegen, dann würden sie einen Dollar verlieren. Die Versuchspersonen wurden in dieser Bedingung mit einer Kein-Verlust-Situation beziehungsweise Verlust-Situation konfrontiert, die mit einem Prevention Focus verknüpft ist. Nach der Aufgabe erhielten die Versuchspersonen unabhängig von der tatsächlichen Leistung ein manipuliertes Feedback, demnach sie die Aufgabe entweder erfolgreich oder nicht erfolgreich gemeistert hätten. Die Personen in einem Promotion Focus zeigten eher Gefühle der Fröhlichkeit beziehungsweise Niedergeschlagenheit als die Personen in einem Prevention Focus, während Probanden in einem Prevention Focus eher Emotionen der Beruhigung und Aufgeregtheit zeigten als die Personen in einem Promotion Focus. Je nach aktiviertem Fokus hatte das manipulierte Feedback unterschiedliche (emotionale) Konsequenzen (Higgins, 1998).

Die Theorie des regulatorischen Fokus bildet das Rahmenmodell für die Einordnung von Erwartungseffekten in einen breiteren theoretischen Kontext. Es wird im folgenden Abschnitt ein Modell vorgestellt, das die klassische Stereotype-Threat-Theorie in die von Higgins entwickelte Theorie des regulatorischen Fokus integriert. In Abhängigkeit des regulatorischen Fokus haben Erwartungen einen unterschiedlichen Effekt auf die Leistung einer Person.

#### 4.5.2 Das MERF-Modell

Das „Moderation of Expectancy Effects by Regulatory Focus“-Modell (MERF) postuliert „an interactive relationship between the type of expectancy (positive vs. negative) and the prevailing mode of self-regulation (promotion vs. prevention) in a given performance situation – that is, all else being equal, positive (and negative) expectations should have differential effects on performance depending on whether a promotion or a prevention focus is activated“ (Keller, 2004, S. 56f).

Nach dem MERF-Modell haben Erwartungen je nach aktiviertem regulatorischen Fokus einen unterschiedlichen Einfluss auf die Leistung. In einem Prevention Focus ist die Sensibilität für Verluste erhöht. Negative Erwartungen werden als Bedrohung gewertet, da die Person in der gegebenen Situation Verluste vermeiden möchte. Der aus dieser Bedrohung resultierende zusätzliche Druck führt zu einer geringeren Testleistung, während eine positive Erwartung im Prevention Focus nicht als Bedrohung interpretiert wird, ein zusätzlicher Druck deshalb ausbleibt und ein besseres Testergebnis erzielt werden kann. Im Promotion Focus stellt sich die Situation anders dar, da hier die Sensitivität für Gewinne betont wird. Eine positive Erwartung ist in einem Promotion Focus mit der Gefahr verbunden, einen möglichen Gewinn zu verpassen. Die sich daraus ergebende Bedrohung führt zu einer zusätzlichen Belastung, die einer guten Testleistung entgegensteht. Eine negative Erwartung wird im Promotion Focus dagegen nicht als Bedrohung interpretiert, eine bessere Leistung ist die Folge.

Nach dem MERF-Modell führt in einem Promotion Focus eine positive Erwartung zu schlechterer Leistung, in einem Prevention Focus zu besserer Leistung. Bei einer negativen Erwartung wird das gegenläufige Muster postuliert: Im Promotion Focus soll eine negative Erwartung zu besserer Leistung, im Prevention Focus zu schlechterer Leistung führen. Mit der Art der Selbstregulation sind darüber hinaus spezifische kognitive, strategische und emotionale Variablen verknüpft, die zwischen positiver und negativer Erwartung und Testleistung vermitteln. Die Abbildung gibt einen Überblick über die im MERF-Modell postulierten Zusammenhänge zwischen Selbstregulation und Art der Erwartung.

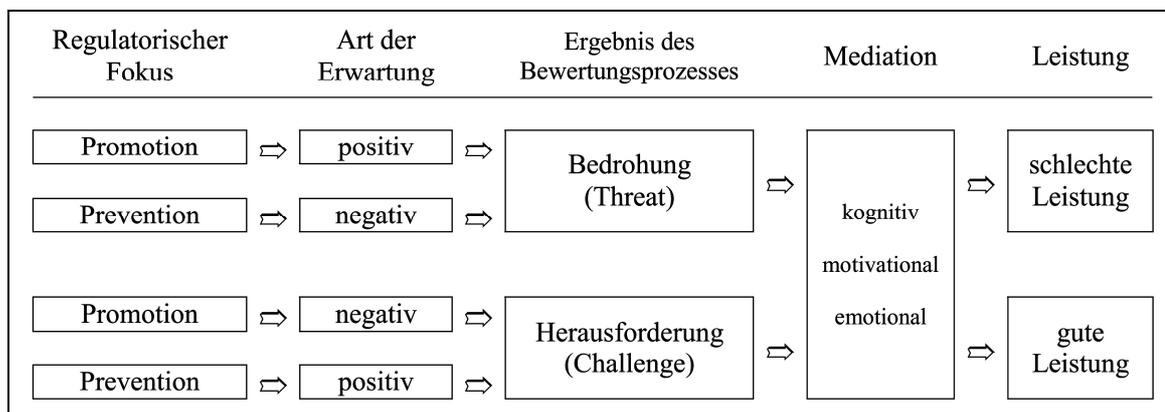


Abbildung 4-9: „Moderation of Expectancy Effects by Regulatory Focus“-Modell (nach Keller, 2004)

Der klassische Stereotype-Threat-Effekt kann als Spezialfall des „Moderation of Expectancy Effects by Regulatory Focus“-Modells interpretiert werden. Wenn sich eine Person in einer Situation befindet, in der die Möglichkeit besteht, einen Verlust zu erleiden, beispielsweise bei einem Mathematiktest, und negative Erwartungen auf das Verhalten der Person anwendbar sind, dann führt dies zu einer reduzierten Leistungsfähigkeit, die in Situationen, in der eine negative Erwartung nicht auf das Verhalten anwendbar ist, ausbleibt.

Stereotype Threat muss jedoch nicht per se als Einbahnstraße – im Sinne einer reduzierten Leistungsfähigkeit – wirken, da nach dem MERF-Modell eine negative Erwartung im Promotion Focus zu einer Leistungssteigerung führen sollte. Im Gegenzug ist eine positive Erwartung nicht zwangsläufig mit einer besseren Leistung verknüpft, da Bedingungen denkbar sind (Promotion Focus), bei der eine positive Erwartung zu einem Leistungseinbruch führen könnte.

Das MERF-Modell wurde in mehreren empirischen Studien überprüft (für einen Überblick: Keller, 2004). Eine Untersuchung soll im folgenden Abschnitt exemplarisch vorgestellt werden, ehe die mittlerweile klassische Studie von Brown und Josephs (1999) vor dem Hintergrund des „Moderation of Expectancy Effects by Regulatory Focus“-Modells nochmals diskutiert werden soll.

#### **4.5.3 Empirische Befunde**

Keller (2004) hat bei einer empirischen Überprüfung des MERF-Modells sowohl die Erwartungen als auch die Zugänglichkeit zu einem regulatorischen Fokus experimentell manipuliert. 179 Studentinnen der Universität Mannheim waren aufgefordert, einen Test zum räumlichen Vorstellungsvermögen zu bearbeiten. Die Probanden wurden dabei zufällig einer von drei Bedingungen des regulatorischen Fokus (Promotion, Prevention, Kontrollbedingung) und einer von zwei Bedingungen der Leistungserwartung (negativ versus positiv) zugewiesen. Um eine erhöhte Zugänglichkeit zu einem Promotion Focus zu manipulieren, haben die Personen folgende Instruktion erhalten.

„Um ein gutes Testergebnis zu erreichen, ist es sinnvoll, zu versuchen, möglichst viele Lösungen zu finden. Für jede richtig gelöste Aufgabe erhalten Sie einen Punkt, für falsche oder fehlende Aufgaben werden Ihnen keine Punkte abgezogen.“ (Keller, 2004, S. 144)

Durch diese Instruktion sollte die Sensibilität für Gewinne erhöht und eine Strategie, bei der Treffer sichergestellt und Auslassungsfehler vermieden werden, aktiviert werden – Variablen,

die mit einem Promotion Focus verknüpft sind. Um die Zugänglichkeit zu einem Prevention Focus zu erhöhen, lasen die Studentinnen folgende Instruktion:

„Um ein schlechtes Testergebnis zu vermeiden, ist es sinnvoll, sorgfältig vorzugehen und zu versuchen, Fehler zu vermeiden. Für jede richtig gelöste Aufgabe erhalten Sie einen Punkt, für jede falsche oder fehlende Aufgabe wird Ihnen ein Punkt abgezogen.“  
(Keller, 2004, S. 144)

Durch diese Beschreibung sollte die Sensibilität für Verluste erhöht und eine Strategie nahe gelegt werden, die sich durch das Vermeiden von Fehlern und das Zurückweisen falscher Alternativen auszeichnet. Sowohl die Sensibilität für negative Ergebnisse als auch Fehler vermeiden/falsche Alternativen zurückweisen sind mit einem Prevention Focus verbunden. Neben der Manipulation des regulatorischen Fokus wurde der Hälfte der Teilnehmerinnen mitgeteilt, bei dem Test zeigten sich geschlechtsspezifische Leistungsunterschiede, während der anderen Gruppe gesagt wurde, es zeigten sich keine Leistungsunterschiede zwischen Männern und Frauen.

Entsprechend den Annahmen des MERF-Modells zeigten Frauen in einem Prevention Focus mit einer negativen Erwartung eine schlechtere Leistung als Frauen im Prevention Focus mit einer positiven Erwartung beziehungsweise einer Fairnessbedingung. Dies entspricht den klassischen Befunden der Stereotype-Threat-Forschung. Das gegenläufige Muster zeigte sich im Promotion Focus: Frauen mit einer negativen Erwartung erzielten bessere Ergebnisse als Frauen mit einer positiven Erwartung. In der Kontrollbedingung (keine Manipulation der Zugänglichkeit zu einem regulatorischen Fokus) zeigten Frauen mit einer negativen Erwartung zwar eine schlechtere Leistung als Frauen mit einer positiven Erwartung, der Effekt war jedoch nicht signifikant (Keller, 2004).

Vor dem Hintergrund des MERF-Modells drängt sich eine Neuinterpretation der bereits berichteten Studie von Brown und Josephs (1999) auf, bei der weibliche Versuchspersonen bessere Leistungen erzielten, wenn ihnen zuvor gesagt wurde, der Mathematiktest identifiziere Personen, die über besonders gute mathematische Fähigkeiten verfügen, während sie schlechter abschnitten, wenn ihnen mitgeteilt wurde, durch den Test können Personen mit schlechten mathematischen Fähigkeiten erkannt werden. Bei den männlichen Probanden zeigte sich das gegenteilige Muster. Die Abbildung auf der folgenden Seite fasst die Ergebnisse der Studie zusammen.

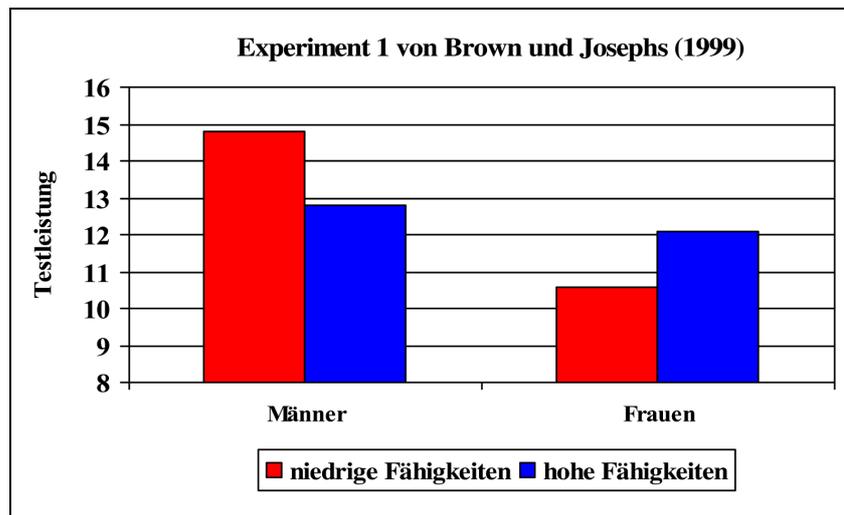


Abbildung 4-10: Testleistung in Abhängigkeit von Testbeschreibung und Geschlecht (nach Brown & Josephs, 1999)

Der Befund von Brown und Josephs (1999) lässt sich dahin gehend interpretieren, dass in der Bedingung, in der Personen identifiziert wurden, die über besonders gute mathematische Fähigkeiten verfügen, ein Promotion Focus induziert wurde, da eine Gewinn-Situation salient gemacht wurde, während in der Bedingung, in der Personen identifiziert wurden, die über niedrige mathematische Fähigkeiten verfügen, die Zugänglichkeit zu einem Prevention Focus erhöht wurde, da die Probanden in dieser Bedingung zumindest die Anerkennung mittelmäßiger mathematischer Fähigkeiten verlieren konnten. Nach dem MERF-Modell führen in einem Promotion Focus niedrige Erwartungen zu besseren Testleistungen als positive Erwartungen, während in einem Prevention Focus positive Erwartungen zu besseren Resultaten führen als negative Erwartungen. Das klassische Stereotyp „Frauen können keine Mathematik“ legt nahe, dass Frauen bei mathematischen Aufgaben eher negative Leistungserwartungen, Männer dagegen eher positive Erwartungen haben. Männer zeigen in einem Prevention Focus (niedrige Fähigkeiten) demnach bessere Leistungen als Männer in einem Promotion Focus (hohe Fähigkeiten), während bei Frauen das gegenteilige Muster zu beobachten ist.

Das MERF-Modell stellt eine gelungene theoretische Weiterentwicklung der Stereotype-Threat-Theorie dar und bietet durch die Integration mit der Theorie des regulatorischen Fokus die Möglichkeit, gezielt zu prüfen, über welche Prozesse die Effekte auf die Leistung vermittelt werden. Je nach aktiviertem regulatorischen Fokus haben positive und negative Erwartungen unterschiedlichen Einfluss auf die Testleistung. Durch den regulatorischen Fokus sind jeweils spezifische Variablen identifiziert, die zwischen positiver/negativer Erwartung und Testleistung vermitteln. Die konsistenten Befunde aus der Stereotype-Threat-Forschung lassen darüber hinaus den Schluss zu, dass sich Mädchen beziehungsweise Frauen bei Mathematiktests tendenziell eher in einem Prevention Focus befinden. Die empirischen Ergebnisse unterstreichen jedoch den starken Einfluss situationaler Aspekte auf die gezeigte Testleistung.

#### 4.6 Kurze Zwischenbetrachtung

Die empirischen Befunde zur Stereotype-Threat-Theorie beziehungsweise dem MERF-Modell stellen dem aus den biologischen und sozialisationstheoretischen Erklärungsansätzen zu geschlechtsspezifischen Leistungsdifferenzen in der Mathematik abgeleiteten eher pessimistischen Handlungsszenario, um Geschlechterdifferenzen in der Mathematik zu beeinflussen, ein optimistischeres Szenario gegenüber. Offensichtlich lassen sich geschlechtsspezifische Leistungsunterschiede in der Mathematik durch eine simple Manipulation der Testsituation aufheben, was die Bedeutung situativer Merkmale der Testsituation unterstreicht.

Der Leistungsanstieg bei Mädchen beziehungsweise Frauen in Bedingungen ohne Stereotype Threat lässt sich durch biologische Mechanismen oder psychosoziale Erklärungsfaktoren, beispielsweise das über mehrere Jahre erworbene autonome Lernverhalten, nicht erklären. Für geschlechtsspezifische Leistungsunterschiede sind demnach weniger stabile intrapersonale Merkmale biologischer oder psychosozialer Art als vielmehr situationale Merkmale der Testsituation entscheidend. Die im psychobiosozialen Modell durch biologische wie auch soziale Faktoren postulierten unterschiedlichen Entwicklungsvoraussetzungen für Mädchen und Jungen scheinen vor dem Hintergrund der empirischen Befunde der Stereotype-Threat-Theorie gering. Wenn durch eine einfache Manipulation der Testbeschreibung Leistungsdifferenzen von Frauen und Männern plötzlich verschwinden, scheidet ein durch die Wirkung von Sexualhormonen von Geburt an unterschiedlich verschaltetes Gehirn von Mädchen und Jungen als Erklärung für geschlechtsspezifische Leistungsunterschiede in der Mathematik eigentlich aus. Möglicherweise sind die Entwicklungsvoraussetzungen für Mädchen und Jungen in der Tat bis zu einem gewissen Grad unterschiedlich; die Annahme, einer weit gehend biologischen Determination geschlechtsspezifischer mathematischer Fähigkeiten ist vor dem Hintergrund der berichteten Studien jedoch nicht haltbar. Die Ausführungen zum Stereotype Threat lassen vielmehr vermuten, dass die situationalen Merkmale der Testsituation einen starken Einfluss auf die Leistungsfähigkeit der Personen haben.

Gegenüber dem Stereotype-Threat-Ansatz wird vielfach eingewendet, dass die Befunde auf kleinen Stichproben mit mathematisch hoch begabten Personen beruhen und deshalb nicht verallgemeinert werden dürfen. So argumentiert Arthur L. Whaley, dass „stereotype threat theory was tested in research on college populations. Research on college populations may be too narrow a base on which to rest social-psychological theories of human behavior“ (Whaley, 1998, S. 679). Diese Kritik wird häufig mit der Forderung nach Untersuchungen bei normalen Testbedingungen und mit repräsentativen Stichproben verbunden (Jensen, 1998), und „research has yet to demonstrate whether and to what degree this effect generalizes beyond the laboratory“ (Sackett, Schmitt, Ellingson & Kabin, 2001, S. 310).

Zwar konnten einige Studien, beispielsweise Keller (2002) oder Keller und Dauenheimer (2003), bereits zeigen, dass sich der Stereotype-Threat-Effekt auch im Schulalltag bei Personen nachweisen lässt, die nicht über ein hohes Maß an Domainidentifikation oder mathematischer Fähigkeit verfügen. Allerdings wurde Stereotype Threat bei diesen Studien absichtlich induziert, indem den Personen der Experimentalgruppe vorab mitgeteilt wurde, bei diesem Test zeigten sich geschlechtsspezifische Leistungsunterschiede. Welche Konsequenzen hat Stereotype Threat, wenn die Personen ohne vorherigen Hinweis auf geschlechtsspezifische Leistungsunterschiede einen Mathematiktest bearbeiten? Nach der berichteten Studie von Smith und White (2002) spielt es keine Rolle, ob das Stereotyp der geringeren mathematischen Begabung von Frauen explizit aktiviert wird. Sowohl in der „explicitly activated stereotype“-Bedingung als auch in der „implicitly activated stereotype“-Bedingung zeigten Frauen im Vergleich zur Fairnessbedingung (nullified stereotype) eine geringere Mathematikleistung. Der Forderung nach einer Untersuchung unter normalen Testbedingungen kommt diese Studie zwar nach, nicht jedoch der Kritik der selektiven Stichprobe, da es sich bei den Probanden um Psychologiestudentinnen handelte.

Im empirischen Teil dieser Arbeit soll deshalb geprüft werden, ob sich bei einer repräsentativen Stichprobe unter normalen Testbedingungen, das heißt, ohne experimentelle Variation der Testbeschreibung, Indizien finden lassen, die eine Erklärung der geschlechtsspezifischen Leistungsunterschiede in Mathematik durch Stereotype Threat nahe legen. Für die empirische Analyse wird dabei auf den internationalen PISA-Datensatz 2003 zurückgegriffen. In Anlehnung an Davies und Spencer (2005) wird dabei argumentiert, dass das Stereotyp „Mädchen beziehungsweise Frauen können keine Mathematik“ in vielen Gesellschaften noch weit verbreitet ist.

„Students taking the AP, SAT, or GRE are well aware of their diagnostic ability and their long history of gender differences. In short, real world settings have real world constraints that don't exist in the laboratory.” (Davies & Spencer, 2005, S. 178)

Wenn Davies und Spencer (2005) mit ihrer Annahme Recht haben, dann sollten die geschlechtsspezifischen Leistungsunterschiede in der internationalen Schulleistungsstudie PISA weitgehend auf Stereotype Threat zurückzuführen sein. Im nächsten Abschnitt sollen zunächst das Design und die Messinstrumente der PISA-Studie vorgestellt werden, ehe Hypothesen aus den theoretischen Annahmen der Stereotype-Threat-Theorie und dem MERF-Modell abgeleitet werden, die mit dem internationalen PISA-Datensatz überprüft werden können. Sind die empirischen Ergebnisse mit der Stereotype-Threat-Theorie und dem MERF-Modell kompatibel, würde dies die große Bedeutung von situationalen Aspekten bei Leistungstests eindrucksvoll unterstreichen.

## 5. Die PISA-Studie: Ein Überblick

Sind die Jugendlichen auf die Herausforderungen der Zukunft gut vorbereitet? Können sie ihre Ideen und Vorstellungen effektiv analysieren, begründen und auch kommunizieren? Haben sie die notwendigen Kompetenzen für ein lebenslanges Lernen? Diese Fragen soll das Programme for International Student Assessment (PISA) beantworten. PISA ist eine vergleichende Studie, welche Basiskompetenzen der nachwachsenden Generation erfassen soll und von der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD) durchgeführt wird. Nach der OECD werden mit PISA die Basiskompetenzen erfasst, „die in modernen Gesellschaften für eine befriedigende Lebensführung in persönlicher und wirtschaftlicher Hinsicht sowie für eine aktive Teilnahme am gesellschaftlichen Leben notwendig sind“ (Baumert, Stanat & Demmrich, 2001, S. 16). Im Mittelpunkt der Schulleistungsstudie stehen die Kompetenzen der Schüler in Mathematik, Lesen und Naturwissenschaften.<sup>11</sup>

Erstmals wurde die PISA-Studie im Jahr 2000 in 32 Ländern (darunter 28 OECD-Mitgliedsstaaten) durchgeführt, danach erfolgten die Erhebungen in einem Dreijahreszyklus. In jedem Zyklus wird ein Kompetenzbereich gründlicher getestet, dem dann zwei Drittel der Testzeit zugeteilt werden, während in den beiden anderen Bereichen jeweils nur zusammenfassende Leistungsprofile ermittelt werden. Die PISA-Studie 2003 wurde in 41 Ländern durchgeführt, darunter alle 30 OECD-Mitgliedsländern. Im Vergleich zu PISA 2000 wurde 2003 der Schwerpunkt von der Lesekompetenz auf die Mathematik verlagert, weshalb bei den empirischen Analysen auf den internationalen PISA-Datensatz 2003 zurückgegriffen wird. In der nächsten Dreijahreserhebung, PISA 2006, wird der Schwerpunkt auf der naturwissenschaftlichen Grundbildung liegen, bevor 2009 dann wieder die Lesekompetenz im Vordergrund stehen wird (OECD, 2004a).<sup>12</sup>

Die PISA-Studie beschränkt sich aber nicht darauf, Staaten in der Art eines olympischen Leistungswettbewerbs zu vergleichen und mit Rangplätzen auszuzeichnen, sondern erfasst auch

---

<sup>11</sup> Die Ergebnisse der PISA-Studie erzielen dabei nicht nur ein beachtliches Medienecho, sondern ihnen wird auch eine große politische Bedeutung zugesprochen. Vor dem Hintergrund der vorgezogenen Bundestagswahl ist der Auftraggeber für die Durchführung der PISA-Studie in Deutschland, die Ständige Konferenz der Kultusminister (KMK), an das PISA-Konsortium herangetreten, um den Veröffentlichungstermin der PISA-Erweiterungsstudie (PISA-E), die einen Ländervergleich der PISA-Studie ermöglicht, erneut abzustimmen. Ursprünglich sollten die Ergebnisse zum 15. September 2005 vorgelegt werden, also neun Monate nach der Veröffentlichung der Ergebnisse des internationalen Vergleichs. Zentrale Ergebnisse des Ländervergleichs der PISA-Erweiterungsstudie 2003 wurden nun bereits am 14. Juli präsentiert, da die KMK diese nicht erst drei Tage vor der Bundestagswahl bekannt geben wollte. Die ausführliche Darstellung der umfangreichen und differenzierten Befunde des Ländervergleichs erfolgt nun in zeitlichem Abstand zur Bundestagswahl auf einer Fachkonferenz am 3. November 2005 (Prenzel, Baumert, Blum, Lehmann, Leutner, Neubrand, Pekrun, Rost & Schiefele, 2005).

<sup>12</sup> Für eine ausführlichere Darstellung der Konzeption der PISA-Studien sei an dieser Stelle auf die Veröffentlichungen der OECD „Lernen für das Leben – Erste Ergebnisse von PISA 2000“ (2001) und „Lernen für die Welt von morgen. Erste Ergebnisse von PISA 2003“ (2004) sowie des Deutschen PISA-Konsortiums „PISA 2000 – Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich“ (2001), „PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland“ (2003) und des PISA-Konsortiums Deutschland „PISA 2003 – Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs“ (2004) verwiesen.

kontextbezogene Indikatoren, mit denen ein Zusammenhang zwischen den Leistungsergebnissen und den Schülermerkmalen hergestellt werden kann. Da der Erhebungsschwerpunkt im Jahr 2003 auf der Mathematik lag, wurden verstärkt Schülermerkmale erfasst, die für die Entwicklung der mathematischen Kompetenz bedeutsam erscheinen.

Durch die gleichzeitige Erfassung des mathematischen Leistungsvermögens der Schüler und der Hintergrundmerkmale (Interesse, Motivation, Selbstwirksamkeit) bietet die internationale PISA-Studie 2003 die Möglichkeit zu prüfen, ob sich die geschlechtsspezifischen Leistungsdifferenzen in der Mathematik mit der Stereotype-Threat-Theorie beziehungsweise dem „Moderation of Expectancy Effects by Regulatory Focus“-Modell erklären lassen. Die Kritikpunkte bisheriger Studien bezüglich Repräsentativität und Testbedingungen können mit der PISA-Studie überwunden werden, so dass die Analysen es ermöglichen, die Relevanz des Stereotype-Threat-Effekts im Schulalltag abzuschätzen.

Zunächst wird das Design der PISA-Studie 2003 (Stichprobe, Durchführung und Testbedingungen) kurz skizziert, ehe die Messinstrumente (abhängige und unabhängige Variablen) vorgestellt werden. Anschließend werden aus der Stereotype-Threat-Theorie und dem MERF-Modell Hypothesen abgeleitet, die mit dem internationalen PISA-Datensatz 2003 überprüft werden können, wobei an dieser Stelle auch auf die Grenzen des PISA-Datensatzes eingegangen werden muss.

## **5.1 Design der Studie**

Um die Vergleichbarkeit der Ergebnisse in den verschiedenen Ländern zu gewährleisten, muss PISA in vergleichbaren Zielpopulationen durchgeführt werden. Vor dem Hintergrund unterschiedlicher schulorganisatorischer Regelungen, beispielsweise Einschulungsalter oder Struktur des Bildungssystems, sind Definitionen der Zielpopulationen, die sich auf bestimmte Klassenstufen beziehen, für internationale Vergleiche ungeeignet (OECD, 2004a). Die Zielpopulation von PISA wird deshalb durch das Lebensalter bestimmt. Bei der PISA-Studie wurden Schüler untersucht, die zum Beginn des Testzeitraums zwischen 15 Jahren/drei Monaten und 16 Jahren/zwei Monate alt waren – unabhängig von der besuchten Jahrgangsstufe oder Art der Bildungseinrichtung.

Die Stichprobenauswahl erfolgte nach strengen Regeln und wurde auf internationaler Ebene überwacht. So war unter anderem eine Beteiligungsquote auf Schulebene von 85 Prozent und auf Schülerebene von 80 Prozent gefordert. Diese Qualitätsanforderungen wurden mit Ausnahme des Vereinigten Königreichs von allen Staaten erfüllt, so dass die PISA-Ergebnisse weitgehend die tatsächlichen Kompetenzen der 15-jährigen Schüler in den jeweiligen Ländern widerspiegeln.

Die PISA-Stichproben in den einzelnen Staaten wurden in der Regel nach dem Konzept der zweistufigen geschichteten Stichprobe konzipiert. In einem ersten Schritt wurde dabei eine Stichprobe einzelner Schulen gezogen, die von 15-jährigen Schülern besucht werden. Im zweiten Schritt des Auswahlprozesses wurden Schülerstichproben innerhalb der teilnehmenden Schulen gezogen. Dazu wurde jeweils eine Liste aller 15-jährigen Schüler erstellt. Aus dieser Liste wurden 35 Schüler mit gleicher Wahrscheinlichkeit ausgewählt; wenn es weniger als 35 Schüler gab, wurden alle 15-jährigen Schüler ausgewählt (OECD, 2004a).<sup>13</sup> Bei PISA ausgeklammert sind 15-Jährige, die nicht mehr zur Schule gehen (Prenzel et al., 2004).<sup>14</sup> In den OECD-Ländern schwankte der Stichprobenumfang zwischen 332 Schülern in Liechtenstein und 29.983 Schülern in Mexiko; insgesamt wurden bei der Schulleistungsstudie mehr als 250.000 Schüler getestet (OECD, 2004a).

Das Design für die internationale Erhebung sah einen Testtag vor, an dem die Schüler in Gruppen an ihrer Schule unter Aufsicht getestet wurden. Um die Testzeit und Testbelastung klein zu halten, dabei dennoch breite Kompetenzbereiche zu untersuchen, wurde bei PISA das Testdesign des Multi-Matrix-Sampling gewählt. Dabei werden nach bestimmten Prinzipien unterschiedliche Testhefte mit systematisch variierenden Aufgabenblöcken erstellt und zufällig den Schülern zugewiesen. Durch das Multi-Matrix-Sampling gewinnt man eine Datenmatrix, die mit der Item-Response-Theorie analysiert werden kann (Prenzel et al., 2004).<sup>15</sup>

Bei PISA 2003 wurden insgesamt 85 Mathematikaufgaben verwendet.<sup>16</sup> Diese Aufgaben können dabei vier mathematischen Inhaltsbereichen zugeordnet werden. Der Inhaltsbereich „Raum und Form“ bezieht sich auf räumliche und geometrische Phänomene und Zusammenhänge, die meist aus dem Stoffgebiet der Geometrie stammen. Der Bereich „Veränderung und Beziehung“ umfasst mathematische Darstellungen von Veränderungen sowie funktionalen

---

<sup>13</sup> Vereinfacht ausgedrückt handelt es sich bei den PISA-Stichproben jeweils um Klumpenstichproben, bei denen das Prinzip der einfachen Zufallsauswahl auf zusammengefasste Elemente (Schulen) angewendet wurde. Klumpenstichproben besitzen Vorteile gegenüber einer einfachen Zufallsstichprobe, wenn keine Liste der Elemente der Grundgesamtheit vorhanden ist, wohl aber eine Liste der zusammengefassten Elemente (Schnell, Hill & Esser, 1999).

<sup>14</sup> In fast allen Staaten befinden sich nahezu alle 15-Jährigen noch in einer Form der schulischen Ausbildung. Es gibt jedoch auch OECD-Staaten, bei denen der Anteil der 15-Jährigen, die sich in keiner schulischen Ausbildung befinden, fünf Prozent und mehr betragen, so in Australien (6,5 Prozent), Mexiko (41,9 Prozent), Österreich (5,7 Prozent), Portugal (5,8 Prozent), Spanien (7,9 Prozent) und der Türkei (46,2 Prozent). Da anzunehmen ist, dass eher leistungsschwächere Schüler mit 15 Jahren das Schulsystem verlassen haben, kann bei diesen Staaten mit einer Überschätzung der tatsächlichen Kompetenz gerechnet werden (Prenzel et al., 2004).

<sup>15</sup> Für eine Diskussion zum Multi-Matrix-Design und der IRT-Skalierung sei auf die Veröffentlichung von Carstensen, Knoll, Rost und Prenzel (2004) verwiesen. Ein Überblick zum Rasch-Modell bietet Rost (1999), eine mathematisch orientierte Einführung in die testtheoretischen Modelle und Verfahren bei der PISA-Studie liefern Lind und Knoche (2004).

<sup>16</sup> Beispielaufgaben finden sich in der OECD-Veröffentlichung „Lernen für die Welt von morgen. Erste Ergebnisse von PISA 2003“ (2004) und der Veröffentlichung des PISA-Konsortiums Deutschland „PISA 2003 – Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs“ (2004). Darüber hinaus sind Beispielaufgaben auf der Homepage des Leibniz-Instituts für die Pädagogik der Naturwissenschaften an der Universität Kiel zu finden ([www.pisa.ipn.uni-kiel.de](http://www.pisa.ipn.uni-kiel.de)).

Abhängigkeiten zwischen Variablen und steht der Algebra am nächsten. Der Inhaltsbereich „Quantitatives Denken“ erstreckt sich auf Zahlenphänomene sowie quantitative Relationen und Muster und ist eng mit der Arithmetik verbunden. Der Inhaltsbereich „Unsicherheit“ umfasst probabilistische und statistische Phänomene, die in der Schulmathematik unter der Rubrik „Wahrscheinlichkeitsrechnung“ behandelt werden. Die Leistungen in den vier Inhaltsbereichen fasst eine Gesamtskala zusammen (OECD, 2004a).

Die Aufgaben bei der PISA-Studie haben darüber hinaus unterschiedliche Antwortformate. In vielen Fällen mussten die Schüler eine Antwort auf die Frage selbst formulieren und dabei auch Rechnungen sowie Begründungen für ihre Ergebnisse liefern. Bei Aufgaben mit einem geschlossenen Antwortformat musste nur das Ergebnis angegeben werden, bei Aufgaben im Mehrfachwahlformat waren die Schüler aufgefordert, eine oder mehrere richtige Lösungen auszuwählen.

Die 85 Mathematikaufgaben entsprachen einer Testzeit von 210 Minuten und wurden auf halbstündige Aufgabenblöcke aufgeteilt. Jeder Schüler erhielt ein Testheft mit vier Aufgabenblöcken, die insgesamt einer Testzeit von zwei Stunden entsprachen. Hinzu kamen 30 Minuten für das Ausfüllen des Schülerfragebogens, mit dem Einstellungen, Engagement und Strategien erfasst wurden.

## 5.2 Messinstrumente

Als abhängige Variable wird bei den empirischen Analysen die erreichte Punktzahl bei dem PISA-Mathematiktest verwendet, wobei der Schwerpunkt auf der Gesamtskala liegt, die die Leistungen in allen vier Inhaltsbereichen zusammenfasst. Der OECD-Durchschnitt ist dabei auf 500 normiert worden, die Standardabweichung auf 100.<sup>17</sup> Nach einer Faustregel können Unterschiede in einer Größenordnung von 40 Punkten in einen Abstand von einem Schuljahr umgerechnet werden (Prenzel et al., 2004).

Dem Schülerfragebogen, den die teilnehmenden Schüler nach dem Mathematiktest bearbeitet haben, sind die unabhängigen Variablen entnommen. Von besonderem Interesse sind dabei die selbstbezogenen Kognitionen in Bezug auf Mathematik, die einzelnen Konstrukte wurden

---

<sup>17</sup> Die individuellen Leistungswerte der Schüler werden bei PISA mit so genannten Plausible Values berechnet. Plausible Values tragen dem Sachverhalt Rechnung, dass die Schüler jeweils nur einen Teil der Aufgaben bearbeitet haben und die Schätzungen von Personenparametern daher mit Unsicherheit behaftet sind. Bei PISA werden für jede Person fünf solcher Plausible Values gezogen (Carstensen et al., 2004). Bei den empirischen Analysen wurden die einzelnen Plausible Values zunächst addiert und anschließend durch fünf geteilt. Der arithmetische Mittelwert der neuen Variable unterscheidet sich aber nur marginal vom Mittelwert der einzelnen Plausible Values. Zur Validierung der empirischen Ergebnisse wurden die multiplen Regressionsmodelle zusätzlich mit den einzelnen Plausible Values berechnet.

dabei über mehrere Fragen erfasst.<sup>18</sup> Aus den Antworten wurden gemäß der Item-Response-Theorie Skalenwerte berechnet (Pekrun & Zirngibl, 2004).

*Interesse und Freude an Mathematik:* Die Schüler wurden zu ihrem Interesse an Mathematik als Fach sowie ihrer Freude am Mathematiklernen befragt.<sup>19</sup> Die Reliabilität<sup>20</sup> der Skala (Cronbach's Alpha) – ermittelt auf der Basis der deutschen PISA-Stichprobe – beträgt 0.90.

*Instrumentelle Motivation in Mathematik:* Die Schüler wurden gefragt, inwieweit sie durch externe Belohnungen, beispielsweise die Notwendigkeit von mathematischen Fähigkeiten für das spätere Berufsleben, zum Lernen ermutigt werden.<sup>21</sup> Cronbach's  $\alpha$  – berechnet auf der Basis der deutschen PISA-Stichprobe – beträgt 0.82.

*Selbstwirksamkeit in Mathematik:* Die Schüler wurden gefragt, inwieweit sie an ihre eigenen Fähigkeiten zur effizienten Bewältigung von Lernsituationen in Mathematik unter Überwindung von Schwierigkeiten glauben.<sup>22</sup> Bei der wahrgenommenen Selbstwirksamkeit handelt es sich um Überzeugungen, spezifische Handlungen im mathematischen Bereich erfolgreich ausführen zu können. Die Reliabilität der Skala beträgt – ermittelt auf der Basis der deutschen PISA-Stichprobe – 0.80.

---

<sup>18</sup> Die von der OECD mit Hilfe einer gewichteten Maximum-Likelihood-Schätzung gebildeten Indizes wurden mit einer explorativen Faktorenanalyse überprüft. Die Ergebnisse entsprachen weitgehend den OECD-Resultaten. Für die theoretischen Überlegungen zur Konstruktion der Indizes und Einzelheiten über die gewählten Methoden sei auf die Veröffentlichung der OECD (2004a) verwiesen.

<sup>19</sup> Der PISA-Index des Interesses und der Freude an Mathematik wurde von den Angaben über die Zustimmung der Schüler zu den folgenden Aussagen abgeleitet: „Ich mag Bücher über Mathematik“, „Ich freue mich auf meine Mathematikstunden“, „Ich mache Mathematik, weil es mir Spaß macht“ und „Mich interessiert das, was ich in Mathematik lerne“. Es wurde eine Vierpunkteskala mit folgenden Antwortkategorien verwendet: „Stimmt ganz genau“, „Stimmt eher“, „Stimmt eher nicht“ und „Stimmt überhaupt nicht“. Alle Aussagen wurden für die Skalierung umgepolt und höhere Werte auf diesem Index weisen auf ein größeres Interesse und Freude an Mathematik hin (OECD, 2004a).

<sup>20</sup> Die Reliabilität gibt den Grad der Genauigkeit an, mit der das Messinstrument ein bestimmtes Merkmal, beispielsweise das Interesse an der Mathematik, misst. Zur Bestimmung der Reliabilität eines Messinstruments gibt es mehrere Ansätze, in dieser Arbeit wird Cronbachs Alpha-Koeffizient verwendet (Bühner, 2004). Cronbachs Alpha kann Werte zwischen 0 und 1 annehmen, empirische Werte über 0.8 werden als akzeptabel betrachtet. In der Praxis werden meist noch niedrigere Koeffizienten akzeptiert (Schnell et al., 1999).

<sup>21</sup> Der PISA-Index der instrumentellen Motivation in Mathematik wurde von den Angaben über die Zustimmung der Schüler zu folgenden Aussagen abgeleitet: „Ich gebe mir in Mathematik Mühe, weil es mir in meinem späteren Job weiterhelfen wird“, „Mathematik zu lernen lohnt sich, weil es mir bei anderen Fächern in den höheren Schulklassen helfen wird“, „Mathematik ist für mich ein wichtiges Fach, weil ich es für mein späteres Studium brauche“ und „Ich werde viele Dinge in Mathematik lernen, die mir dabei helfen werden, einen Job zu bekommen“. Es wurde eine Vierpunkteskala mit folgenden Antwortkategorien verwendet: „Stimmt ganz genau“, „Stimmt eher“, „Stimmt eher nicht“ und „Stimmt überhaupt nicht“. Alle Aussagen wurden für die Skalierung umgepolt und größere Werte auf diesem Index weisen auf ein höheres Niveau an instrumenteller Motivation in Mathematik hin (OECD, 2004a).

<sup>22</sup> Der PISA-Index der Selbstwirksamkeit in Mathematik wurde von den Angaben über das Vertrauen der Schüler abgeleitet, folgende Mathematikaufgaben lösen zu können: „Anhand eines Zugfahrplans ausrechnen, wie lange die Fahrt von einem Ort x zu einem Ort y dauern würde“, „Ausrechnen, wie viel billiger ein Fernseher bei 30 Prozent Rabatt wäre“, „Ausrechnen, wie viele Quadratmeter Fliesen gebraucht werden, um einen Fußboden damit auszulegen“, „Diagramme in Zeitungen verstehen“, „Eine Gleichung wie  $3x + 5 = 17$  lösen“, „Auf einer Karte mit einem Maßstab von 1:10.000 die tatsächliche Entfernung zwischen zwei Orten bestimmen“, „Eine Gleichung wie  $2(x+3) = (x+3)(x-3)$  lösen“ und „Den Benzinverbrauch eines Autos berechnen“. Es wurde eine Vierpunkteskala mit den Antwortkategorien „Sehr sicher“, „Sicher“, „Nicht sehr sicher“ und „Gar nicht sicher“ verwendet. Alle Aussagen wurden für die Skalierung umgepolt und größere Werte auf diesem Index weisen auf ein höheres Niveau an Selbstwirksamkeit hin (OECD, 2004a).

*Selbstkonzept in Mathematik:* Beim mathematischen Selbstkonzept handelt es sich um Überzeugungen zum eigenen Leistungsvermögen in der Mathematik. Die Reliabilität des PISA-Index – berechnet auf der Basis der deutschen PISA-Stichprobe – liegt bei 0.91.<sup>23</sup>

*Mathematikangst:* Die Schüler wurden gefragt, inwieweit sie sich im Umgang mit Mathematikaufgaben emotional gestresst fühlen.<sup>24</sup> Es hat sich gezeigt, dass sich Mathematikangst negativ auf das Leistungsniveau auswirkt. Statt aufgabengerechte Kognitionen zu verarbeiten, sind Schüler mit Mathematikangst häufig mit für die Aufgabenbewältigung irrelevanten Kognitionen und emotionalem Stress beschäftigt. Beides reduziert die Fähigkeit zur korrekten Aufgabenbewältigung und hat ein schwächeres Leistungsniveau zur Folge. Cronbach's  $\alpha$  – berechnet auf der Basis der deutschen PISA-Stichprobe – beträgt für den PISA-Index Mathematikangst 0.86.

Neben den selbst bezogenen Kognitionen in Bezug auf Mathematik misst die OECD den Lernstrategien der Schüler eine große Bedeutung bei (OECD, 2004a), weshalb sie bei den Analysen als Kontrollvariablen berücksichtigt werden. Die Lernstrategien werden in Memorier- und Aufsaagestrategien, Elaborationsstrategien und Kontrollstrategien unterteilt.

*Memorier- und Aufsaagestrategien:* Die Schüler wurden über die Verwendung von Lernstrategien in Mathematik befragt, die Repräsentation von Kenntnissen und Verfahren mit sich bringen, die unverarbeitet oder nicht weiter verarbeitet im Gedächtnis gespeichert werden.<sup>25</sup> Die Reliabilität der Skala – ermittelt auf der Basis der deutschen PISA-Stichprobe – beträgt 0.68.

---

<sup>23</sup> Der PISA-Index zum mathematischen Selbstkonzept wurde vom Niveau der Zustimmung der Schüler zu folgenden Aussagen abgeleitet: „Ich bin einfach nicht gut in Mathematik“, „Im Fach Mathematik bekomme ich gute Noten“, „In Mathematik lerne ich schnell“, „Ich war schon immer überzeugt, dass Mathematik eines meiner besten Fächer ist“ und „Im Mathematikunterricht verstehe ich sogar die schwierigsten Aufgaben“. Es wurde eine Vierpunkteskala mit den Antwortkategorien „Stimmt ganz genau“, „Stimmt eher“, „Stimmt eher nicht“ und „Stimmt überhaupt nicht“ verwendet. Bis auf die erste Aussage wurden die Items für die Skalierung umgepolt und größere Werte weisen auf ein höheres Selbstkonzept in Mathematik hin (OECD, 2004a).

<sup>24</sup> Der PISA-Index der Mathematikangst wurde von den Angaben über die Zustimmung der Schüler zu den folgenden Aussagen abgeleitet: „Ich mache mir oft Sorgen, dass es für mich im Mathematikunterricht schwierig sein wird“, „Ich bin sehr angespannt, wenn ich Mathematikhausaufgaben machen muss“, „Beim Lösen von Aufgaben in Mathematik werde ich ganz unruhig“, „Beim Lösen von Mathematikaufgaben fühle ich mich hilflos“ und „Ich mache mir Sorgen, dass ich in Mathematik schlechte Noten bekomme“. Es wurde eine Vierpunkteskala mit den Antwortkategorien „Stimmt ganz genau“, „Stimmt eher“, „Stimmt eher nicht“ und „Stimmt überhaupt nicht“ verwendet. Alle Aussagen wurden für die Skalierung umgepolt und größere Werte auf diesem Index weisen auf ein höheres Niveau an Mathematikangst hin (OECD, 2004a).

<sup>25</sup> Der PISA-Index des Memorierens/Aufsagens wurde vom Niveau der Zustimmung der Schüler zu den folgenden Aussagen abgeleitet: „Manche Aufgaben in Mathematik rechne ich so oft durch, dass ich sie auch im Schlaf lösen könnte“, „Wenn ich für Mathematik lerne, lerne ich so viel wie möglich auswendig“, „Um mir den Lösungsweg einzuprägen, rechne ich die Mathematikaufgaben immer wieder durch“ und „Um für Mathematik zu lernen, versuche ich mir jeden einzelnen Lösungsschritt einzuprägen“. Es wurde eine Vierpunkteskala „Stimmt ganz genau“, „Stimmt eher“, „Stimmt eher nicht“ und „Stimmt überhaupt nicht“ verwendet. Alle Aussagen wurden für die Skalierung umgepolt und größere Werte weisen auf eine stärkere Präferenz der Lernstrategie des Memorierens/Aufsagens in (OECD, 2004a).

*Elaborationsstrategien:* Die Schüler wurden über den Einsatz von Lernstrategien befragt, bei denen neue Informationen mit bereits Gelerntem verknüpft werden.<sup>26</sup> Cronbach's  $\alpha$  – berechnet auf der Basis der deutschen PISA-Stichprobe – beträgt 0.73.

*Kontrollstrategien:* Die Schüler wurden über die Verwendung von Lernstrategien befragt, mit denen sie prüfen, was sie gelernt haben, und herauszufinden versuchen, was sie noch lernen müssen.<sup>27</sup> Die Reliabilität für diese Skala – ermittelt auf der Basis der deutschen PISA-Stichprobe – liegt bei 0.71.

### 5.3 Hypothesen

Bei der PISA-Studie wurde weder die Testbeschreibung noch die Zugänglichkeit zu einem regulatorischen Fokus experimentell variiert. Die situationalen Testbedingungen waren für die Schüler – von zufälligen Abweichungen vom vorgesehenen Reglement einmal abgesehen – identisch. PISA ist kein Experiment, folglich können keine Kausalschlüsse gezogen werden. Ein direkter (unmittelbarer) Test der Stereotype-Threat-Theorie oder des MERF-Modells ist mit den PISA-Daten grundsätzlich nicht möglich.<sup>28</sup>

---

<sup>26</sup> Der PISA-Index der Elaborationsstrategien wurde von den Angaben über die Zustimmung der Schüler zu folgenden Aussagen abgeleitet: „Bei Mathematikaufgaben überlege ich mir oft neue Lösungswege“, „Ich überlege mir, wie das, was ich in Mathematik gelernt habe, im Alltag angewendet werden kann“, „Neues in Mathematik versuche ich besser zu verstehen, indem ich Verbindungen zu Dingen herstelle, die ich schon kenne“, „Wenn ich eine Mathematikaufgabe löse, überlege ich oft, wie die Lösung für andere interessantere Fragestellungen verwendet werden könnte“ und „Wenn ich für Mathematik lerne, versuche ich den Stoff mit Dingen zu verbinden, die ich in anderen Fächern gelernt habe“. Es wurde eine Vierpunkteskala „Stimmt ganz genau“, „Stimmt eher“, „Stimmt eher nicht“ und „Stimmt überhaupt nicht“ verwendet. Alle Aussagen wurden für die Skalierung umgepolt und höhere Werte weisen auf eine stärkere Präferenz der Elaborationsstrategie hin (OECD, 2004a).

<sup>27</sup> Der PISA-Index der Kontrollstrategien wurde von den Angaben über die Zustimmung der Schüler zu den folgenden Aussagen abgeleitet: „Wenn ich mich für eine Mathematikprüfung vorbereite, versuche ich herauszufinden, welches die wichtigsten Dinge sind, die ich lernen muss“, „Wenn ich etwas für Mathematik lerne, zwingen mich zu prüfen, ob ich das Gelernte auch behalten habe“, „Wenn ich für Mathematik lerne, versuche ich herauszufinden, was ich noch nicht richtig verstanden habe“, „Wenn ich in Mathematik etwas nicht verstehe, suche ich nach zusätzlichen Informationen, um das Problem zu klären“ und „Wenn ich für Mathematik lerne, überlege ich mir zuerst, was genau ich üben muss“. Es wurde eine Vierpunkteskala „Stimmt ganz genau“, „Stimmt eher“, „Stimmt eher nicht“ und „Stimmt überhaupt nicht“ verwendet. Alle Aussagen wurden für die Skalierung umgepolt und höhere Werte weisen auf eine stärkere Präferenz der Kontrollstrategie hin (OECD, 2004a).

<sup>28</sup> Die PISA-Studie wird zwar alle drei Jahre wiederholt, die einzelnen Erhebungsrounden sind aber querschnittlich angelegt: Von jedem Teilnehmer liegen Daten vor, die zu einem einzelnen Zeitpunkt erhoben wurden. Eine solche Datenstruktur erlaubt es, Zusammenhänge von Variablen festzustellen, Schlüsse auf eine Verursachung der einen durch eine andere Variable sind methodisch problematisch, da vielfach auch die umgekehrte Bedingungsrichtung denkbar ist. Ein Beispiel: Die geringere Lesekompetenz der Jungen wurde bei der PISA-Studie 2000 vielfach auf das geringere Leseinteresse der Jungen zurückgeführt. Das geringere Leseinteresse habe also zu einer geringeren Lesekompetenz geführt. Die PISA-Daten lassen jedoch auch die umgekehrte Bedingungsrichtung zu: Geringere Lesekompetenzen haben dazu geführt, dass leistungsschwache Schüler ungern lesen, während leistungsstarke Schüler mehr Spaß am Lesen haben, weil sie über eine höhere Lesekompetenz verfügen, die ihnen das Lesen leichter macht. Theoretisch denkbar ist auch, dass es zwischen Lesekompetenz und Leseinteresse überhaupt keinen Zusammenhang gibt, vielmehr könnte sowohl das Leseinteresse als auch die Lesekompetenz durch eine dritte Variable verursacht sein, die in der Studie gar nicht erhoben wurde. Bei der Darstellung der empirischen Ergebnisse wird an geeigneten Stellen eine Bedingungsrichtung unterstellt, bei der Interpretation sollte man sich der methodischen Probleme einer Kausalinterpretation jedoch bewusst sein.

Die empirischen Analysen sind deshalb als eine Art Indizienprozess zu verstehen. Aus der Stereotype-Threat-Theorie und den Annahmen des MERF-Modells werden Hypothesen abgeleitet, die mit dem internationalen PISA-Datensatz 2003 überprüft werden können. Die einzelnen Hypothesen entsprechen dabei jeweils einem Indiz. Die Summe der Indizien soll ein Urteil darüber ermöglichen, ob eine Erklärung der geschlechtsspezifischen Leistungsunterschiede in der Mathematik durch Stereotype Threat beziehungsweise das MERF-Modell plausibel erscheint.

In Übereinstimmung mit Smith und White (2002) sowie Davies und Spencer (2005) kann der PISA-Mathematiktest als eine Situation angesehen werden, in der die teilnehmenden Schüler stereotype Erwartungen auf ihr Verhalten anwenden konnten. Diese Annahme wird durch das standardisierte Verfahren der PISA-Testsituation gestützt, demnach die Leiter der Testsitzungen nicht zugleich Fachlehrer der teilnehmenden Schüler in den Bereichen Mathematik, Lesen und Naturwissenschaften sein durften und darüber hinaus nicht Mitglied des Lehrerkollegiums einer Schule sein sollten, an der sie die Leitung des PISA-Tests übernahmen (OECD, 2004a). Konkret: Die PISA-Testsituation stellte für die Teilnehmerinnen eine Situation dar, bei der sie annehmen mussten, auf der Grundlage des in vielen Gesellschaften noch immer weit verbreiteten Stereotyps „Mädchen beziehungsweise Frauen können keine Mathematik“ beurteilt zu werden.

Nach der Stereotype-Threat-Theorie führt die wahrgenommene Befürchtung zu einer Leistungsminderung, die besonders bei den Personen, die sich mit Mathematik identifizieren, ausgeprägt ist. Daraus lässt sich folgende These ableiten, die mit dem internationalen PISA-Datensatz überprüft werden kann:

Bei weiblichen Testpersonen wird ein negativer Zusammenhang zwischen  
H 1: Mathematikinteresse und Testleistung erwartet. Je höher das Mathematikinteresse, desto geringer die Testleistung.

Der klassische Stereotype-Threat-Effekt zeigte sich bei experimentellen Studien nur bei schwierigen Leistungstests (Spencer et al., 1999). Deshalb stellt die Hypothese einen besonders kritischen Test dar, da bei PISA sowohl schwierige als auch leichte Mathematikaufgaben verwendet wurden. Sollte sich bei den weiblichen Testteilnehmern trotzdem ein negativer Zusammenhang zwischen der Mathematikleistung und dem mathematischem Interesse zeigen, würde dies die Relevanz von Stereotype Threat auch außerhalb von Laborbedingungen unterstreichen.

Auch aus dem MERF-Modell und der damit verbundenen Theorie des regulatorischen Fokus sind Hypothesen ableitbar, die mit dem internationalen PISA-Datensatz überprüft werden

können. Die Theorie des regulatorischen Fokus unterscheidet mit dem Promotion und dem Prevention Focus zwei Möglichkeiten, angenehme Zustände zu erreichen und unangenehme Zustände zu vermeiden. Nach Higgins gibt es für die Zugänglichkeit zu einem Fokus einerseits eine chronische Veranlagung, andererseits kann bei einer Person die Zugänglichkeit zu einem Fokus auch situational bedingt variieren. Da es sich bei PISA um eine standardisierte Leistungsmessung handelt, kann eine absichtliche Manipulation der Zugänglichkeit zu einem Fokus ausgeschlossen werden. Die erhobenen Schülermerkmale dienen in dieser Arbeit jedoch als ein Hinweis auf die chronische Prädisposition zu einem regulatorischen Fokus.<sup>29</sup>

So umfasst der PISA-Index zur instrumentellen Motivation Items, die die Notwendigkeit mathematischer Fähigkeiten im Schulalltag beziehungsweise im späteren Berufsalltag herausstellen. Die Fragen, die das Konstrukt instrumentelle Motivation erfassen sollen, legen nahe, dass eine mathematische Grundkompetenz als Pflicht beziehungsweise als Norm anzusehen ist, die es zu erfüllen gilt, um dem Bedürfnis nach Sicherheit – im Sinne eines Arbeitsplatzes – gerecht zu werden. Hohe Werte auf dem Index zur instrumentellen Motivation werden deshalb als ein Hinweis für eine chronische Zugänglichkeit zu einem Prevention Focus angesehen. Darüber hinaus stellte der PISA-Mathematiktest eine Situation dar, bei der die weiblichen Testteilnehmer mit einer negativen stereotypen Erwartung konfrontiert waren. Nach dem MERF-Modell wird in einem Prevention Focus eine negative Erwartung als Bedrohung gewertet. Der daraus resultierende Druck führt schließlich zu einer geringeren Testleistung. Gemäß dieser Argumentation lässt sich folgende These formulieren:

H 2: Bei weiblichen Testteilnehmern wird ein negativer Zusammenhang zwischen der instrumentellen Motivation (Indikator für chronische Zugänglichkeit zu einem Prevention Focus) und der Mathematikleistung erwartet. Je höher die instrumentelle Motivation, desto geringer die Punktzahl.

Bietet der PISA-Index zur instrumentellen Motivation die Möglichkeit, die Disposition zu einem Prevention Focus zumindest annäherungsweise zu erfassen, kann das mathematische Selbstkonzept als Indikator für die Orientierung an hoher mathematischer Fähigkeit interpretiert werden. Die Items, die das mathematische Selbstkonzept erfassen, zielen auf Eigenschaften ab, die eine Person idealerweise besitzen möchte (Beispiel: „Im Mathematikunterricht verstehe ich sogar die schwierigsten Aufgaben“). Der PISA-Index zum mathematischen Selbstkonzept beschreibt Hoffnungen, Wünsche und Sehnsüchte, die eine Zugänglichkeit zu

---

<sup>29</sup> Diese Interpretation ist aus konzeptioneller Sicht zweifellos nicht unproblematisch. Bei der PISA-Studie wurden jedoch keine Skalen verwendet, die eine chronische Disposition zu einem regulatorischen Fokus besser erfassen, weshalb in dieser Arbeit auf die Schülermerkmale zurückgegriffen wird.

einem chronischen Promotion Focus nahe legen. Nach dem MERF-Modell führen in einem Promotion Focus positive Erwartungen zu einer Leistungsreduktion. Da die männlichen Testteilnehmer beim PISA-Mathematiktest tendenziell mit einer positiven Erwartung konfrontiert waren, lässt sich folgende Hypothese formulieren:

Bei männlichen Testteilnehmern wird ein negativer Zusammenhang zwischen dem mathematischen Selbstkonzept (Indikator für chronische Zugänglichkeit zu einem Promotion Focus) und der Mathematikleistung erwartet. Je ausgeprägter das Selbstkonzept, desto geringer die Punktzahl beim PISA-Mathematiktest.

Für die Schülerinnen sollte sich dagegen kein negativer Zusammenhang zwischen dem mathematischen Selbstkonzept und der Mathematikleistung zeigen, da eine negative Erwartung im Promotion Focus nicht als Bedrohung interpretiert wird, ein zusätzlicher Druck ausbleibt und ein besseres Testergebnis erzielt werden kann.

Dieser zusätzliche Druck lässt sich als emotionale Angespanntheit beziehungsweise Stress charakterisieren. Statt aufgabengerechte Kognitionen zu verarbeiten, sind die Schüler mit für die Aufgabenbewältigung irrelevanten Kognitionen und emotionalem Stress beschäftigt. Beides reduziert die Fähigkeit, die Aufgabe korrekt zu lösen, und hat ein niedrigeres Leistungs niveau zur Folge. Der PISA-Index Mathematikangst erfasst die emotionale Angespanntheit der Schüler bei Mathematikaufgaben. Je wichtiger die Fähigkeit eingeschätzt wird – bei PISA über die Skala der instrumentellen Motivation erfasst –, desto stärker sollte die Person beim Lösen der Mathematikaufgaben emotional angespannt sein. Diese emotionale Anspannung führt zu einer geringeren Testleistung. Aus dieser Argumentation lässt sich folgende Hypothese ableiten:

Je wichtiger die mathematische Fähigkeit beurteilt wird, desto stärker war die Person beim PISA-Mathematiktest emotional angespannt. Es wird ein negativer Zusammenhang zwischen instrumenteller Motivation/Mathematikangst und Testleistung erwartet.

Diese Hypothesen stehen im Mittelpunkt des nun folgenden empirischen Teils dieser Arbeit. Es gilt zu prüfen, ob die Hypothesen bei den Analysen bestätigt werden können oder verworfen werden müssen. Lassen sich die geschlechtsspezifischen Leistungsunterschiede in der Mathematik durch Stereotype Threat beziehungsweise das MERF-Modell erklären? Das nächste Kapitel soll diese Frage beantworten.

## 6. Empirische Analysen

Im empirischen Teil dieser Arbeit soll geprüft werden, ob sich Indizien finden, die eine Erklärung der geschlechtsspezifischen Leistungsunterschiede in der Mathematik durch Stereotype Threat beziehungsweise das MERF-Modell nahe legen. Für die Analysen steht der internationale PISA-Datensatz 2003<sup>30</sup> zur Verfügung, als Methode dient die lineare (multiple) Regression. Dem Regressionsmodell werden schrittweise Variablen hinzugefügt, von denen vermutet wird, dass sie einen Einfluss auf die mathematische Leistung haben und/oder im Zusammenhang mit den Erklärungsansätzen stehen. Zunächst werden Regressionsmodelle mit den Daten der deutschen Schüler aus der internationalen Schulleistungsstudie berechnet, anschließend folgt der internationale Vergleich.

### 6.1 Deutschland

Bei der PISA-Studie 2003 beteiligten sich in der Bundesrepublik Deutschland 4660 Schüler aus 216 Schulen. Da 46 Personen keinem Geschlecht zuzuordnen waren, reduziert sich die Fallzahl auf 4614 Teilnehmer. Bei der PISA-Studie erzielten die deutschen Schüler eine durchschnittliche Leistung von 503 Punkten, die nur geringfügig vom OECD-Durchschnitt (500 Punkte) abweicht.

Der Startpunkt der Analyse ist eine Einfachregression mit der abhängigen Variable Mathematikleistung (Gesamtskala) und der unabhängigen Variable Geschlecht, die dummykodiert ist. Die Dummykodierung des Geschlechts folgt in dieser Arbeit einem Vorschlag von Fox. Die Referenzkategorie bilden die Mädchen (0), die Jungen erhielten bei der Rekodierung die 1 (Fox, 1999).

Mathematik	Koeffizient	t-Wert	p-Wert
Jungen	8.98	3.07	0.002
Konstante	498.90	240.76	0.000

Fallzahl: 4614

$r^2 = 0.0020$

Tabelle 6-1: Regressionsmodell 1 (Deutschland)

<sup>30</sup> Der internationale PISA-Datensatz 2003 steht auf der Homepage des internationalen PISA-Konsortiums zur Verfügung und ist unter der Adresse [http://pisaweb.acer.edu.au/oecd\\_2003/oecd\\_pisa\\_data.html](http://pisaweb.acer.edu.au/oecd_2003/oecd_pisa_data.html) als txt-Datei abrufbar. Die txt-Datei wurde mit einer ebenfalls auf der Homepage abrufbaren SPSS-Syntax in einen SPSS-Datensatz konvertiert. Im Gegensatz zu dem Statistikpaket SPSS lädt das Statistikprogramm Stata den Datensatz direkt in den Arbeitsspeicher des Computers, was einen schnelleren Datenzugriff und eine höhere Rechengeschwindigkeit gewährleistet. Deshalb wurde der knapp 500 Megabyte große SPSS-Datensatz mit dem Programm Stat/Transfer in einen Stata-Datensatz (Version 7.0) überführt. Mit dem Statistikprogramm Stata wurden weitgehend die Analysen durchgeführt. Auf der beiliegenden DVD sind der internationale PISA-Datensatz als txt-Datei und zugehöriger SPSS-Syntax sowie als SPSS- und Stata-Datensatz enthalten.

Die Jungen erzielten beim PISA-Mathematiktest durchschnittlich neun Punkte mehr als die Mädchen. Die Mädchen kamen beim Mathematiktest durchschnittlich auf 498.90 Punkte, während die Jungen 507.88 Punkte erreichen konnten. Diese Leistungsdifferenz ist statistisch abgesichert.<sup>31</sup> Die erklärte Varianz liegt mit einem  $r^2$  von 0.0020 deutlich unter einem Prozent. Insofern ist Janet Hyde zuzustimmen, die bereits bei einer Metaanalyse vor über 20 Jahren, das Geschlecht als schlechten Prädiktor für mathematische Leistungen bezeichnete (Hyde, 1981). Mit einem d-Wert von +0.09 ist der Effekt zwar klein, unter dem Gesichtspunkt der Chancengleichheit beider Geschlechter jedoch durchaus als bedeutsam zu bezeichnen (Zimmer, Burba & Rost, 2004). In einem zweiten Regressionsmodell wird neben dem Geschlecht auch das Mathematikinteresse (Interesse) berücksichtigt.

<b>Mathematik</b>	<b>Koeffizient</b>	<b>t-Wert</b>	<b>p-Wert</b>
Jungen	10.97	3.87	0.000
Interesse	9.00	7.11	0.000
Konstante	490.79	172.78	0.000

Fallzahl: 4379

$r^2 = 0.0175$

Tabelle 6-2: Regressionsmodell 2 (Deutschland)

Bei der Kontrolle des Mathematikinteresses steigt der Leistungsvorsprung der Jungen von neun auf elf Punkte leicht an. Es zeigt sich ein positiver Zusammenhang zwischen dem Mathematikinteresse und der Mathematikleistung: Steigt das Interesse bei dem PISA-Index um einen Punkt, dann nimmt die durchschnittliche Mathematikleistung um neun Punkte zu. Insgesamt erklärt das Interesse aber nur einen geringen Teil der Varianz. Mit einem  $r^2$  von 0.0175 werden nicht einmal zwei Prozent der gezeigten Mathematikleistung bei der PISA-Studie durch das Geschlecht und das Interesse erklärt.

<sup>31</sup> Die Schätzung von Populationskennwerten aus Stichproben ist immer fehlerbehaftet. Die übliche Berechnung des Standardfehlers führt nur unter der Voraussetzung einer einfachen Zufallsstichprobe, bei der die Elemente voneinander unabhängig sind, zu vertretbar genauen Schätzungen. Wie bereits berichtet, handelt es sich bei der PISA-Studie jedoch um eine Klumpenstichprobe, die Jugendlichen einer Schule werden sich (statistisch gesehen) weniger stark voneinander unterscheiden als Jugendliche aus verschiedenen Schulen (vgl. Kapitel 5.1). Zur Berechnung des Stichprobenfehlers werden bei PISA die komplexen Replikationsmethoden verwendet, die auf der Idee basieren, aus einer vorhandenen Stichprobe künstlich Stichprobenreplikationen zu bilden und die Varianz eines Stichprobenkennwertes über diese Replikationen hinweg zu berechnen (Carstensen et al., 2004). Konkret wird in der PISA-Studie das so genannte Balanced-repeated-Replication-Verfahren (BRR) in der Variante nach Fay mit Faktor 0.5 eingesetzt (E-Mail von Professor Claus H. Carstensen vom 3. Juni 2005). Zu den Details des BRR-Verfahrens in der Variante nach Fay sei auf Judkins (1990) verwiesen. Bei multivariaten Zusammenhangsanalysen ist dieses Verfahren jedoch nicht ohne weiteres anwendbar, mit unverhältnismäßig großem Aufwand verbunden oder durch die Verringerung von Freiheitsgraden extrem konservativ, so dass die Wahrscheinlichkeit der Entdeckung von Zusammenhängen stark verringert wird. In Übereinstimmung mit dem deutschen PISA-Konsortium wird in dieser Arbeit deshalb die übliche Prüfstatistik berichtet, die Unabhängigkeit der Stichprobenelemente voraussetzt. Dabei ist zu berücksichtigen, dass der  $\alpha$ -Fehler systematisch unterschätzt wird. Für die meisten multivariaten Analysen wird dieses Vorgehen als unproblematisch bezeichnet, da die Befunde auch auf der Basis der effektiven Stichprobengröße zufallskritisch abgesichert werden können (Sibbern & Baumert, 2001). Signifikanztests werden gemäß der üblichen Praxis in dieser Arbeit auf dem Signifikanzniveau von fünf Prozent durchgeführt.

Das nächste Regressionsmodell wird deshalb um die instrumentelle Motivation (wichtig) der Schüler erweitert.

<b>Mathematik</b>	<b>Koeffizient</b>	<b>t-Wert</b>	<b>p-Wert</b>
Jungen	13.33	4.66	0.000
Interesse	13.31	8.90	0.000
wichtig	-8.56	-5.02	0.000
Konstante	502.00	137.61	0.000

Fallzahl: 4371

$r^2 = 0.0242$

Tabelle 6-3: Regressionsmodell 3 (Deutschland)

Der Vorsprung der Jungen vergrößert sich bei Kontrolle des Interesses und der Beurteilung der Bedeutung der Mathematik von elf Punkten um zwei Punkte auf 13 Punkte. Der positive Zusammenhang zwischen Interesse und Mathematikleistung bleibt bestehen, während sich ein negativer Zusammenhang zwischen der instrumentellen Motivation und der Mathematikleistung zeigt. Steigt der PISA-Index instrumentelle Motivation um einen Punkt, dann sinkt die durchschnittliche Leistung um 8.5 Punkte. Eine solche kausale Annahme ist zwar mit den PISA-Daten eigentlich nicht möglich, sie erscheint aber plausibler als die umgekehrte Bedingungsrichtung, die eine niedrigere instrumentelle Motivation durch höhere mathematische Kompetenz unterstellt. Die erklärte Varianz ist mit 2.4 Prozent immer noch recht gering.

Im vierten Regressionsmodell wird das erhobene Schülermerkmal des mathematischen Selbstkonzepts berücksichtigt, das die Überzeugungen zur eigenen Leistungsfähigkeit in Mathematik erfasst.

<b>Mathematik</b>	<b>Koeffizient</b>	<b>t-Wert</b>	<b>p-Wert</b>
Jungen	5.72	2.05	0.040
Interesse	-8.17	-4.42	0.000
wichtig	-10.65	-6.47	0.000
Selbstkonzept	31.72	18.37	0.000
Konstante	478.13	127.50	0.000

Fallzahl: 4356

$r^2 = 0.0946$

Tabelle 6-4: Regressionsmodell 4 (Deutschland)

Das Selbstkonzept erhöht die aufgeklärte Varianz um rund sieben Prozentpunkte auf 9.46 Prozent. Mit den Variablen Geschlecht, Interesse, instrumentelle Motivation (wichtig) und Selbstkonzept sind knapp zehn Prozent der Varianz der Mathematikleistung bei den deutschen Schülern erklärt. Interessanter sind jedoch die negativen Zusammenhänge des Mathematikinteresses und der Punktzahl sowie der Bedeutung der Mathematik (wichtig) und der gezeigten Testleistung. Das mathematische Selbstkonzept zeigt dagegen einen positiven Zusammenhang mit der Mathematikleistung; steigt das Selbstkonzept um einen Punkt, dann erhöht sich die Leistung um 31 Punkte.

Im folgenden Regressionsmodell wird zusätzlich die Selbstwirksamkeit berücksichtigt. Bei der Selbstwirksamkeit handelt es sich um Überzeugungen, spezifische Handlungen erfolgreich ausführen zu können. Um die Selbstwirksamkeit in Mathematik zu erfassen, wurden den Testteilnehmern Aufgaben vorgelegt, bei denen sie einschätzen mussten, ob sie diese „sehr sicher“, „sicher“, „nicht sehr sicher“ oder „gar nicht sicher“ lösen können.

<b>Mathematik</b>	<b>Koeffizient</b>	<b>t-Wert</b>	<b>p-Wert</b>
Jungen	-5.28	-2.14	0.032
Interesse	-8.06	-4.96	0.000
wichtig	-12.61	-8.72	0.000
Selbstkonzept	11.52	7.12	0.000
Selbstwirksamkeit	50.70	35.65	0.000
Konstante	329.09	61.70	0.000

Fallzahl: 4344

$r^2 = 0.3006$

Tabelle 6-5: Regressionsmodell 5 (Deutschland)

Bei Berücksichtigung der Selbstwirksamkeit zeigt sich ein geringer Leistungsvorsprung der Mädchen, das heißt, bei Kontrolle des Interesses, der instrumentellen Motivation, des Selbstkonzepts und der Selbstwirksamkeit schneiden die Mädchen um fünf Punkte besser ab als die männlichen Testteilnehmer. Beachtlich ist der stark positive Zusammenhang der Selbstwirksamkeit und der erreichten Punktzahl im PISA-Mathematiktest. Steigt die Selbstwirksamkeit um einen Punkt, dann nimmt die Punktzahl um 50 Punkte zu. Auch bei Kontrolle der Selbstwirksamkeit bleibt der negative Zusammenhang des Interesses beziehungsweise der instrumentellen Motivation (wichtig) und der mathematischen Leistung bestehen. Das  $r^2$  erreicht mit einem Wert von 0.3006 einen neuen Höchststand, das Regressionsmodell erklärt rund 30 Prozent der Gesamtvarianz.

Beim sechsten Regressionsmodell wird das Modell um die Mathematikangst erweitert. Dabei wird ein negativer Zusammenhang zwischen der Mathematikangst und der Mathematikleistung erwartet, da Mathematikangst mit aufgabenirrelevanten Kognitionen verbunden ist, die eine korrekte Lösung der Aufgabe weniger wahrscheinlich machen.

<b>Mathematik</b>	<b>Koeffizient</b>	<b>t-Wert</b>	<b>p-Wert</b>
Jungen	-5.53	-2.28	0.023
Interesse	-7.62	-4.77	0.000
wichtig	-10.27	-7.17	0.000
Selbstkonzept	-4.14	-2.05	0.040
Selbstwirksamkeit	49.02	34.91	0.000
Mathematikangst	-19.48	-12.60	0.000
Konstante	408.57	49.72	0.000

Fallzahl: 4341

$r^2 = 0.3256$

Tabelle 6-6: Regressionsmodell 6 (Deutschland)

In der Tat ist ein negativer Zusammenhang zwischen der erzielten Punktzahl und der Mathematikangst zu beobachten. Steigt die Mathematikangst um einen Punkt, dann sinkt die Leistung um 20 Punkte. Es zeigt sich nur noch zwischen der Selbstwirksamkeit und der Testleistung ein positiver Zusammenhang; beim Mathematikinteresse, der instrumentellen Motivation und dem mathematischen Selbstkonzept sind jeweils negative Zusammenhänge mit der Mathematikleistung festzustellen. Mit einem  $r^2$  von 0.3256 erhöht sich die erklärte Varianz durch die Berücksichtigung der Mathematikangst um 2.5 Punkte auf 32.56 Prozent.

Bei der PISA-Studie 2003 schenkte die OECD Lernstrategien eine besondere Aufmerksamkeit, die in den Analysen als so genannte Kontrollvariablen berücksichtigt werden sollen. Mit jeweils vier beziehungsweise fünf Items wurden die Schüler nach Strategien befragt, die Repräsentation von Kenntnissen und Verfahren ermöglichen, die unverarbeitet im Gedächtnis gespeichert werden (Memorier- und Aufsaagestrategien), bei denen neue Informationen mit bereits Gelerntem verknüpft werden (Elaborationsstrategien) und mit denen sie prüfen, was sie gelernt haben, und herauszufinden versuchen, was sie noch lernen müssen (Kontrollstrategien). Diese drei Indizes werden dem Regressionsmodell als Kontrollvariablen (Kontrolle, Elaboration und Memorieren) hinzugefügt.

<b>Mathematik</b>	<b>Koeffizient</b>	<b>t-Wert</b>	<b>p-Wert</b>
Jungen	-8.14	-3.37	0.001
Interesse	-4.68	-2.99	0.003
wichtig	-4.58	-3.18	0.001
Selbstkonzept	-1.16	-0.59	0.553
Selbstwirksamkeit	50.44	36.58	0.000
Mathematikangst	-12.72	-8.22	0.000
Kontrolle	1.26	1.00	0.319
Elaboration	-6.94	-5.63	0.000
Memorieren	-14.63	-12.65	0.000
Konstante	429.86	52.01	0.000

Fallzahl: 4321

$r^2 = 0.3682$

Tabelle 6-7: Regressionsmodell 7 (Deutschland)

Mit einem  $r^2$  von 0.3682 sind 36.8 Prozent der Gesamtvarianz durch die im Regressionsmodell enthaltenen Variablen erklärt. Die Kontrollstrategien und das mathematische Selbstkonzept zeigen keinen signifikanten Zusammenhang mit der Mathematikleistung. Überraschend ist der negative Zusammenhang zwischen den Elaborationsstrategien und der erreichten Gesamtpunktzahl, während der negative Zusammenhang zwischen den Memorier- und Aufsaagestrategien und der Mathematikleistung zu erwarten war. Das Auswendiglernen von mathematischen Konzepten kann zwar kurzfristig Erfolg versprechend sein, da die PISA-Aufgaben aber darauf abzielten, die gelernte Mathematik in problemhaltigen Situationen flexibel an-

wenden zu können, ist das Auswendiglernen von einzelnen mathematischen Rechenfolgen für das Lösen von PISA-Aufgaben wohl nicht Ziel führend.

Die bisherigen Regressionsmodelle unterstellen, dass die einzelnen Variablen, beispielsweise das mathematische Interesse oder die instrumentelle Motivation, einen vom Geschlecht unabhängigen Einfluss auf die Mathematikleistung haben. Die aus der Stereotype-Threat-Theorie und dem MERF-Modell abgeleiteten Hypothesen postulieren jedoch je nach Geschlecht unterschiedliche Effekte. So wird bei den Mädchen ein negativer Zusammenhang zwischen dem Interesse und der erzielten Punktzahl erwartet, während ein solcher Zusammenhang bei den Jungen nicht zu beobachten sein sollte. Aus dem MERF-Modell wurde für die Mädchen ein negativer Zusammenhang zwischen der instrumentellen Motivation (Indikator für Prevention Focus) und der Mathematikleistung abgeleitet, während für die Jungen ein negativer Zusammenhang zwischen dem mathematischen Selbstkonzept (Indikator für Promotion Focus) und der erzielten Punktzahl zu erwarten ist. Ausgehend von dieser Argumentation werden im nächsten Regressionsmodell Interaktionsterme zwischen dem Geschlecht und den einzelnen Variablen aufgenommen.

<b>Mathematik</b>	<b>Koeffizient</b>	<b>t-Wert</b>	<b>p-Wert</b>
Jungen	54.97	3.29	0.001
Interesse	-8.85	-3.86	0.000
Interesse*Jungen	7.07	2.26	0.024
wichtig	-7.28	-3.59	0.000
wichtig*Jungen	5.77	2.01	0.044
Selbstkonzept	6.73	2.35	0.019
Selbstkonzept*Jungen	-14.79	-3.76	0.000
Selbstwirksamkeit	56.39	26.59	0.000
Selbstwirksamkeit*Jungen	-10.02	-3.60	0.000
Mathematikangst	-9.84	-4.18	0.000
Mathematikangst*Jungen	-4.89	-1.57	0.116
Kontrolle	-1.60	-0.89	0.373
Kontrolle*Jungen	5.46	2.16	0.030
Elaboration	-7.88	-4.45	0.000
Elaboration*Jungen	2.37	0.96	0.335
Memorieren	-9.86	-6.13	0.000
Memorieren*Jungen	-9.62	-4.16	0.000
Konstante	394.15	31.26	0.000

Fallzahl: 4321

$r^2 = 0.3765$

Tabelle 6-8: Regressionsmodell 8 (Deutschland)

Bei den Schülerinnen ist ein negativer Zusammenhang zwischen dem Mathematikinteresse und der Mathematikleistung zu beobachten. Steigt das Interesse um einen Punkt, dann sinkt das durchschnittliche Testergebnis um 8.85 Punkte. Bei den Jungen hat das Interesse – Verrechnung des Haupteffekts (-8.85) und des Interaktionsterms (7.07) – praktisch keinen Ein-

fluss auf die erreichte Punktzahl. Dieses Ergebnis ist mit der Stereotype-Threat-Theorie kompatibel. Stereotype Threat entsteht, wenn sich eine Person in einer Situation befindet, in der negative stereotype Erwartungen auf ihr Verhalten anwendbar sind, und ist mit der Befürchtung verbunden, negativ beurteilt zu werden. Diese wahrgenommene Befürchtung führt zu einer Verringerung der Leistungsfähigkeit, die insbesondere bei Personen zu beobachten ist, die sich für den Bereich interessieren. Das Regressionsmodell stützt den erwarteten Zusammenhang: Je stärker sich die Mädchen für Mathematik interessieren, desto geringer ist die Mathematikleistung. Steigt das Interesse um einen Punkt, dann sinkt die durchschnittliche Leistung um knapp neun Punkte. Da die Jungen beim PISA-Mathematiktest gerade einmal neun Punkte mehr erzielten als die Mädchen, unterstreicht dieses Ergebnis die Stärke des Stereotype-Threat-Effekts.

Interessant ist auch der geschlechtsspezifische Effekt der instrumentellen Motivation (wichtig); die Variable, die in den Analysen als Indikator für einen chronischen Prevention Focus dient. Der negative Zusammenhang zwischen Mathematikleistung und instrumenteller Motivation ist weitgehend auf die Mädchen beschränkt. Steigt der PISA-Index der instrumentellen Motivation um einen Punkt, dann sinkt die erreichte Punktzahl um 7.28 Punkte. Bei den Jungen zeigt sich nach Verrechnung des Haupteffekts (-7.28) und des Interaktionsterms (5.77) im Prinzip kein relevanter Zusammenhang zwischen der Einschätzung der Notwendigkeit der Mathematik im späteren Berufsleben und der erreichten Punktzahl im PISA-Mathematiktest. Dieser Befund steht im Einklang mit dem MERF-Modell: Personen mit einer erhöhten Zugänglichkeit zu einem Prevention Focus werten negative Erwartungen als Bedrohung. Der aus dieser Bedrohung resultierende Druck führt zu einer geringeren Testleistung, die bei Personen, die nicht mit einer negativen Erwartung konfrontiert sind, ausbleibt. Die Ergebnisse des Regressionsmodells entsprechen dieser Argumentation. So waren eher die Mädchen als die Jungen beim PISA-Mathematiktest mit einer negativen Erwartung konfrontiert, die sich regressionsanalytisch in einem negativen Zusammenhang zwischen instrumenteller Motivation und Testleistung niederschlägt.

Das mathematische Selbstkonzept dient in dieser Arbeit als Indikator für einen chronischen Promotion Focus. Personen in einem Promotion Focus wollen so genannte Idealziele erreichen. Übertragen auf die Mathematik stellt beispielsweise das Verstehen von schwierigen Aufgaben ein solches Idealziel dar. Nach dem MERF-Modell führen in einem Promotion Focus positive Erwartungen zu einem zusätzlichen Druck, der sich negativ auf die Leistung auswirkt, während eine negative Erwartung im Promotion Focus nicht als Bedrohung interpretiert wird, ein zusätzlicher Druck ausbleibt und ein besseres Ergebnis erzielt werden kann. Ausgehend von dieser Argumentation sollte das mathematische Selbstkonzept je nach Ge-

schlecht einen unterschiedlichen Einfluss auf die erreichte Punktzahl im PISA-Mathematiktest haben. Bei den Mädchen ist ein positiver Zusammenhang zwischen Selbstkonzept und Leistung zu beobachten, bei den Jungen zeigt sich ein negativer Zusammenhang zwischen mathematischem Selbstkonzept und Testleistung. Dieses Ergebnis entspricht den theoretischen Erwartungen.

Für beide Geschlechter zeigt sich ein positiver Zusammenhang zwischen der Selbstwirksamkeit und der Testleistung. Steigt bei den Mädchen die Selbstwirksamkeit um einen Punkt, dann steigt die durchschnittliche Leistung um 56 Punkte, bei den Jungen um 46 Punkte. Die Selbstwirksamkeit hat bei den Mädchen also einen stärkeren Einfluss auf die Testleistung als bei den Jungen. Da die Selbstwirksamkeit – wie die anderen Variablen auch – per Selbstausskunft erhoben wurde, spiegelt dieses Ergebnis vermutlich die Tendenz der Mädchen wieder, ihre Leistungsfähigkeit zu unterschätzen, während die Jungen dazu tendieren, ihre Leistungsstärke zu überschätzen.

Darüber hinaus zeigt sich ein negativer Zusammenhang zwischen der Mathematikangst und der Mathematikleistung, wobei der Interaktionsterm der Jungen statistisch nicht abgesichert werden kann. Bei den Mädchen zeigt sich kein signifikanter Zusammenhang zwischen den Kontrollstrategien und der Testleistung, während bei den Jungen ein positiver Zusammenhang zu beobachten ist. Der negative Zusammenhang zwischen dem Auswendiglernen und der Mathematikleistung überrascht in diesem Kontext nicht, wohl aber der negative Zusammenhang zwischen der Elaboration, also dem Verknüpfen von Lerninhalten mit anderen Informationen, und der erreichten Punktzahl im Mathematiktest. Möglicherweise verbinden die Schüler die neuen Konzepte mit dem, was sie bereits wissen, ohne aber weiter über diese nachzudenken. So stellt Elaboration zwar eine bessere Strategie als das reine mechanische Auswendiglernen dar, trägt aber nicht wirklich zum besseren Verständnis bei.

Künftiger Forschung bleibt es vorbehalten, den Zusammenhang zwischen Lernstrategien und der Leistung näher zu untersuchen. In dieser Analyse sind die Lernstrategien in erster Linie Kontrollvariablen, um zu prüfen, ob die Zusammenhänge zwischen Mathematikinteresse und Mathematikleistung oder instrumenteller Motivation und Mathematikleistung auch bei Kontrolle der Lernstrategien erhalten bleiben.

Im theoretischen Teil dieser Arbeit wurde argumentiert, dass Personen, denen Mathematik wichtig ist, beim PISA-Mathematiktest möglicherweise besonders angespannt waren. Durch die emotionale Anspannung waren die Schüler mit für die Aufgabenbewältigung irrelevanten Kognitionen beschäftigt, die die Fähigkeit, eine Aufgabe korrekt zu lösen, reduziert haben. Das nächste Regressionsmodell wird deshalb um weitere Interaktionsterme (Angst\*wichtig und Angst\*wichtig\*Jungen) ergänzt.

Mathematik	Koeffizient	t-Wert	p-Wert
Jungen	50.49	2.62	0.009
Interesse	-8.97	-3.92	0.000
Interesse*Jungen	7.19	2.31	0.021
wichtig	1.57	0.41	0.682
wichtig*Jungen	5.33	1.07	0.284
Selbstkonzept	7.08	2.48	0.013
Selbstkonzept*Jungen	-14.53	-3.70	0.000
Selbstwirksamkeit	56.14	26.50	0.000
Selbstwirksamkeit*Jungen	-10.26	-3.68	0.000
Mathematikangst	-2.11	-0.57	0.566
Mathematikangst*Jungen	-1.29	-0.24	0.810
Kontrolle	-1.80	-1.00	0.315
Kontrolle*Jungen	5.60	2.23	0.026
Elaboration	-7.92	-4.48	0.000
Elaboration*Jungen	2.44	0.99	0.321
Memorieren	-9.85	-6.13	0.000
Memorieren*Jungen	-9.36	-4.05	0.000
Angst*wichtig	-3.56	-2.72	0.006
Angst*wichtig*Jungen	-.75	-0.42	0.677
Konstante	375.32	26.14	0.000

Fallzahl: 4321

$r^2 = 0.3792$

Tabelle 6-9: Regressionsmodell 9 (Deutschland)

Der Regressionskoeffizient Angst\*wichtig lässt sich entsprechend den theoretischen Annahmen interpretieren. Es zeigt sich ein negativer Zusammenhang zwischen instrumenteller Motivation/Mathematikangst und Testleistung. Je höher die instrumentelle Motivation, desto stärker war die Person beim PISA-Mathematiktest emotional angespannt und mit aufgabenirrelevanten Kognitionen konfrontiert. Dies führte zu einer geringeren Punktzahl beim PISA-Mathematiktest.

Zusammenfassend lassen sich bei den deutschen Schülern deutliche Hinweise finden, die im Sinne der Stereotype-Threat-Theorie und des MERF-Modells interpretiert werden können. So ist der negative Zusammenhang zwischen dem Mathematikinteresse und der Mathematikleistung bei den Mädchen ebenso mit den theoretischen Annahmen kompatibel, wie der negative Zusammenhang zwischen der instrumentellen Motivation und der erzielten Punktzahl bei den Mädchen und der negative Zusammenhang zwischen dem mathematischen Selbstkonzept und der Testleistung bei den Jungen (vgl. Tabelle 6-8 auf Seite 78).

Bei den bisherigen Regressionsmodellen wurde nur die Gesamtskala Mathematik betrachtet; mit der folgenden Analyse wird die Betrachtung auf die einzelnen Bereiche – Geometrie, Stochastik, Algebra und Arithmetik – ausgeweitet.

Mathematik	Gesamtskala	Geometrie	Stochastik	Algebra	Arithmetik
Jungen	54.97	36.91	57.87	60.81	57.85
Interesse	-8.85	-6.96	-9.73	-9.27	-7.63
Interesse*Jungen	7.07	9.15	5.32 <sup>ns</sup>	6.62 <sup>ns</sup>	8.92
wichtig	-7.28	-5.67	-6.78	-6.77	-8.97
wichtig*Jungen	5.77	4.29 <sup>ns</sup>	5.41 <sup>ns</sup>	5.98 <sup>ns</sup>	4.49 <sup>ns</sup>
Selbstkonzept	6.73	7.01	3.95 <sup>ns</sup>	6.19 <sup>ns</sup>	7.77
Selbstkonzept*Jungen	-14.79	-13.86	-13.63	-16.62	-16.76
Selbstwirksamkeit	56.39	57.39	58.95	61.10	51.54
Selbstwirksamkeit*Jungen	-10.02	-8.55	-9.94	-10.08	-9.53
Mathematikangst	-9.84	-12.84	-8.04	-9.22	-10.15
Mathematikangst*Jungen	-4.89 <sup>ns</sup>	-3.70 <sup>ns</sup>	-5.86 <sup>ns</sup>	-6.33 <sup>ns</sup>	-6.46
Kontrolle	-1.60 <sup>ns</sup>	-4.09	-2.05 <sup>ns</sup>	-1.22 <sup>ns</sup>	-3.57 <sup>ns</sup>
Kontrolle*Jungen	5.46	5.84	6.94	7.20	6.02
Elaboration	-7.88	-4.51	-6.88	-10.79	-7.78
Elaboration*Jungen	2.37 <sup>ns</sup>	-1.77 <sup>ns</sup>	2.33 <sup>ns</sup>	1.66 <sup>ns</sup>	1.23 <sup>ns</sup>
Memorieren	-9.86	-13.41	-10.40	-11.10	-5.64
Memorieren*Jungen	-9.62	-8.57	-8.85	-10.40	-10.07
Konstante	394.15	398.31	369.54	387.33	421.25
Fallzahl	4321	4321	4321	4321	4321
r <sup>2</sup>	0.3765	0.3873	0.3906	0.3524	0.3223

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-10: Regressionsmodell 10 (Deutschland)

Zwar sind in einzelnen Regressionsmodellen einige Interaktionsterme auf dem Fünf-Prozent-Signifikanzniveau nicht signifikant, doch insgesamt können die Ergebnisse im Sinne der Stereotype-Threat-Theorie beziehungsweise des MERF-Modells interpretiert werden. Die „Richtung“ der einzelnen Regressionsparameter stimmt in allen Modellen mit den theoretisch erwarteten Zusammenhängen überein.

Der negative Zusammenhang zwischen dem mathematischen Interesse und der Testleistung ist weit gehend auf die Mädchen beschränkt. Bei den Jungen zeigt sich entweder kein oder ein geringer positiver Zusammenhang zwischen dem mathematischen Interesse und der Leistung. Ausnahmen bilden die Subskalen Stochastik und Algebra, bei denen der jeweilige Interaktionsterm nicht signifikant ist.

Auch finden sich Hinweise, die im Einklang mit dem MERF-Modell stehen. So zeigt sich bei den Mädchen durchweg ein negativer Zusammenhang der instrumentellen Motivation (Indikator für Prevention Focus) und der Testleistung sowie ein positiver Zusammenhang des mathematischen Selbstkonzepts (Indikator für Promotion Focus) und der erzielten Punktzahl. Bei den Jungen ist das gegenläufige Muster zu beobachten.

In der Debatte um geschlechtsspezifische Unterschiede in der Mathematik wird häufig angemerkt, dass Mädchen in der Schule meist bessere Noten erhalten, bei standardisierten Leistungstests in Mathematik jedoch schlechter abschneiden als Jungen (Beerman et al., 1992).

Deshalb soll an dieser Stelle die Punktzahl beim PISA-Mathematiktest mit der Mathematiknote verglichen werden, ehe die Analysen auf andere Länder ausgeweitet werden.

Da Schulnoten durch vielfältige Verhaltensweisen beeinflusst werden, beispielsweise Fleiß, Aufmerksamkeit oder auch Betragen, sind sie kein geeigneter Indikator für die mathematische Fähigkeit. Diese Argumentation wird durch die Korrelation von 0.32 zwischen der erreichten Punktzahl beim PISA-Test und der Mathematiknote gestützt. Entgegen amerikanischen Befunden ist zumindest bei der deutschen PISA-Studie auch kein geschlechtsspezifischer Vorteil bei den Mathematiknoten für die Mädchen auszumachen. Im Gegenteil: Die Jungen erhalten auf der Basis der deutschen PISA-Stichprobe eine durchschnittliche Mathematiknote von 2.97, die Mädchen eine Mathematiknote von 3.11. Die Differenz von 0.14 ist statistisch abgesichert, der geschlechtsspezifische Unterschied ist mit einem d-Wert von 0.14 bei den Schulnoten größer als beim PISA-Test (d-Wert: 0.09). Die Tabelle zeigt die Mathematiknoten der Schüler, die bei der deutschen PISA-Studie teilgenommen haben.

Note	Mädchen	Jungen
1	4.38 Prozent	5.06 Prozent
2	25.06 Prozent	30.10 Prozent
3	34.69 Prozent	34.88 Prozent
4	27.07 Prozent	23.52 Prozent
5	8.58 Prozent	5.77 Prozent
6	0.23 Prozent	0.66 Prozent
Fallzahl	2191	2113

Tabelle 6-11: Mathematiknote nach Geschlecht

Haben die motivationalen Merkmale der Schüler jeweils den gleichen Einfluss auf die Mathematiknote und die erreichte Punktzahl beim PISA-Mathematiktest? Um diese Frage zu beantworten, wurden zwei getrennte Regressionen berechnet – einmal mit der Mathematiknote als abhängiger Variable, einmal mit der erreichten Punktzahl beim PISA-Mathematiktest als abhängiger Variable. Um die Ergebnisse besser vergleichen zu können, wurde die Mathematiknote so kodiert, dass eine höhere Mathematiknote einer besseren Mathematiknote entspricht (sehr gut = 6). Um die Einflussstärken der verschiedenen Variablen besser vergleichen zu können, enthält die folgende Tabelle standardisierte Regressionskoeffizienten, so genannte Beta-Koeffizienten. Da in der Gegenwart von Interaktionstermen die Berechnung von standardisierten Regressionskoeffizienten ungültig ist, wurden – um dieses Problem zu lösen – die an der Bildung der Interaktionsterme beteiligten Variablen (mit Ausnahme der dichotomen Geschlechtsvariable) im Vorfeld der Regression z-standardisiert (Kohler & Kreuter, 2001).

Die Tabelle enthält die standardisierten Regressionskoeffizienten, bei der die einzelnen Variablen vor Bildung der Interaktionsterme z-standardisiert wurden.<sup>32</sup>

<b>Mathematik</b>	<b>Punktzahl</b>	<b>Mathenote</b>
Jungen	-0.05	-0.11
Interesse	-0.11	-0.10
Interesse*Jungen	0.06	0.10
wichtig	-0.08	0.00 <sup>ns</sup>
wichtig*Jungen	0.04	0.02 <sup>ns</sup>
Selbstkonzept	0.08	0.66
Selbstkonzept*Jungen	-0.12	-0.11
Selbstwirksamkeit	0.59	0.06
Selbstwirksamkeit*Jungen	-0.08	0.02 <sup>ns</sup>
Mathematikangst	-0.12	-0.10
Mathematikangst*Jungen	-0.04 <sup>ns</sup>	-0.02 <sup>ns</sup>
Kontrolle	-0.02 <sup>ns</sup>	0.03 <sup>ns</sup>
Kontrolle*Jungen	0.05	-0.03 <sup>ns</sup>
Elaboration	-0.09	-0.01 <sup>ns</sup>
Elaboration*Jungen	0.02 <sup>ns</sup>	0.00 <sup>ns</sup>
Memorieren	-0.13	-0.01 <sup>ns</sup>
Memorieren*Jungen	-0.09	-0.01 <sup>ns</sup>
Konstante	520.06	4.06
Fallzahl	4321	4236
r <sup>2</sup>	0.3765	0.4741

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-12: Mathematiknote und PISA-Punktzahl im Vergleich (standardisierte Regressionskoeffizienten)

Die standardisierten Regressionskoeffizienten können wie folgt interpretiert werden: Eine Erhöhung der Selbstwirksamkeit um eine Standardabweichung erhöht beim PISA-Test die Punktzahl um 0.59 Standardabweichungen, die Mathematiknote um 0.06 Standardabweichungen. Während beim PISA-Test die Selbstwirksamkeit den stärksten Einfluss auf die Punktzahl hat, hat das Selbstkonzept bei der Mathematiknote den größten Einfluss. Die Lernstrategien haben – im Gegensatz zum PISA-Test – auf die Mathematiknote überhaupt keinen Einfluss. Ebenso hat die instrumentelle Motivation (wichtig), der Indikator für die Zugänglichkeit zu einem chronischen Prevention Focus, keinen Einfluss auf die Mathematiknote. Der negative Zusammenhang zwischen dem Mathematikinteresse und der Leistung zeigt sich auch bei der Mathematiknote, wobei dieser Effekt ausschließlich bei den Mädchen zu beobachten ist. Bei den Jungen hat das Mathematikinteresse keinen Einfluss auf die Mathematiknote. Die Befunde sind zumindest als Hinweis zu werten, dass die Mädchen auch im Unterrichtsalltag mit negativen stereotypen Erwartungen konfrontiert sind.

<sup>32</sup> Die Verwendung standardisierter Regressionskoeffizienten bei dichotomen Variablen ist unzulässig, da die Standardabweichung einer dichotomen Variablen eine Funktion ihrer Schiefe ist, die standardisierten Regressionskoeffizienten deshalb um so niedriger werden, je schief die Variable ist (Kohler & Kreuter, 2001). Deshalb sollte die Geschlechtsvariable (Jungen) nicht interpretiert werden.

Nach dem Vergleich zwischen Mathematiknote und erreichter Punktzahl auf der Basis der deutschen PISA-Stichprobe sollen die empirischen Analysen auf den internationalen Vergleich ausgeweitet werden.

## 6.2 Internationaler Vergleich

An der PISA-Studie 2003 haben sich insgesamt 41 Staaten<sup>33</sup> beteiligt. Damit ist es möglich, Vergleiche zwischen Staaten vorzunehmen. Ein Ländervergleich mit 40 Staaten erscheint allerdings nicht sinnvoll, da internationale Vergleiche von Selbstaussagen nicht unproblematisch sind. Selbstaussagen zu Einstellungen lassen sich interkulturell nur begrenzt zueinander in Beziehung setzen. Durch die Übersetzung der Fragebögen in verschiedene Sprachen können Aussagen unterschiedliche Akzentuierungen erhalten, die es unterschiedlich leicht fallen lassen, sie zu bejahen oder zu verneinen. Hinzu kommen kulturspezifische Regeln des Umgangs mit selbstbezogenen Aussagen, die eine eher bescheidene oder eher offensive Selbstdarstellung als sozial erwünscht scheinen lassen (Pekrun & Zirngibl, 2004). Für den internationalen Vergleich wurden deshalb westeuropäische Länder ausgewählt, bei denen eine gemeinsame Kultur zumindest ansatzweise unterstellt werden kann. Die ausgewählten Länder wurden in geografische Ländergruppen eingeteilt, die der Tabelle zu entnehmen sind.

Zentraleuropa	Skandinavien	Südeuropa	Beneluxstaaten
Deutschland	Dänemark	Spanien	Niederlande
Schweiz	Finnland	Portugal	Belgien
Österreich	Norwegen	Italien	Luxemburg
	Schweden		

Tabelle 6-13: Untersuchungsländer

In einem ersten Schritt soll geprüft werden, ob sich auch in anderen Ländern Indizien finden lassen, die mit Stereotype Threat und dem MERF-Modell kompatibel sind. Die geschlechtsspezifischen Leistungsunterschiede zwischen den Ländern variieren zum Teil recht deutlich. So zeigt sich beispielsweise in den Niederlanden ein Leistungsvorsprung der Jungen von fünf Punkten, während in Luxemburg die Jungen beim PISA-Mathematiktest durchschnittlich 17 Punkte besser abschneiden als die Mädchen. Der zweite Schritt zielt deshalb darauf ab, die geschlechtsspezifischen Unterschiede in der Mathematikleistung zwischen den Staaten zu erklären. Da die Aufgliederung der Mathematikleistung nach einzelnen Inhaltsbereichen – Geometrie, Stochastik, Algebra, Arithmetik – keine wesentlich neuen Erkenntnisse erbrachte, bleibt der Ländervergleich auf die Gesamtskala Mathematik begrenzt.

<sup>33</sup> Wie bereits in Kapitel 3 berichtet, hat das Vereinigte Königreich die vom PISA-Konsortium festgelegten Beteiligungsquoten nicht erfüllt, weshalb ein Vergleich mit anderen Staaten ausscheidet (OECD, 2004a).

### 6.2.1 Ländervergleich mit Deutschland, Österreich und Schweiz

Bei der PISA-Studie 2003 erzielten die deutschen Schüler durchschnittlich 503 Punkte, die österreichischen Teilnehmer 506 Punkte und die Schüler aus der Schweiz 527 Punkte. Während sich Deutschland und Österreich nicht signifikant vom OECD-Durchschnitt (500 Punkte) unterscheiden, liegt die Schweiz oberhalb des OECD-Durchschnitts. Die geschlechtsspezifischen Leistungsdifferenzen sind in Deutschland mit neun Punkten im internationalen Vergleich eher unauffällig, in Österreich beträgt der Abstand 7.5 Punkte und ist statistisch nicht abgesichert. In Deutschland erzielten die Jungen 508 Punkte, die Mädchen 499 Punkte. In Österreich kamen die Mädchen durchschnittlich auf 502, die Jungen auf 509 Punkte. Mit einer Differenz von 17 Punkten können in der Schweiz größere geschlechtsspezifische Unterschiede nachgewiesen werden. Die Jungen erzielten in der Schweiz 535 Punkte, die Mädchen dagegen 518 Punkte. Finden sich – analog zu Deutschland – auch in der Schweiz und in Österreich Hinweise, die mit Stereotype Threat und dem MERF-Modell kompatibel sind? Um die Frage zu beantworten, wurden nach Ländern getrennte Regressionen berechnet.

<b>Mathematik</b>	<b>Schweiz</b>	<b>Österreich</b>	<b>Deutschland</b>
Jungen	27.21	28.65 <sup>ns</sup>	54.97
Interesse	-6.22	-7.58	-8.85
Interesse*Jungen	8.12	6.61	7.07
wichtig	-11.01	-10.77	-7.28
wichtig*Jungen	2.19 <sup>ns</sup>	3.07 <sup>ns</sup>	5.77
Selbstkonzept	8.99	21.87	6.73
Selbstkonzept*Jungen	-9.56	-13.53	-14.79
Selbstwirksamkeit	57.78	46.81	56.39
Selbstwirksamkeit*Jungen	-7.74	-4.71 <sup>ns</sup>	-10.02
Mathematikangst	-6.29	-.84 <sup>ns</sup>	-9.84
Mathematikangst*Jungen	-5.11	-4.92 <sup>ns</sup>	-4.89 <sup>ns</sup>
Kontrolle	-3.02	-1.62 <sup>ns</sup>	-1.60 <sup>ns</sup>
Kontrolle*Jungen	9.78	8.95	5.46
Elaboration	-10.99	-10.68	-7.88
Elaboration*Jungen	.88 <sup>ns</sup>	5.43	2.37 <sup>ns</sup>
Memorieren	-7.14	-8.03	-9.86
Memorieren*Jungen	-8.05	-13.26	-9.62
Konstante	389.27	376.28	394.15
Fallzahl	8253	4485	4321
r <sup>2</sup>	0.4078	0.3820	0.3765

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-14: Regressionsmodell Deutschland, Österreich und Schweiz 1

In allen drei Ländern zeigt sich für die Mädchen ein negativer Zusammenhang zwischen dem Mathematikinteresse und der Mathematikleistung. Für die Jungen sind bei Verrechnung des Haupteffekts und des Interaktionsterms in Deutschland (-1.78) und in Österreich (-0.97) marginale negative Zusammenhänge, in der Schweiz (1.9) ein positiver Zusammenhang zwischen Interesse und Punktzahl festzustellen.

Der aus dem MERF-Modell abgeleitete Zusammenhang zwischen instrumenteller Motivation (wichtig) und Mathematikleistung sowie Selbstkonzept und Mathematikleistung kann weitgehend regressionsanalytisch nachgewiesen werden. Allerdings ist in der Schweiz und in Österreich der erwartete geschlechtsspezifische Effekt der instrumentellen Motivation bei einem Signifikanzniveau von fünf Prozent statistisch nicht abgesichert. Die Zugänglichkeit zu einem Prevention Focus hat demnach sowohl für Jungen als auch für Mädchen einen negativen Effekt auf die Mathematikleistung, theoretisch sollte der leistungsmindernde Effekt jedoch auf die weiblichen Personen begrenzt sein. Dies ist möglicherweise auf die konzeptionell problematische Erfassung der Zugänglichkeit zu einem Prevention Focus zurückzuführen. Allerdings muss bei der Bewertung der Ergebnisse berücksichtigt werden, dass zumindest die „Richtung“ der Regressionskoeffizienten mit den theoretischen Erwartungen übereinstimmt. Beim mathematischen Selbstkonzept – Indikator für einen chronischen Promotion Focus – zeigt sich das theoretisch erwartete Bild: ein positiver Zusammenhang zwischen Selbstkonzept und Mathematikleistung für die Mädchen, ein negativer beziehungsweise für Österreich ein deutlich schwächerer positiver Zusammenhang zwischen Selbstkonzept und erreichter Punktzahl im PISA-Mathematiktest bei den Jungen. Das nächste Regressionsmodell wird um die Interaktionsterme Angst\*wichtig sowie Angst\*wichtig\*Jungen ergänzt.

<b>Mathematik</b>	<b>Schweiz</b>	<b>Österreich</b>	<b>Deutschland</b>
Jungen	52.93	20.14 <sup>ns</sup>	50.49
Interesse	-6.42	-7.91	-8.97
Interesse*Jungen	8.32	6.69	7.19
wichtig	.10 <sup>ns</sup>	-4.80 <sup>ns</sup>	1.57 <sup>ns</sup>
wichtig*Jungen	-11.43	5.78 <sup>ns</sup>	5.33 <sup>ns</sup>
Selbstkonzept	9.06	21.90	7.08
Selbstkonzept*Jungen	-9.72	-13.32	-14.53
Selbstwirksamkeit	57.34	46.67	56.14
Selbstwirksamkeit*Jungen	-7.25	-4.70 <sup>ns</sup>	-10.26
Mathematikangst	2.29 <sup>ns</sup>	2.92 <sup>ns</sup>	-2.11 <sup>ns</sup>
Mathematikangst*Jungen	-17.10	.45 <sup>ns</sup>	-1.29 <sup>ns</sup>
Kontrolle	-3.20	-1.66 <sup>ns</sup>	-1.80 <sup>ns</sup>
Kontrolle*Jungen	10.03	8.73	5.60
Elaboration	-11.04	-10.61	-7.92
Elaboration*Jungen	.95 <sup>ns</sup>	5.11	2.44 <sup>ns</sup>
Memorieren	-7.29	-7.99	-9.85
Memorieren*Jungen	-7.88	-13.29	-9.36
Angst*wichtig	-4.44	-2.52	-3.56
Angst*wichtig*Jungen	5.70	-1.73 <sup>ns</sup>	-.75 <sup>ns</sup>
Konstante	370.30	367.64	375.32
Fallzahl	8253	4485	4321
r <sup>2</sup>	0.4092	0.3842	0.3792

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-15: Regressionsmodell Deutschland, Österreich und Schweiz 2

In Österreich und in Deutschland ist nur der Interaktionsterm Angst\*<sup>2</sup>wichtig signifikant. Aufschlussreich ist der Befund für die Schweizer PISA-Daten: Es zeigt sich ein geschlechtsspezifischer Zusammenhang zwischen der Einschätzung der Bedeutsamkeit von Mathematik (wichtig) und der Mathematikangst. Bei Frauen, die eine erhöhte Zugänglichkeit zu einem Prevention Focus haben, ist der Leistungsabfall über die Mathematikangst vermittelt. Sie waren beim PISA-Mathematiktest emotional angespannt und mit aufgabenirrelevanten Kognitionen konfrontiert, was zu einer geringeren Punktzahl beim PISA-Mathematiktest führte. Insgesamt sind die Befunde mit den Annahmen der Stereotype-Threat-Theorie und dem MERF-Modell durchaus kompatibel. Besonders beeindruckend bleibt der – weitgehend auf die Mädchen begrenzte – negative Zusammenhang zwischen dem Mathematikinteresse und der erreichten Punktzahl beim PISA-Test. Schon dieser Zusammenhang unterstreicht die Relevanz des Einflusses von Stereotypen bei einer normalen Testsituation.

### **6.2.2 Ländervergleich mit Dänemark, Finnland, Norwegen und Schweden**

Die skandinavischen Länder liegen mit Ausnahme Norwegens oberhalb des OECD-Durchschnitts. Die finnischen Schüler erzielten bei der PISA-Studie durchschnittlich 544 Punkte, das beste Ergebnis bei der PISA-Studie 2003. In Dänemark kamen die Schüler auf 514 Punkte und in Schweden auf 509 Punkte. Norwegen gehört mit einem durchschnittlichen Testergebnis von 495 Punkten zum OECD-Durchschnitt.

Finnland gelingt es recht gut, hohe Gesamtleistungen zu erreichen und gleichzeitig nur relativ geringe Leistungsunterschiede zwischen Jungen und Mädchen entstehen zu lassen. Die Mädchen erreichten in Finnland 541 Punkte, die Jungen erzielten durchschnittlich sieben Punkte mehr und kamen so auf 548 Punkte. Die Ergebnisse in Finnland sind zumindest als Hinweis zu werten, dass es mit steigender mathematischer Kompetenz nicht unbedingt zu einer Zunahme geschlechtsspezifischer Leistungsdifferenzen kommen muss und diese als unvermeidlich angesehen werden müssen.

Die Jungen in Dänemark erzielten 523 Punkte, die Mädchen 506 Punkte. Durchschnittlich konnten die Jungen in Dänemark 17 Punkte mehr erreichen als die Mädchen, der stärkste geschlechtsspezifische Unterschied in den skandinavischen Ländern. In Schweden beträgt die Differenz zwischen Mädchen und Jungen sieben Punkte, in Norwegen ist der Abstand mit sechs Punkten statistisch nicht signifikant.

Auch für die skandinavischen Länder wurden getrennte Regressionen gerechnet, um einen möglichen Einfluss der Bedrohung durch Stereotype auf die gezeigte Mathematikleistung zu entdecken.

<b>Mathematik</b>	<b>Dänemark</b>	<b>Finnland</b>	<b>Norwegen</b>	<b>Schweden</b>
Jungen	3.90 <sup>ns</sup>	-2.32 <sup>ns</sup>	-1.08 <sup>ns</sup>	-10.96 <sup>ns</sup>
Interesse	-15.51	-11.79	-5.19	-11.19
Interesse*Jungen	9.53	7.34	7.22	7.45
wichtig	2.57 <sup>ns</sup>	2.12 <sup>ns</sup>	9.42	3.11 <sup>ns</sup>
wichtig*Jungen	-5.41 <sup>ns</sup>	3.10 <sup>ns</sup>	-2.05 <sup>ns</sup>	3.91 <sup>ns</sup>
Selbstkonzept	31.92	39.08	28.27	23.39
Selbstkonzept*Jungen	-5.41 <sup>ns</sup>	-3.68 <sup>ns</sup>	-3.94 <sup>ns</sup>	-5.23 <sup>ns</sup>
Selbstwirksamkeit	34.20	29.03	26.21	42.28
Selbstwirksamkeit*Jungen	-3.35 <sup>ns</sup>	-4.36 <sup>ns</sup>	2.27 <sup>ns</sup>	-2.53 <sup>ns</sup>
Mathematikangst	-15.25	-1.19 <sup>ns</sup>	-6.63	-8.25
Mathematikangst*Jungen	-0.59 <sup>ns</sup>	-7.12	-5.23 <sup>ns</sup>	-4.95 <sup>ns</sup>
Kontrolle	-9.57	-6.08	-10.32	-16.53
Kontrolle*Jungen	1.99 <sup>ns</sup>	6.43	11.32	5.98 <sup>ns</sup>
Elaboration	-10.23	-4.73	-8.88	-6.10
Elaboration*Jungen	5.69 <sup>ns</sup>	-3.14 <sup>ns</sup>	-7.19	-0.41 <sup>ns</sup>
Memorieren	-3.67 <sup>ns</sup>	-5.58	4.32	1.02 <sup>ns</sup>
Memorieren*Jungen	-5.74 <sup>ns</sup>	-2.39 <sup>ns</sup>	-7.41	-3.30 <sup>ns</sup>
Konstante	451.31	428.32	397.45	395.46
Fallzahl	4081	5657	3932	4552
r <sup>2</sup>	0.4264	0.4480	0.4583	0.4199

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-16: Regressionsmodell Dänemark, Finnland, Norwegen und Schweden 1

Beachtlich ist die durch die einzelnen Regressionsmodelle erklärte Varianz der Mathematikleistung in den einzelnen Ländern. In Norwegen sind mit einem  $r^2$  von 0.4583 über 45.8 Prozent der Gesamtvarianz der Mathematikleistung durch die im Modell enthaltenen Variablen erklärt. In Schweden liegt die erklärte Varianz bei 41.99 Prozent, in Dänemark bei 42.64 Prozent und in Finnland bei 44.88 Prozent.

Von besonderem Interesse ist an dieser Stelle wieder die geschlechtsspezifische Wirkung des Mathematikinteresses: In allen Ländern ist für die Mädchen ein negativer Zusammenhang zwischen Interesse und Punktzahl zu beobachten. Zwar zeigt sich – mit Ausnahme von Norwegen – auch bei den Jungen ein negativer Zusammenhang zwischen Interesse und erzielter Punktzahl beim PISA-Mathematiktest, der jedoch betragsmäßig deutlich niedriger ausfällt als bei den Mädchen.

Ein geschlechtsspezifischer Effekt der instrumentellen Motivation und des Selbstkonzepts kann in keinem Land nachgewiesen werden, die entsprechenden Regressionskoeffizienten sind jeweils nicht signifikant. Lediglich in Norwegen zeigt sich ein positiver Zusammenhang zwischen der instrumentellen Motivation und der Mathematikleistung.

Die Annahmen des MERF-Modells finden in den skandinavischen Ländern keine empirische Bestätigung. Dies ist möglicherweise auf die konzeptionell nicht unproblematische Erfassung des Prevention Focus beziehungsweise des Promotion Focus zurückzuführen. Allerdings bietet die PISA-Studie keine besseren Indikatoren, weshalb in dieser Arbeit auf die PISA-Indizes

zur instrumentellen Motivation und zum mathematischen Selbstkonzept zurückgegriffen wird. Der negative Zusammenhang zwischen Mathematikinteresse und Mathematikleistung kann jedoch dahin gehend interpretiert werden, dass die skandinavischen Schülerinnen bei der PISA-Testsituation mit dem Stereotyp „Frauen können keine Mathematik“ konfrontiert waren. Je stärker das Mathematikinteresse, desto größer die Gefahr, das kulturelle Stereotyp mit dem eigenen Verhalten zu bestätigen. Die Befürchtung kommt der Fähigkeit, in der Situation eine gute Leistung erbringen zu können, in die Quere. Die Stereotype-Threat-Theorie bietet die Möglichkeit, den negativen Zusammenhang zwischen Interesse und Leistung zu erklären, sozialisationstheoretische oder biologische Erklärungsansätze können diesen – alltagspsychologisch unerwarteten Befund – nicht begreiflich machen.

Das nächste Regressionsmodell wird erneut um die Interaktionsterme Angst\*wichtig sowie Angst\*wichtig\*Jungen ergänzt.

<b>Mathematik</b>	<b>Dänemark</b>	<b>Finnland</b>	<b>Norwegen</b>	<b>Schweden</b>
Jungen	14.03 <sup>ns</sup>	2.26 <sup>ns</sup>	7.33 <sup>ns</sup>	-19.56 <sup>ns</sup>
Interesse	-15.45	-11.80	-5.39	-11.12
Interesse*Jungen	9.38	7.37	7.23	7.77
wichtig	9.31	2.81 <sup>ns</sup>	18.33	8.33
wichtig*Jungen	-9.34 <sup>ns</sup>	1.16 <sup>ns</sup>	-3.94 <sup>ns</sup>	6.71 <sup>ns</sup>
Selbstkonzept	31.98	39.06	27.76	23.21
Selbstkonzept*Jungen	-5.12 <sup>ns</sup>	-3.61 <sup>ns</sup>	-4.01 <sup>ns</sup>	-5.35 <sup>ns</sup>
Selbstwirksamkeit	34.25	29.04	26.38	42.04
Selbstwirksamkeit*Jungen	-3.56 <sup>ns</sup>	-4.34 <sup>ns</sup>	1.92 <sup>ns</sup>	-2.33 <sup>ns</sup>
Mathematikangst	-7.15 <sup>ns</sup>	.48 <sup>ns</sup>	1.39 <sup>ns</sup>	-2.58 <sup>ns</sup>
Mathematikangst*Jungen	-3.90 <sup>ns</sup>	-9.44 <sup>ns</sup>	-5.92 <sup>ns</sup>	1.31 <sup>ns</sup>
Kontrolle	-9.96	-6.10	-10.58	-16.60
Kontrolle*Jungen	2.28 <sup>ns</sup>	6.47	11.19	6.13
Elaboration	-10.37	-4.75	-8.61	-6.18
Elaboration*Jungen	5.80 <sup>ns</sup>	-3.09 <sup>ns</sup>	-7.51	-.68 <sup>ns</sup>
Memorieren	-3.40 <sup>ns</sup>	-5.56	4.79	1.22 <sup>ns</sup>
Memorieren*Jungen	-6.06	-2.43 <sup>ns</sup>	-7.71	-3.74 <sup>ns</sup>
Angst*wichtig	-2.96	-.30 <sup>ns</sup>	-3.37	-2.48 <sup>ns</sup>
Angst*wichtig*Jungen	1.45 <sup>ns</sup>	.93 <sup>ns</sup>	.38 <sup>ns</sup>	-2.01 <sup>ns</sup>
Konstante	432.37	426.79	374.03	383.79
Fallzahl	4081	5657	3932	4552
r <sup>2</sup>	0.4271	0.4481	0.4602	0.4215

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-17: Regressionsmodell Dänemark, Finnland, Norwegen und Schweden 2

Je wichtiger die Mathematik, desto größer die Befürchtung, das kulturelle Stereotyp zu bestätigen. Die Personen sind dann eher mit aufgabenirrelevanten Kognitionen konfrontiert, was die Fähigkeit reduziert, die Aufgaben korrekt zu lösen. Für Dänemark und Norwegen zeigt sich ein negativer Zusammenhang zwischen instrumenteller Motivation/Angst und Mathematikleistung, der statistisch abgesichert ist.

Mit Ausnahme von Dänemark sind die geschlechtsspezifischen Leistungsunterschiede in den skandinavischen Ländern eher unauffällig. Dennoch lassen sich nicht nur in Dänemark, sondern auch in Finnland, Schweden und Norwegen Indizien finden, die darauf hinweisen, dass Stereotype Threat die Leistungsfähigkeit der skandinavischen Mädchen negativ beeinflusst hat. Dafür spricht der negative Zusammenhang des Mathematikinteresses und der Mathematikleistung. Wie sehen die Befunde für südeuropäische Länder aus? Diese Frage wird im folgenden Abschnitt beantwortet.

### 6.2.3 Ländervergleich mit Spanien, Italien und Portugal

Im Gegensatz zu den skandinavischen Ländern befinden sich die südeuropäischen Staaten unterhalb des OECD-Durchschnitts. Die spanischen Schüler erzielten durchschnittlich 485 Punkte, die italienischen und portugiesischen Schüler jeweils 466 Punkte. Der Leistungsvorsprung der Jungen liegt in Italien bei 18 Punkten (475 zu 457), in Portugal bei zwölf Punkten (472 zu 460) und in Spanien bei neun Punkten (490 zu 481). Multiple Regressionen sollen eine Beurteilung ermöglichen, welchen Einfluss die motivationalen Merkmale der Schüler auf die erzielte Punktzahl im PISA-Mathematiktest haben.

Mathematik	Spanien	Italien	Portugal
Jungen	-2.54 <sup>ns</sup>	3.53 <sup>ns</sup>	22.46 <sup>ns</sup>
Interesse	-7.36	-4.48	-17.99
Interesse*Jungen	1.37 <sup>ns</sup>	-7.69	-.80 <sup>ns</sup>
wichtig	1.83 <sup>ns</sup>	-4.82	4.25
wichtig*Jungen	6.76	13.87	2.23 <sup>ns</sup>
Selbstkonzept	21.83	10.38	25.31
Selbstkonzept*Jungen	-5.79	.94 <sup>ns</sup>	-3.23 <sup>ns</sup>
Selbstwirksamkeit	39.02	51.55	48.12
Selbstwirksamkeit*Jungen	-7.08	-1.91 <sup>ns</sup>	-4.44 <sup>ns</sup>
Mathematikangst	-5.86	-14.40	-6.09
Mathematikangst*Jungen	-3.49 <sup>ns</sup>	-1.36 <sup>ns</sup>	-1.10 <sup>ns</sup>
Kontrolle	1.72 <sup>ns</sup>	4.31	7.99
Kontrolle*Jungen	6.99	3.24 <sup>ns</sup>	7.18
Elaboration	-12.41	-20.24	-10.04
Elaboration*Jungen	6.29	-.42 <sup>ns</sup>	-5.01 <sup>ns</sup>
Memorieren	-1.91 <sup>ns</sup>	-7.99	-11.42
Memorieren*Jungen	-4.70	-4.73	-2.33 <sup>ns</sup>
Konstante	364.34	384.46	317.14
Fallzahl	10.596	11.432	4492
r <sup>2</sup>	0.2787	0.3149	0.3932

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-18: Regressionsmodell Spanien, Italien und Portugal 1

Im Vergleich zu den skandinavischen Ländern oder auch Deutschland, Österreich und der Schweiz fällt die deutlich geringere Varianzerklärung in Spanien und Italien auf. In Spanien sind nicht einmal 28 Prozent der Gesamtvarianz erklärt, in Italien gerade einmal 31.5 Prozent.

Wenn man berücksichtigt, dass in Norwegen mit den gleichen Variablen rund 15 Prozent mehr an Varianz erklärt ist, lässt dies den Schluss zu, dass die Leistungen im PISA-Mathematiktest in Spanien und Italien im weitaus stärkeren Maße von sozialstrukturellen Variablen abhängen als beispielsweise in den skandinavischen Ländern.

Insgesamt sind die empirischen Ergebnisse in den südeuropäischen Ländern schwierig einzuordnen. Zwar zeigt sich in allen Ländern der theoretisch erwartete negative Zusammenhang zwischen dem Mathematikinteresse und der Mathematikleistung, in Spanien und Portugal ist jedoch ein vom Geschlecht unabhängiger Einfluss auf die Mathematikleistung festzustellen. Konkret: Sowohl für die Mädchen als auch für die Jungen ist ein negativer Zusammenhang zwischen dem Mathematikinteresse und der Mathematikleistung zu beobachten. In Italien sind die Jungen sogar stärker betroffen als die Mädchen (signifikanter negativer Interaktionsterm). Gerade in Italien entspricht jedoch der Zusammenhang zwischen der Einschätzung der Bedeutung des Fachs (wichtig) und der Punktzahl im Mathematiktest den aus dem MERF-Modell abgeleiteten theoretischen Erwartungen. Für die Mädchen zeigt sich ein negativer, für die Jungen ein positiver Zusammenhang. In dieser Arbeit werden hohe Werte auf dem Index zur instrumentellen Motivation als Hinweis für eine Zugänglichkeit zu einem Prevention Focus angesehen. Der PISA-Mathematiktest stellte darüber hinaus eine Situation dar, bei der die weiblichen Testteilnehmer mit einer negativen stereotypen Erwartung konfrontiert waren. Nach dem MERF-Modell werden in einem Prevention Focus negative Erwartungen als Bedrohung interpretiert. Der daraus resultierende Druck führt schließlich bei den weiblichen Testpersonen zu einer geringeren Testleistung, die bei den Männern – die beim PISA-Mathematiktest nicht mit einer negativen stereotypen Erwartung konfrontiert waren – nicht zu beobachten ist. Die Ergebnisse in Spanien lassen sich tendenziell analog zu Italien interpretieren, wenn auch der Zusammenhang zwischen der Bedeutung der Mathematik (wichtig) und der Mathematikleistung bei den Mädchen nicht signifikant ist. In Portugal hat das instrumentelle Interesse einen vom Geschlecht unabhängigen Einfluss auf die Mathematikleistung.

Für die männlichen Testteilnehmer wurde vor dem Hintergrund des MERF-Modells ein negativer Zusammenhang zwischen dem mathematischen Selbstkonzept (Indikator für chronische Zugänglichkeit zu einem Promotion Focus) und der Mathematikleistung erwartet. Für die Schülerinnen wurde dagegen kein negativer Zusammenhang zwischen dem mathematischen Selbstkonzept und der Mathematikleistung erwartet, da eine negative Erwartung im Promotion Focus nicht als Bedrohung interpretiert wird, ein zusätzlicher Druck ausbleibt und ein besseres Testergebnis erzielt werden kann. Für die Mädchen in den südeuropäischen Ländern zeigt sich in der Tat ein positiver Zusammenhang zwischen Selbstkonzept und Mathematikleistung, jedoch lässt sich die aus dem MERF-Modell abgeleitete Hypothese für die Jungen

nicht bestätigen. In Italien und Portugal zeigt sich ein vom Geschlecht unabhängiger Einfluss des mathematischen Selbstkonzepts auf die Mathematikleistung. In Spanien ist dieser Zusammenhang für die Jungen zwar schwächer, aber immer noch deutlich positiv. Steigt der PISA-Index zum mathematischen Selbstkonzept um einen Punkt, dann erhöht sich die durchschnittliche Leistung unter Kontrolle der im Regressionsmodell enthaltenen Variablen um 16.04 Punkte. Die Ergebnisse sprechen zwar nur im begrenzten Maße für einen Einfluss der Bedrohung der Stereotype auf die Testleistung, können eine durch Stereotype Threat verursachte geschlechtsspezifische Leistungsdifferenz aber nicht grundsätzlich ausschließen. Auch die Berücksichtigung weiterer Interaktionsterme (Angst\*wichtig und Angst\*wichtig\*Jungen) kann dieses nebulöse Bild nicht aufhellen.

<b>Mathematik</b>	<b>Spanien</b>	<b>Italien</b>	<b>Portugal</b>
Jungen	6.30 <sup>ns</sup>	-25.02 <sup>ns</sup>	24.13 <sup>ns</sup>
Interesse	-7.43	-4.52	-18.01
Interesse*Jungen	1.52 <sup>ns</sup>	-7.85	-.78 <sup>ns</sup>
wichtig	23.46	-1.38 <sup>ns</sup>	4.73 <sup>ns</sup>
wichtig*Jungen	5.64 <sup>ns</sup>	27.09	1.53 <sup>ns</sup>
Selbstkonzept	21.03	10.36	25.30
Selbstkonzept*Jungen	-6.45	.55 <sup>ns</sup>	-3.22 <sup>ns</sup>
Selbstwirksamkeit	38.42	51.46	48.11
Selbstwirksamkeit*Jungen	-7.43	-1.97 <sup>ns</sup>	-4.42 <sup>ns</sup>
Mathematikangst	9.93	-11.87	-5.64 <sup>ns</sup>
Mathematikangst*Jungen	-2.72 <sup>ns</sup>	9.94	-1.77 <sup>ns</sup>
Kontrolle	1.79 <sup>ns</sup>	4.27	7.99
Kontrolle*Jungen	6.19	3.21 <sup>ns</sup>	7.19
Elaboration	-12.75	-20.31	-10.04
Elaboration*Jungen	5.83	-.56 <sup>ns</sup>	-5.01 <sup>ns</sup>
Memorieren	-2.23 <sup>ns</sup>	-7.93	-11.42
Memorieren*Jungen	-4.58	-4.82	-2.32 <sup>ns</sup>
Angst*wichtig	-7.24	-1.19 <sup>ns</sup>	-.18 <sup>ns</sup>
Angst*wichtig*Jungen	-.25 <sup>ns</sup>	-4.88	.26 <sup>ns</sup>
Konstante	321.73	377.56	315.97
Fallzahl	10.596	11.432	4492
r <sup>2</sup>	0.2869	0.3171	0.3932

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-19: Regressionsmodell Spanien, Italien und Portugal 2

In Portugal haben die neuen Interaktionsterme überhaupt keinen Effekt. In Spanien ist ein vom Geschlecht unabhängiger negativer Zusammenhang zwischen Angst\*wichtig und der Mathematikleistung zu beobachten, in Italien beschränkt sich dieser negative Zusammenhang auf das männliche Geschlecht. Die Mathematikangst hat bei den italienischen Schülern insbesondere dann einen negativen Einfluss auf die Mathematikleistung, wenn ihnen Mathematik wichtig ist. Die Ergebnisse können die theoretische Argumentation nicht stützen, aber auch nicht widerlegen.

## 6.2.4 Ländervergleich mit den Niederlanden, Belgien und Luxemburg

Der internationale Ländervergleich soll mit den Beneluxstaaten abgerundet werden. Beim PISA-Mathematiktest lagen die niederländischen und belgischen Schüler oberhalb des OECD-Durchschnitts, die Schüler aus Luxemburg unterhalb des OECD-Durchschnitts. Konkret erzielten die Jungen in Holland 540 Punkte, die Mädchen 535 Punkte (Mittelwert: 538 Punkte). Der geschlechtsspezifische Unterschied von fünf Punkten ist statistisch nicht signifikant. In Belgien lag die durchschnittliche Testleistung bei 529 Punkten. Die Jungen kamen auf 533 Punkte, die Mädchen erzielten 525 Punkte. Die Differenz von acht Punkten ist statistisch nicht abgesichert. In Luxemburg zeigten sich mit 17 Punkten größere geschlechtsspezifische Leistungsunterschiede zugunsten der Jungen (502 zu 485). Der Mittelwert lag bei 493 Punkten. Für jedes Land wurden einzelne multiple Regressionsmodelle berechnet, die Tabelle gibt einen Überblick über die Ergebnisse.

Mathematik	Niederlande	Belgien	Luxemburg
Jungen	12.96 <sup>ns</sup>	1.80 <sup>ns</sup>	24.40 <sup>ns</sup>
Interesse	-2.89 <sup>ns</sup>	6.76	-3.87 <sup>ns</sup>
Interesse*Jungen	5.13 <sup>ns</sup>	-3.12 <sup>ns</sup>	9.51
wichtig	-9.24	.69 <sup>ns</sup>	-3.87 <sup>ns</sup>
wichtig*Jungen	7.74	5.49	1.45 <sup>ns</sup>
Selbstkonzept	10.30	7.10	4.79 <sup>ns</sup>
Selbstkonzept*Jungen	-12.48	-6.32	-8.47
Selbstwirksamkeit	47.75	45.36	42.44
Selbstwirksamkeit*Jungen	1.53 <sup>ns</sup>	-2.40 <sup>ns</sup>	-2.33 <sup>ns</sup>
Mathematikangst	.58 <sup>ns</sup>	-9.26	-11.51
Mathematikangst*Jungen	-4.54 <sup>ns</sup>	-3.35 <sup>ns</sup>	-2.31 <sup>ns</sup>
Kontrolle	-13.04	-5.02	-6.37
Kontrolle*Jungen	9.87	6.08	8.42
Elaboration	-17.71	-25.60	-14.84
Elaboration*Jungen	-1.20 <sup>ns</sup>	4.76	-.30 <sup>ns</sup>
Memorieren	21.58	-.17 <sup>ns</sup>	3.88
Memorieren*Jungen	-15.90	-5.22	-10.79
Konstante	396.85	460.73	411.47
Fallzahl	3822	8338	3826
r <sup>2</sup>	0.2819	0.2553	0.3306

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-20: Regressionsmodell Niederlande, Belgien und Luxemburg 1

In den Niederlanden zeigt sich kein signifikanter Zusammenhang zwischen Punktzahl im Mathematiktest und Interesse. In Belgien ist ein geschlechtsunabhängiger Zusammenhang zu beobachten und in Luxemburg ist ausschließlich bei den Jungen ein positiver Zusammenhang zwischen Mathematikinteresse und Mathematikleistung zu erkennen. Sowohl in Holland als auch in Luxemburg stimmt jedoch die Richtung der Regressionskoeffizienten mit den theoretisch erwarteten Werten überein.

Der in den Niederlanden gefundene Zusammenhang zwischen instrumenteller Motivation (wichtig) und Mathematikleistung sowie mathematischem Selbstkonzept und Mathematikleistung entspricht nahezu perfekt den theoretischen Annahmen. Für die Mädchen ist ein negativer Zusammenhang zwischen dem Indikator für einen Prevention Focus und der Mathematikleistung zu beobachten und ein positiver Zusammenhang zwischen dem Indikator für einen Promotion Focus und der Mathematikleistung. Bei den Jungen zeigt sich das gegenteilige Muster. Grundsätzlich lässt sich diese Interpretation auch auf das Großherzogtum Luxemburg und auf Belgien übertragen, allerdings sind die betreffenden Regressionskoeffizienten teilweise nicht signifikant.

Insgesamt stützen die Ergebnisse in den Beneluxstaaten die Annahme, dass sich auch bei einer repräsentativen Stichprobe unter normalen Testbedingungen Hinweise finden lassen, die mit der Stereotype-Threat-Theorie und den aus dem MERF-Modell abgeleiteten Hypothesen kompatibel sind. Im nächsten Schritt werden die Analysen durch die Berücksichtigung der bereits bekannten Interaktionsterme ergänzt.

<b>Mathematik</b>	<b>Niederlande</b>	<b>Belgien</b>	<b>Luxemburg</b>
Jungen	15.41 <sup>ns</sup>	-5.58 <sup>ns</sup>	26.27 <sup>ns</sup>
Interesse	-3.02 <sup>ns</sup>	6.17	-4.69 <sup>ns</sup>
Interesse*Jungen	5.16 <sup>ns</sup>	-2.61 <sup>ns</sup>	9.81
wichtig	5.86 <sup>ns</sup>	17.86	11.93
wichtig*Jungen	4.21 <sup>ns</sup>	7.61 <sup>ns</sup>	-2.25 <sup>ns</sup>
Selbstkonzept	9.76	6.97	5.52
Selbstkonzept*Jungen	-11.93	-6.36	-8.71
Selbstwirksamkeit	47.26	44.70	41.33
Selbstwirksamkeit*Jungen	1.50 <sup>ns</sup>	-2.55 <sup>ns</sup>	-1.69 <sup>ns</sup>
Mathematikangst	12.43	2.01 <sup>ns</sup>	-1.73 <sup>ns</sup>
Mathematikangst*Jungen	-2.38 <sup>ns</sup>	1.70 <sup>ns</sup>	-.35 <sup>ns</sup>
Kontrolle	-13.20	-5.50	-6.34
Kontrolle*Jungen	9.88	6.18	7.91
Elaboration	-17.31	-25.55	-14.68
Elaboration*Jungen	-1.89 <sup>ns</sup>	4.04 <sup>ns</sup>	-.67 <sup>ns</sup>
Memorieren	21.45	-.36 <sup>ns</sup>	4.07
Memorieren*Jungen	-16.27	-4.87 <sup>ns</sup>	-10.90
Angst*wichtig	-6.60	-6.09	-5.72
Angst*wichtig*Jungen	.70 <sup>ns</sup>	-1.71 <sup>ns</sup>	.41 <sup>ns</sup>
Konstante	371.71	433.63	385.83
Fallzahl	3822	8338	3826
r <sup>2</sup>	0.2857	0.2616	0.3384

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-21: Regressionsmodell Niederlande, Belgien und Luxemburg 2

In allen drei Ländern ist der Interaktionsterm Angst\*wichtig signifikant. Personen, denen Mathematik wichtig ist, waren beim PISA-Mathematiktest besonders angespannt. Durch diese Anspannung waren die Schüler mit für die Aufgabenbewältigung irrelevanten Kognitionen

beschäftigt, die die Fähigkeit, eine Aufgabe korrekt zu lösen, reduziert haben. Die Ergebnisse bestätigen den Verdacht, dass situationale Aspekte einer Testsituation einen Einfluss auf die Testleistung haben.

### **6.2.5 Kurze Zwischenbetrachtung**

Der internationale Vergleich sollte Anhaltspunkte liefern, ob die geschlechtsspezifischen Leistungsunterschiede auch in anderen Ländern auf Stereotype Threat zurückgeführt werden können. In zehn der 13 ausgewählten Länder zeigen sich signifikante negative Zusammenhänge zwischen dem Mathematikinteresse und der erreichten Punktzahl beim PISA-Mathematiktest, die weitgehend auf die Mädchen begrenzt sind. In zwei Ländern (Niederlande und Luxemburg) sind diese negativen Zusammenhänge nicht signifikant, in Belgien ist der Zusammenhang positiv.

Aufschlussreich sind auch die geschlechtsspezifischen Zusammenhänge bei der instrumentellen Motivation und dem Testergebnis sowie dem mathematischen Selbstkonzept und der erreichten Punktzahl. In dieser Arbeit dient die instrumentelle Motivation als Indikator für einen chronischen Prevention Focus, das mathematische Selbstkonzept als Indikator für einen chronischen Promotion Focus. Vor dem Hintergrund dieser Annahme sind die empirischen Befunde weitgehend mit dem MERF-Modell kompatibel. Das MERF-Modell postuliert je nach Zugänglichkeit zu einem regulatorischen Fokus einen unterschiedlichen Einfluss der Erwartung auf die Leistung. In einem Prevention Focus werden negative Erwartungen als Bedrohung gewertet. Der aus dieser Bedrohung resultierende Druck führt schließlich zu einer geringeren Testleistung. Da beim PISA-Mathematiktest primär die Schülerinnen mit negativen stereotypen Erwartungen konfrontiert waren, sollte der negative Zusammenhang zwischen instrumenteller Motivation und Testleistung im Wesentlichen auf die weiblichen Personen beschränkt sein. Für die Mädchen zeigen sich in fünf der 13 Länder signifikante negative Zusammenhänge zwischen der instrumentellen Motivation und der Mathematikleistung, in sieben Ländern sind die Zusammenhänge nicht signifikant, in Norwegen ist der Zusammenhang positiv. Für die Schülerinnen mit einer erhöhten Zugänglichkeit zu einem Promotion Focus wurde ein positiver Zusammenhang zwischen dem mathematischem Selbstkonzept und der Testleistung erwartet. Mit Ausnahme von Luxemburg sind in zwölf der 13 Länder bei den weiblichen Testpersonen positive Zusammenhänge zwischen dem mathematischen Selbstkonzept und der erreichten Punktzahl im PISA-Test zu beobachten. Bei den Jungen, die beim PISA-Test tendenziell mit einer positiven Erwartung konfrontiert waren, sind diese Zusammenhänge negativ beziehungsweise deutlich schwächer ausgeprägt als bei den Mädchen. Auch dieser Befund steht im Einklang mit den theoretischen Annahmen.

Jedoch zeigen sich in einigen Ländern – insbesondere in den skandinavischen Ländern – entweder keine signifikanten Zusammenhänge zwischen den Indikatoren der regulatorischen Fokusse und der Punktzahl im PISA-Mathematiktest oder es sind den theoretischen Annahmen entgegenlaufende Befunde zu beobachten. Dies ist möglicherweise auf die konzeptionell problematische Erfassung der regulatorischen Fokusse zurückzuführen. Andererseits ist anzunehmen, dass die geschlechtsspezifischen Erwartungen der Schüler zumindest in ihrer Stärke von Land zu Land variieren. Bei der PISA-Studie wurde die wahrgenommene Stereotypisierung von Mathematik als männliche Domäne bedauerlicherweise nicht erfasst, so dass die empirischen Analysen an dieser Stelle begrenzt sind.

Zusammenfassend lassen die Ergebnisse jedoch den Schluss zu, dass die geschlechtsspezifischen Leistungsunterschiede in der Mathematik auf Stereotype Threat und damit auf situationalen Faktoren zurückgeführt werden können. Da bei PISA sowohl schwierige als auch leichte Aufgaben verwendet wurden, unterstreichen die Befunde die Stärke des Stereotype-Threat-Effekts bei nicht selektiven Stichproben und unter normalen Testbedingungen. Stereotype Threat ist folglich nicht auf Laborbedingungen begrenzt, sondern stellt auch außerhalb des Labors eine Bedrohung der Leistungsfähigkeit dar.

Zwischen den Ländern zeigen sich jedoch beachtliche Unterschiede in dem Ausmaß der Leistungsdifferenzen zwischen Mädchen und Jungen. Diese Länderunterschiede stehen im Mittelpunkt des folgenden Abschnitts.

### 6.3 Länderunterschiede in den geschlechtsspezifischen Leistungsdifferenzen

Die Tabelle gibt einen Überblick über die Stärke der geschlechtsspezifischen Leistungsdifferenzen im PISA-Mathematiktest in den einzelnen Ländern.

Land	d-Wert	Punktzahldifferenz zugunsten der Jungen
Italien	0.19	18
Luxemburg	0.19	17
Dänemark	0.18	17
Schweiz	0.17	17
Portugal	0.14	12
Spanien	0.10	9
Deutschland	0.09	9
Finnland	0.09	7
Österreich	0.08	8 (nicht signifikant)
Belgien	0.08	8 (nicht signifikant)
Schweden	0.07	7
Norwegen	0.07	6 (nicht signifikant)
Niederlande	0.06	5 (nicht signifikant)

Anmerkung: nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-22: Geschlechtsspezifische Leistungsunterschiede in den Ländern (sortiert nach d-Werten)

In Italien, Luxemburg, Dänemark oder der Schweiz sind die geschlechtsspezifischen Leistungsunterschiede deutlich stärker ausgeprägt als in Österreich, Belgien, den Niederlanden oder Norwegen. Die große Varianz, die bei den geschlechtsspezifischen Unterschieden zwischen den Ländern besteht, lässt darauf schließen, dass die gegenwärtigen Leistungsunterschiede zwischen Mädchen und Jungen nicht unabdingbar sind. Können die geschlechtsspezifischen Leistungsdifferenzen zwischen den Ländern durch Unterschiede in den motivationalen Schülermerkmalen erklärt werden? Um einen vermittelnden Effekt der motivationalen Variablen zu klären, müssen drei verschiedene Regressionsmodelle berechnet werden.

$$Y_{(\text{Punktzahl})} = \beta_{10} + \beta_{11}\text{Land} + \beta_{12}\text{Geschlecht} + \beta_{13}\text{Land*Geschlecht} + \varepsilon_1 \quad (1)$$

$$Y_{(\text{Mediator})} = \beta_{20} + \beta_{21}\text{Land} + \beta_{22}\text{Geschlecht} + \beta_{23}\text{Land*Geschlecht} + \varepsilon_2 \quad (2)$$

$$Y_{(\text{Punktzahl})} = \beta_{30} + \beta_{31}\text{Land} + \beta_{32}\text{Geschlecht} + \beta_{33}\text{Mediator} + \beta_{34}\text{Land*Geschlecht} + \varepsilon_3 \quad (3)$$

Dem Vorschlag von Muller, Judd und Yzerbyt (in Druck) zur Durchführung von Mediationsanalysen folgend, muss in einem ersten Schritt geprüft werden, ob sich die geschlechtsspezifischen Leistungsunterschiede zwischen zwei Ländern signifikant unterscheiden. Dazu wird ein Regressionsmodell mit der abhängigen Variable Punktzahl im PISA-Mathematiktest und den unabhängigen Variablen Land, Geschlecht und dem Interaktionsterm Land\*Geschlecht berechnet. Voraussetzung für weitere Analysen ist, dass der Interaktionsterm  $\beta_{13}\text{Land*Geschlecht}$  in der ersten Gleichung signifikant ist.

Im zweiten Schritt werden potenzielle Mediatoren der Unterschiede der geschlechtsspezifischen Leistungsdifferenzen zwischen den Ländern getestet. Als abhängige Variable dienen dabei jeweils die motivationalen Merkmale der Schüler, also Mathematikinteresse, instrumentelle Motivation, Selbstkonzept, Selbstwirksamkeit und Mathematikangst. In der zweiten Gleichung sollte der Interaktionsterm  $\beta_{23}\text{Land*Geschlecht}$  signifikant sein.

Im dritten Schritt der Analyse wird der Effekt der potenziellen Mediatorvariable auf den Interaktionsterm getestet. In diesem dritten Regressionsmodell sollte die potenzielle Mediatorvariable ( $\beta_{33}\text{Mediator}$ ) signifikant sein und der Interaktionsterm zwischen Land\*Geschlecht ( $\beta_{34}\text{Land*Geschlecht}$ ) einen geringeren Betrag aufweisen als der Interaktionsterm Land\*Geschlecht in der ersten Gleichung ( $\beta_{13}\text{Land*Geschlecht}$ ).

Die folgenden Analysen werden wieder getrennt nach den Ländergruppen durchgeführt. Für die Analysen wurden jeweils die Länder einer Gruppe ausgewählt, die sich im Ausmaß der geschlechtsspezifischen Leistungsdifferenzen am stärksten unterscheiden: in der Gruppe Zentraleuropa die Schweiz und Österreich, in der Gruppe der skandinavischen Länder Däne-

mark und Norwegen, in der Gruppe der südeuropäischen Länder Spanien und Italien sowie in der Ländergruppe der Beneluxstaaten Luxemburg und die Niederlande/Belgien.

### 6.3.1 Vergleich Schweiz und Österreich

In der Schweiz beträgt der geschlechtsspezifische Leistungsunterschied 17 Punkte, in Österreich lediglich acht Punkte. Bei dem Regressionsmodell bilden Österreich und die Mädchen die Referenzkategorien, die bei der Rekodierung jeweils die 0 erhielten.

Mathematik	Koeffizient	t-Wert	p-Wert
Jungen	7.57	3.31	0.001
Schweiz	16.14	6.99	0.000
Schweiz*Jungen	9.05	2.80	0.005
Konstante	501.82	309.42	0.000

Fallzahl: 13.017

$r^2 = 0.0175$

Tabelle 6-23: Vergleich Österreich und Schweiz 1

Die Schülerinnen in Österreich erhielten beim PISA-Mathematiktest durchschnittlich 501.82 Punkte, die Jungen in Österreich 7.57 Punkte mehr.<sup>34</sup> In der Schweiz erzielten die Schülerinnen durchschnittlich 16.14 Punkte mehr als die Schülerinnen in Österreich, der geschlechtsspezifische Leistungsvorsprung zugunsten der Jungen vergrößert sich gegenüber Österreich um neun Punkte und beträgt knapp 17 Punkte. Der Interaktionsterm Schweiz\*Jungen ist signifikant, die Grundvoraussetzung für die weiteren Analysen ist gegeben.

Im nächsten Schritt wird geprüft, ob sich die Schüler in Österreich und in der Schweiz in den motivationalen Merkmalen signifikant unterscheiden. Dazu werden Regressionen berechnet, bei denen die motivationalen Merkmale der Schüler – Interesse, instrumentelle Motivation (wichtig), Konzept, Selbstwirksamkeit und Angst – jeweils als abhängige Variable dienen.

	Interesse	wichtig	Konzept	Selbstwirksamkeit	Angst
Jungen	.42	.58	.47	.46	-.42
Schweiz	.30	.38	-.07	.11	.01 <sup>ns</sup>
Schweiz*Jungen	.18	.12	.24	.08	-.06 <sup>ns</sup>
Konstante	1.29	1.59	1.96	3.82	2.42
Fallzahl	12.911	12.896	12.870	12.875	12.872
$r^2$	0.0962	0.1407	0.0823	0.0693	0.0399

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-24: Vergleich Österreich und Schweiz 2

Es zeigt sich, dass lediglich Angst als Mediator ausscheidet, weil der betreffende Interaktionsterm nicht signifikant ist. Bei den Analysen mit dem mathematischen Interesse, der instru-

<sup>34</sup> Die Leistungsdifferenz zwischen Mädchen und Jungen ist in Österreich nur bei Anwendung eines üblichen t-Tests auf dem Fünf-Prozent-Niveau signifikant. Unter Berücksichtigung des Stichprobenfehlers (Klumpeneffekt) kann der Unterschied nicht zufallskritisch abgesichert werden.

mentellen Motivation (wichtig), dem Selbstkonzept und der Selbstwirksamkeit als abhängige Variable ist der Interaktionsterm Schweiz\*Jungen jeweils signifikant.

Im dritten Schritt wird jeweils der Einfluss eines potenziellen Mediators auf den Interaktionsterm berechnet. In den jeweiligen Modellen sollte die Mediatorvariable signifikant werden und der Interaktionsterm einen geringeren Betrag aufweisen als in der Ausgangsanalyse. Die erste Spalte der Tabelle fasst die Ergebnisse der Ausgangsanalyse zusammen, in den anderen Spalten wurde jeweils eine potenzielle Mediatorvariable dem Regressionsmodell hinzugefügt.

Mathematik	Koeffizient	Koeffizient	Koeffizient	Koeffizient	Koeffizient
Jungen	7.57	5.12	12.02	-3.77 <sup>ns</sup>	-14.43
Schweiz	16.14	12.48	17.53	16.29	9.16
Interesse		8.51			
wichtig			-5.57		
Selbstkonzept				25.21	
Selbstwirksamkeit					50.92
Schweiz*Jungen	9.05	6.69	8.46	3.12 <sup>ns</sup>	3.87 <sup>ns</sup>
Konstante	501.82	492.18	512.05	454.29	308.81
Fallzahl	13.017	12.911	12.896	12.870	12.875
r <sup>2</sup>	0.0175	0.0249	0.0199	0.0942	0.3045

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-25: Vergleich Österreich und Schweiz 3

Der stärkste betragsmäßige Rückgang des Interaktionsterms ist bei Berücksichtigung des mathematischen Selbstkonzepts zu beobachten. Hier sinkt der Interaktionsterm von 9.05 auf 3.12. Bei Kontrolle des mathematischen Selbstkonzepts unterscheiden sich die Schweiz und Österreich hinsichtlich des Ausmaßes der Leistungsdifferenzen zwischen Mädchen und Jungen nicht mehr. Ein starker betragsmäßiger Rückgang des Interaktionsterms ist auch bei Berücksichtigung der Selbstwirksamkeit (von 9.05 auf 3.87) zu sehen, bei Kontrolle des mathematischen Interesses und der instrumentellen Motivation ist der Rückgang schwächer.

Die Analysen liefern klare Evidenz für die These, dass die Unterschiede in den geschlechtsspezifischen Leistungsdifferenzen zwischen Österreich und der Schweiz auf Unterschiede in den motivationalen Merkmalen der Schüler zurückzuführen sind. Wie sieht es bei den skandinavischen Ländern aus?

### 6.3.2 Vergleich Dänemark und Norwegen

Die geschlechtsspezifischen Leistungsunterschiede sind in den skandinavischen Ländern insgesamt eher unauffällig. In Finnland und Schweden beträgt die Differenz zwischen Mädchen und Jungen jeweils sieben Punkte, in Norwegen sogar nur sechs Punkte. Einzige Ausnahme ist Dänemark. Dort liegen zwischen dem durchschnittlichen Ergebnis der Mädchen und dem durchschnittlichen Testergebnis der Jungen 17 Punkte.

Für den Vergleich wurden Norwegen und Dänemark ausgewählt, die Tabelle gibt einen Überblick über die Ergebnisse des ersten Regressionsmodells.

Mathematik	Koeffizient	t-Wert	p-Wert
Jungen	6.22	2.29	0.022
Dänemark	14.10	5.19	0.000
Dänemark*Jungen	10.36	2.69	0.007
Konstante	492.05	255.49	0.000

Fallzahl: 8282

$r^2 = 0.0166$

Tabelle 6-26: Vergleich Dänemark und Norwegen 1

Die geschlechtsspezifische Leistungsdifferenz nimmt in Dänemark gegenüber Norwegen um rund 10.36 Punkte zu, der Interaktionsterm Dänemark\*Jungen ist signifikant. Im nächsten Schritt werden wieder potenzielle Mediatoren des Effekts getestet. Dazu werden jeweils Regressionsmodelle gerechnet, bei denen die motivationalen Merkmale die abhängige Variable bilden.

	Interesse	wichtig	Konzept	Selbstwirksamkeit	Angst
Jungen	.26	.24	.47	.40	-.39
Dänemark	.57	.15	.41	-.03 <sup>ns</sup>	-.40
Dänemark*Jungen	.02 <sup>ns</sup>	.15	.02 <sup>ns</sup>	.02 <sup>ns</sup>	-.01 <sup>ns</sup>
Konstante	1.49	2.41	1.71	3.65	2.62
Fallzahl	8164	8170	8147	8151	8149
$r^2$	0.0907	0.0399	0.0852	0.0408	0.0673

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-27: Vergleich Dänemark und Norwegen 2

Nur bei der instrumentellen Motivation (wichtig) ist der Interaktionsterm signifikant. Abschließend wird geprüft, ob sich der Interaktionsterm Dänemark\*Jungen der Ausgangsanalyse bei Berücksichtigung des potenziellen Mediators „wichtig“ verändert. Die Tabelle zeigt die Ergebnisse der Ausgangsanalyse und die Ergebnisse des um die potenzielle Mediatorvariable erweiterten Regressionsmodells.

Mathematik	Koeffizient	Koeffizient
Jungen	6.22	1.89 <sup>ns</sup>
Dänemark	14.10	10.72
wichtig		24.72
Dänemark*Jungen	10.36	5.89 <sup>ns</sup>
Konstante	492.05	432.77
Fallzahl	8282	8170
$r^2$	0.0166	0.0887

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-28: Vergleich Dänemark und Norwegen 3

Es ist ein starker betragsmäßiger Rückgang von 10.36 auf 5.89 zu beobachten. Der Interaktionsterm Dänemark\*Jungen ist bei Berücksichtigung der Bedeutung der Mathematik (wichtig) im Vergleich zur Ausgangsanalyse nicht mehr signifikant. Die Unterschiede in den geschlechtsspezifischen Leistungsdifferenzen zwischen Dänemark und Norwegen sind demnach auf Unterschiede in der instrumentellen Motivation zurückzuführen.

### 6.3.3 Vergleich Italien und Spanien

In Italien beträgt die geschlechtsspezifische Leistungsdifferenz 18 Punkte zugunsten der Jungen, in Spanien ist der Abstand zwischen Mädchen und Jungen mit neun Punkten nur halb so groß. Die Tabelle gibt einen Überblick über die Ergebnisse der Ausgangsanalyse.

Mathematik	Koeffizient	t-Wert	p-Wert
Jungen	8.86	4.83	0.000
Italien	-23.65	-14.10	0.000
Italien*Jungen	8.96	3.73	0.000
Konstante	480.74	373.74	0.000

Fallzahl: 22.429

$r^2 = 0.0183$

Tabelle 6-29: Vergleich Italien und Spanien 1

Die spanischen Schülerinnen erreichten beim PISA-Mathematiktest durchschnittlich 481 Punkte, die Jungen erzielten knapp neun Punkte mehr. In Italien kamen die Mädchen durchschnittlich auf 457 Punkte, die Jungen erreichten 474.9 Punkte. Die Mädchen in Spanien waren bei der internationalen Schulleistungsstudie also erfolgreicher als die Jungen in Italien. Die geschlechtsspezifische Differenz ist in Italien jedoch deutlich größer als in Spanien. Die folgende Tabelle gibt einen Überblick über die Ergebnisse der Regressionsmodelle, bei denen die motivationalen Merkmale der Schüler jeweils die abhängige Variable darstellen.

	Interesse	wichtig	Konzept	Selbstwirksamkeit	Angst
Jungen	.03 <sup>ns</sup>	.09	.25	.25	-.29
Italien	.10	-.16	.25	-.09	-.07
Italien*Jungen	.08	.13	-.11	.05	.15
Konstante	1.70	2.29	1.81	3.73	2.90
Fallzahl	22.299	22.307	22.267	22.260	22.265
$r^2$	0.0068	0.0107	0.0181	0.0278	0.0160

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-30: Vergleich Italien und Spanien 2

Vor dem Hintergrund der großen Fallzahl sind alle Interaktionsterme signifikant. Es kommen demnach alle motivationalen Merkmale als Mediatoren in Frage.

Die folgende Tabelle enthält in der ersten Spalte wieder die Ergebnisse der Ausgangsanalyse, in den anderen Spalten sind die Ergebnisse der Regressionen mit den entsprechenden potenziellen Mediatorvariablen eingetragen. Dabei sollte die potenzielle Mediatorvariable signifikant sein und der Interaktionsterm Italien\*Jungen einen geringeren Betrag aufweisen als in der Ausgangsanalyse (Spalte 1).

Mathematik	Koeffizient	Koeffizient	Koeffizient	Koeffizient	Koeffizient	Koeffizient
Jungen	8.86	9.25	8.17	2.72 <sup>ns</sup>	-1.81 <sup>ns</sup>	.96 <sup>ns</sup>
Italien	-23.65	-25.37	-22.15	-30.84	-20.01	-26.28
Interesse		14.34				
wichtig			12.90			
Selbstkonzept				27.63		
Selbstwirksamkeit					47.81	
Mathematikangst						-29.99
Italien*Jungen	8.96	7.66	7.48	11.63	5.74	13.14
Konstante	480.74	456.88	451.87	431.46	303.27	568.52
Fallzahl	22.429	22.299	22.307	22.267	22.260	22.265
r <sup>2</sup>	0.0183	0.0430	0.0394	0.1160	0.2284	0.1007

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 6-31: Vergleich Italien und Spanien 3

Der stärkste betragsmäßige Rückgang des Interaktionsterms Italien\*Jungen ist bei Berücksichtigung der Selbstwirksamkeit festzustellen (8.96 auf 5.74). Der betragsmäßige Rückgang des Interaktionsterms ist bei den Variablen Interesse (8.96 auf 7.66) und wichtig (8.96 auf 7.48) geringer. Bei einer zusätzlich gerechneten Regression mit den Variablen Interesse, wichtig und Selbstwirksamkeit waren die beiden erstgenannten Variablen nicht mehr signifikant, der Interaktionsterm verringerte sich von 8.96 auf 5.88.

Da der Interaktionsterm weiterhin signifikant bleibt, kann der Länderunterschied in den geschlechtsspezifischen Leistungen zwischen Italien und Spanien nicht komplett auf Unterschiede in den untersuchten Schülermerkmalen zurückgeführt werden. Es handelt sich im vorliegenden Fall um eine partielle Mediation.

### 6.3.4 Vergleich Luxemburg und Niederlande/Belgien

Die geschlechtsspezifische Leistungsdifferenz beträgt in den Niederlanden fünf und in Belgien acht Punkte, wobei der Unterschied jeweils statistisch nicht abgesichert werden kann. In Luxemburg erzielten die Jungen mit 502 Punkten durchschnittlich 17 Punkte mehr als die Mädchen, die 485 Punkte erreichen konnten.

Bei der Analyse der Länderunterschiede in den geschlechtsspezifischen Leistungsdifferenzen war der Interaktionsterm Geschlecht\*Land weder für die Niederlande ( $p = 0.386$ ) noch für Belgien ( $p = 0.344$ ) signifikant, so dass die Grundvoraussetzung für weitere Analysen nicht erfüllt war.

### 6.3.5 Kurze Zwischenbetrachtung

Die Analysen in diesem Abschnitt sollten erste Hinweise liefern, ob die Länderunterschiede in den geschlechtsspezifischen Leistungsdifferenzen auf Unterschiede in den motivationalen Merkmalen der Schüler zurückzuführen sind. Bei dem Vergleich zwischen der Schweiz und Österreich sowie Dänemark und Norwegen konnten Mediatoren identifiziert werden, die die Länderunterschiede in den Leistungsdifferenzen zwischen Mädchen und Jungen komplett erklären. Bei dem Vergleich zwischen der Schweiz und Österreich war der betreffende Interaktionsterm Schweiz\*Jungen nach Berücksichtigung des mathematischen Selbstkonzepts beziehungsweise der Selbstwirksamkeit nicht mehr signifikant, der Länderunterschied wurde durch die Unterschiede in den motivationalen Merkmalen der Schüler erklärt. Bei dem Vergleich zwischen Dänemark und Norwegen konnte die instrumentelle Motivation (wichtig) als Mediatorvariable identifiziert werden.

Eine partielle Mediation konnte bei der Analyse der Länderunterschiede zwischen Italien und Spanien nachgewiesen werden. Der stärkste betragsmäßige Rückgang des Interaktionsterms Italien\*Jungen war bei Berücksichtigung der Selbstwirksamkeit festzustellen (von 8.96 auf 5.74). Der Interaktionsterm blieb jedoch signifikant, so dass die Länderunterschiede nicht komplett aufgeschlüsselt werden konnten. Bei der Analyse der Länderunterschiede in den Beneluxstaaten war weder für den Vergleich Niederlande/Luxemburg noch für den Vergleich Belgien/Luxemburg der Interaktionsterm Land\*Geschlecht signifikant. Alles in allem legen die Ergebnisse in diesem Abschnitt nahe, dass die Länderunterschiede in den geschlechtsspezifischen Leistungsdifferenzen in erster Linie auf Unterschiede in den motivationalen Schülermerkmalen zurückzuführen sind.

Die bisherigen Analysen haben sich auf die geschlechtsspezifischen Leistungsunterschiede in der Mathematik beschränkt. Im nächsten Kapitel sollen diese Analysen durch eine Erweiterung der Perspektive auf geschlechtsspezifische Leistungsunterschiede im naturwissenschaftlichen Bereich ergänzt werden, ehe die empirischen Ergebnisse zusammenfassend erörtert werden.

## 7. Von der Mathematik zu den Naturwissenschaften

Im Mittelpunkt der bisherigen Analysen standen die geschlechtsspezifischen Leistungsunterschiede in der Mathematik. Dabei konnten regressionsanalytisch Hinweise gefunden werden, die mit der Stereotype-Threat-Theorie und dem MERF-Modell kompatibel sind. Das Phänomen der Bedrohung durch Stereotype ist aber nicht auf die Mathematik beschränkt, sondern bezieht sich auf alle Situationen, in denen negative stereotype Erwartungen auf das Verhalten der Personen anwendbar sind. Deshalb werden die bisherigen Analysen durch eine Untersuchung der geschlechtsspezifischen Leistungsunterschiede in den naturwissenschaftlichen Fächern ergänzt.

Dabei ist zwischen den naturwissenschaftlichen Fächern zu differenzieren: Mädchen und Jungen haben unterschiedliche Präferenzen für die naturwissenschaftlichen Fächer und zeigen auch deutlich unterschiedliche Leistungen. Mädchen schneiden typischerweise in Biologie besser ab als Jungen, während Jungen in Physik und Chemie den Mädchen überlegen sind (Prenzel, Carstensen, Rost & Senkbeil, 2002). Biologie gilt allgemein eher als weibliche Domäne, Physik und Chemie gelten als männliche Domänen. Entsprechend dieser Argumentation sind Mädchen eher in Physik und Chemie, die Jungen eher in Biologie mit negativen stereotypen Erwartungen konfrontiert.

Bei den Analysen der geschlechtsspezifischen Leistungsunterschiede in den Naturwissenschaften kann jedoch nicht der internationale PISA-Datensatz 2003 verwendet werden. Angelsächsische Länder, aber auch viele andere OECD-Staaten kennen die im deutschen Schulsystem vorgenommene Aufgliederung der Naturwissenschaften in die Fächer Biologie, Chemie und Physik nicht, so dass die internationalen PISA-Science-Aufgaben nicht diesen drei Fächern zugeordnet waren. Bei der deutschen PISA-Erweiterungsstudie 2000 (PISA-E), die primär darauf abzielte, die Leistungen der deutschen Schüler auch auf der Ebene der Bundesländer zu vergleichen, wurde der internationale Naturwissenschaftstest durch ein nationales Erhebungsinstrument ergänzt, das grundsätzlich eine Differenzierung nach den Fächern Biologie, Chemie und Physik erlaubt. Der nationale Test stellt damit zusätzliche Informationen bereit, die Aussagen über die in den einzelnen naturwissenschaftlichen Fächern erzielten Kompetenzen erlauben. Außerdem wurden motivationale Merkmale – Selbstwirksamkeit, das Interesse an Biologie, Chemie und Physik oder auch das akademische Selbstkonzept – erfasst. Bei den folgenden Analysen wird deshalb auf den deutschen PISA-Erweiterungsdatensatz<sup>35</sup> 2000 zurückgegriffen.

---

<sup>35</sup> Der Datensatz der nationalen PISA-Erweiterungsstudie 2000 (PISA-E) steht auf der Homepage des Sekretariats der deutschen Kultusministerkonferenz unter <http://www.kmk.org/schul/pisa/Datensaetze/daten.htm> zur Verfügung. Die SPSS-Datei wurde mit dem Programm Stat/Transfer in einen Stata-Datensatz (Version 7.0) überführt. Auf der beiliegenden DVD ist der Datensatz als SPSS-Datei und als Stata-Datei enthalten.

Zunächst werden das Design der Studie und die Messinstrumente kurz vorgestellt, ehe der regressionsanalytische Indizienprozess auf die naturwissenschaftlichen Fächer ausgeweitet werden soll.

### **7.1 Design und Messinstrumente**

Die Konzeption, der Testansatz sowie die Messinstrumente der deutschen PISA-Erweiterungsstudie 2000 (PISA-E) stimmen weitgehend mit denen der internationalen PISA-Studie 2003 überein, weshalb auf eine ausführliche Darstellung des Designs der PISA-Erweiterungsstudie verzichtet werden kann. An dieser Stelle werden lediglich die wesentlichen Merkmale der PISA-E-Studie skizziert.<sup>36</sup>

Die nationale Erweiterungsstudie unterscheidet sich von den internationalen PISA-Studien 2000 und 2003 in erster Linie durch die Vergrößerung des Stichprobenumfangs, um die Leistungen der Schüler auch zwischen den Ländern der Bundesrepublik Deutschland vergleichen zu können. Für jedes Land und jeden Stadtstaat wurde eine Schülerstichprobe gezogen, die repräsentativ für die 15-Jährigen des betreffenden Landes ist und statistisch abgesicherte Aussagen über die Ergebnisse in den einzelnen Bundesländern und pro Schulform erlaubt. Insgesamt sind im Rahmen der PISA-Erweiterungsstudie über 48.000 Schüler aus 1479 Schulen erfasst worden (Baumert, Artelt, Carstensen, Sibberns & Stanat, 2002). Da bei der PISA-Studie 2000 jedoch die Lesekompetenz den Schwerpunkt der Erhebung bildete, reduziert sich die Fallzahl bei den Analysen der naturwissenschaftlichen Kompetenzen – insbesondere bei Berücksichtigung der motivationalen Merkmale der Schüler – deutlich.

Für den Naturwissenschaftstest der PISA-Erweiterungsstichprobe wurden insgesamt neun thematische Aufgabenblöcke mit insgesamt 46 Items entwickelt. Bei der Konstruktion der Aufgaben für die nationale Zusatzstudie wurde das internationale Aufgabenformat übernommen. Ein kurzer Vorspann diente zur Einbettung der Aufgaben in einen für die Schüler vertrauten und anregenden Kontext. Der Test bestand aus Aufgabenblöcken für die Fächer Biologie (vier Aufgabenblöcke mit jeweils vier Items), Chemie (ein Aufgabenblock mit acht Items) und Physik (vier Aufgabenblöcke mit jeweils vier bis sechs Items). Die einzelnen Items hatten dabei unterschiedliche Schwierigkeitsgrade, es gab also leichte und schwierige Aufgaben (Prenzel, Rost, Senkbeil, Häußler & Klopp, 2001).

Nach dem Bearbeiten der Aufgaben waren die Schüler aufgefordert, einen Fragebogen auszufüllen, der kontextbezogene Merkmale erfasst, mit denen ein Zusammenhang zwischen den

---

<sup>36</sup> Für eine ausführliche Diskussion zur Konzeption der deutschen PISA-Erweiterungsstudie 2000 (PISA-E) sei an dieser Stelle auf die Veröffentlichungen des Deutschen PISA-Konsortiums „PISA 2000 – Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich“ (2001), „PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich“ (2002) und „PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland“ (2003) verwiesen.

Leistungsergebnissen und den Schülermerkmalen hergestellt werden kann. Die folgende Tabelle gibt einen Überblick über die bei den Analysen verwendeten Messinstrumente<sup>37</sup> und ihre Reliabilitäten.

Skalenbezeichnung	Anzahl der Items	Beispielaussage	Reliabilität (Cronbach's Alpha)
akademisches Selbstkonzept	3	In den meisten Schulfächern schneide ich in Klassenarbeiten gut ab.	0.7633
Selbstkonzept Mathematik	3	Mathematik ist eines meiner besten Fächer.	0.8925
Selbstkonzept Biologie	3	Ich bin für Biologie begabt.	0.8746
Selbstkonzept Chemie	3	Mir fällt Chemie leicht.	0.9095
Selbstkonzept Physik	3	Ich bringe in Physik gute Leistungen.	0.9113
Interesse Mathematik	3	Mathematik ist mir persönlich wichtig.	0.7531
Interesse Biologie	2	Ich lerne in Biologie etwas, das für mich sehr wichtig ist.	0.6190
Interesse Chemie	2	Mir macht Chemieunterricht Spaß.	0.7310
Interesse Physik	2	Ich lerne in Physik etwas, das für mich sehr wichtig ist.	0.7832
Selbstwirksamkeit	4	Ich bin überzeugt, dass ich die Fertigkeiten, die gelehrt werden, beherrschen kann.	0.7581
Lernstrategien: Wiederholung	4	Wenn ich lerne, lerne ich so viel wie möglich auswendig.	0.7474
Lernstrategien: Kontrolle	5	Wenn ich lerne, überlege ich mir zuerst, was genau ich lernen muss.	0.7254
Lernstrategien: Elaboration	4	Wenn ich lerne, überlege ich, wie der Stoff mit dem zusammenhängt, was ich schon gelernt habe.	0.7535

Tabelle 7-1: Verwendete Messinstrumente bei den Analysen mit dem PISA-E-Datensatz 2000

Die einzelnen Skalen wurden jeweils als Mittelwert der Items gebildet. Es erfolgte keine Indexbildung, wenn ein Wert auf einem zur Indexbildung benötigten Item fehlte. Die Skalen wurden jeweils so kodiert, dass höhere Werte auf der Skala eine größere Zustimmung bedeuten.

<sup>37</sup> Für ausführliche Informationen zu den verwendeten Messinstrumenten sei an dieser Stelle auf „PISA 2000: Dokumentation der Erhebungsinstrumente“ von Kunter, Schümer, Artelt, Baumert, Klieme, Neubrand, Prenzel, Schiefele, Schneider, Stanat, Tillmann und Weiß (2002) verwiesen.

Bei den folgenden Analysen soll geprüft werden, ob sich auch bei den naturwissenschaftlichen Fächern Hinweise finden lassen, die im Einklang mit der Stereotype-Threat-Theorie stehen. Im Mittelpunkt steht dabei das fachspezifische Interesse. Der Stereotype-Threat-Effekt ist insbesondere bei Personen zu erwarten, die sich durch eine hohe Domainidentifizierung – bei PISA durch das Interesse konzeptionalisiert – auszeichnen.

## 7.2 Empirische Analysen

Der Ausgangspunkt der Analysen ist eine Einfachregression mit der Leistung in Biologie, Chemie und Physik als abhängigen Variablen und der unabhängigen Variable Geschlecht, das – analog zu den Analysen mit dem internationalen PISA-Datensatz 2003 – dummykodiert ist. Die Referenzkategorie bilden die Mädchen (0), die Jungen erhielten bei der Rekodierung die 1. Ergänzend zu den Analysen der naturwissenschaftlichen Fächer werden auch die Ergebnisse der mathematischen Grundbildung bei der PISA-Erweiterungsstudie 2000 berichtet.

	<b>Mathematik</b>	<b>Biologie</b>	<b>Chemie</b>	<b>Physik</b>
Jungen	13.38	-1.19 <sup>ns</sup>	22.00	29.23
Konstante	493.14	504.16	478.73	477.65
Fallzahl	17.436	31.887	31.887	31.887
r <sup>2</sup>	0.0048	0.000	0.0108	0.0292

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 7-2: Regressionsmodell 1 (PISA-E-Studie)

Bei der deutschen PISA-Erweiterungsstudie 2000 zeigten sich in Mathematik, Chemie und Physik Leistungsvorteile zugunsten der Jungen. In Mathematik kamen die Schülerinnen durchschnittlich auf 493 Punkte, die Jungen konnten durchschnittlich 13 Punkte mehr erzielen und kamen auf rund 506 Punkte. In Chemie beträgt der Vorsprung der Jungen 22 Punkte (478 zu 500) und in Physik 29 Punkte (477 zu 506). Lediglich in Biologie zeigen sich keine geschlechtsspezifischen Leistungsunterschiede. Die erklärte Varianz der vier Regressionsmodelle ist als gering einzuschätzen. In Mathematik ist weniger als ein Prozent der Gesamtvarianz erklärt, in Chemie gerade einmal ein Prozent und in Physik sind es knapp drei Prozent.

In Deutschland ist die Leistungsdifferenz zwischen Mädchen und Jungen ganz erheblich davon abhängig, ob die Gesamtpopulation betrachtet oder die Leistungen innerhalb von Bildungsgängen analysiert werden. Durch die ungleiche Verteilung der Geschlechter auf die Bildungsgänge – der Anteil der Mädchen beträgt auf dem Gymnasium 57 Prozent und der Anteil der Jungen auf der Hauptschule 60,3 Prozent (Zimmer, Burba & Rost, 2004) – wird der Mittelwert für die Gesamtgruppe aller Mädchen erhöht. Dem ungleichen Mischungsverhältnis von Mädchen und Jungen in den Schulformen soll durch die Berücksichtigung der Bildungsgänge in den Regressionsmodellen Rechnung getragen werden. Darüber hinaus stellen sowohl

Schulformen als auch Einzelschulen innerhalb derselben Schulform institutionell vorgeformte differenzielle Entwicklungsmilieus dar. Schüler mit gleicher Begabung, gleichen Fachleistungen und gleicher Sozialschichtzugehörigkeit erhalten je nach Schulformzugehörigkeit und je nach besuchter Einzelschule unterschiedliche Entwicklungschancen (Baumert, Trautwein & Artelt, 2003). Es ist denkbar, dass verschiedene Schulumwelten auf unterschiedliche Weise auf geschlechtsspezifische Stärken und Schwächen einwirken. Deshalb werden in den folgenden Regressionsmodellen nicht nur die Bildungsgänge<sup>38</sup> – Haupt- und Realschule, integrierte Gesamtschule (IGS) und Gymnasium – berücksichtigt, sondern auch Interaktionsterme zwischen den Bildungsgängen und dem Geschlecht.

	<b>Mathematik</b>	<b>Biologie</b>	<b>Chemie</b>	<b>Physik</b>
Jungen	24.86	8.95	33.02	40.73
Realschule	91.68	91.35	89.36	76.48
Gymnasium	166.17	173.28	166.35	138.37
IGS	49.90	52.17	55.34	43.90
Realschule*Jungen	-2.30 <sup>ns</sup>	1.60 <sup>ns</sup>	-.18 <sup>ns</sup>	-4.86
Gymnasium*Jungen	3.93 <sup>ns</sup>	8.73	6.57	3.43 <sup>ns</sup>
IGS*Jungen	-3.47 <sup>ns</sup>	-2.62 <sup>ns</sup>	-4.95 <sup>ns</sup>	-6.48
Konstante	402.81	411.16	388.38	401.98
Fallzahl	17.436	31.887	31.887	31.887
r <sup>2</sup>	0.4462	0.4625	0.3855	0.4204

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 7-3: Regressionsmodell 2 (PISA-E-Studie)

Es zeigt sich in allen Fächern eine deutliche Zunahme der geschlechtsspezifischen Leistungsunterschiede. Die geringsten Unterschiede sind mit knapp neun Punkten in der Biologie zu beobachten, in der Mathematik steigt die Leistungsdifferenz zwischen Mädchen und Jungen bei Kontrolle der Bildungsgänge auf knapp 25 Punkte, in Chemie auf 33 Punkte und in Physik auf 40 Punkte.

Die Hauptschülerinnen erreichen in Mathematik durchschnittlich 402.8 Punkte, die Hauptschüler durchschnittlich 24.86 Punkte mehr. In der integrierten Gesamtschule (IGS) erzielen die Schülerinnen durchschnittlich 49.90 Punkte, in der Realschule 91.68 Punkte und im Gymnasium 166.17 Punkte mehr als die Schülerinnen der Hauptschule (Referenzkategorie). Ähnliche Zusammenhänge sind auch in den naturwissenschaftlichen Fächern auszumachen. In Biologie und Chemie zeigt sich zugunsten der Jungen, in Physik zugunsten der Mädchen ein Effekt der differenziellen Entwicklungsmilieus. Mit anderen Worten: Die geschlechtsspezifischen

<sup>38</sup> Dazu wurden die Bildungsgänge Haupt- und Realschule sowie Gymnasium und integrierte Gesamtschule (IGS) dummykodiert. Die Hauptschule dient bei den Analysen als Referenzkategorie. Da bei der PISA-Studie Schüler untersucht wurden, die zum Beginn des Testzeitraums 15 Jahre/drei Monate und 16 Jahre/zwei Monate alt waren, wurden die 211 Berufsschüler als Hauptschüler rekodiert. Die Sonderschüler (752 Personen) wurden aus den Analysen ausgeschlossen.

schen Leistungsunterschiede vergrößern sich im Gymnasium gegenüber der Hauptschule um 8.7 Punkte (Biologie) und 6.6 Punkte (Chemie), während es der Realschule und der integrierten Gesamtschule im Vergleich zur Hauptschule gelingt, die Leistungsunterschiede zwischen den Geschlechtern leicht zu verringern.

Von besonderem Interesse sind natürlich wieder die Zusammenhänge zwischen den motivationalen Merkmalen der Schüler und der erreichten Punktzahl bei der nationalen PISA-Erweiterungsstudie. Deshalb werden im nächsten Schritt die Selbstwirksamkeit, das akademische Selbstkonzept, welches sich allgemein auf die Schulleistungen bezieht, das fachspezifische Konzept und das fachspezifische Interesse in den Regressionsmodellen berücksichtigt. Bedauerlicherweise liegen diese motivationalen Merkmale für die naturwissenschaftlichen Leistungsbereiche nur von einem Bruchteil der Schüler vor, so dass sich die Fallzahlen an dieser Stelle deutlich reduzieren.

	<b>Mathematik</b>	<b>Biologie</b>	<b>Chemie</b>	<b>Physik</b>
Jungen	14.69	1.14 <sup>ns</sup>	33.00	30.40
Realschule	82.33	82.87	82.46	69.71
Gymnasium	153.92	163.26	147.15	125.18
IGS	42.09	56.86	36.16 <sup>ns</sup>	52.72
Realschule*Jungen	.34 <sup>ns</sup>	1.05 <sup>ns</sup>	-28.82 <sup>ns</sup>	-5.84 <sup>ns</sup>
Gymnasium*Jungen	7.53	14.73 <sup>ns</sup>	9.24 <sup>ns</sup>	20.33 <sup>ns</sup>
IGS*Jungen	3.75 <sup>ns</sup>	-17.15 <sup>ns</sup>	-9.79 <sup>ns</sup>	-23.11 <sup>ns</sup>
Selbstwirksamkeit	7.67	16.21	19.94	10.96
akademisches Konzept	12.33	21.13	30.41	19.20
fachspezifisches Konzept	21.60	17.91	11.38 <sup>ns</sup>	14.53
fachspezifisches Interesse	-6.46	-6.02 <sup>ns</sup>	-9.56 <sup>ns</sup>	-8.73
Konstante	325.93	295.45	274.98	327.34
Fallzahl	14.835	672	662	670
r <sup>2</sup>	0.4957	0.5091	0.3710	0.4758

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 7-4: Regressionsmodell 3 (PISA-E-Studie)

In Mathematik, Biologie und Physik ist bei Kontrolle der motivationalen Merkmale ein Rückgang der geschlechtsspezifischen Leistungsunterschiede auszumachen, in Chemie bleibt die Leistungsdifferenz zwischen Mädchen und Jungen mit 33 Punkten weitgehend stabil. Der Leistungsvorsprung für die Jungen beträgt in Mathematik 14.69 Punkte und in Physik 30.4 Punkte. In Biologie sind keine signifikanten Geschlechtsunterschiede auszumachen. Die erreichte Punktzahl bleibt erwartungsgemäß im starken Maße vom Bildungsgang abhängig. So erzielen Gymnasiasten eine höhere Punktzahl als Realschüler und Realschüler eine höhere Punktzahl als Schüler der integrierten Gesamtschule. Hauptschüler erzielen durchschnittlich schlechtere Leistungen als Schüler der integrierten Gesamtschule, der Realschule und des Gymnasiums.

Interessant bleibt der (allerdings nicht mehr signifikante) Einfluss der Bildungsgänge auf geschlechtsspezifische Unterschiede bei der erreichten Punktzahl im PISA-Test. Die geschlechtsspezifischen Leistungsunterschiede sind im Gymnasium deutlich stärker ausgeprägt als in der Hauptschule. In der Realschule und in der integrierten Gesamtschule sind die Leistungsunterschiede zwischen Mädchen und Jungen tendenziell schwächer als in der Hauptschule. Ein Beispiel: In der Hauptschule erzielten die Jungen in Chemie durchschnittlich 33 Punkte mehr als die Mädchen, im Gymnasium vergrößert sich der Vorsprung der Jungen gegenüber den Mädchen sogar um neun auf 42 Punkte. In der Realschule reduzieren sich die geschlechtsspezifischen Leistungsunterschiede von 33 Punkten um 28.82 auf vier Punkte, in der integrierten Gesamtschule immerhin um knapp zehn Punkte auf 23 Punkte. Innerhalb der Bildungsgänge gibt es demnach Prozesse, die die geschlechtsspezifischen Leistungsunterschiede verstärken oder abschwächen. Es ist interessant, dass sich im leistungsstärksten Bildungsgang Gymnasium und im leistungsschwächsten Bildungsgang Hauptschule stärkere Geschlechtsunterschiede zeigen als in den leistungsmittleren Bildungsgängen Realschule und IGS.

Bei den motivationalen Merkmalen zeigen sich positive Zusammenhänge zwischen der Selbstwirksamkeit, dem akademischen und dem fachspezifischen Selbstkonzept und der erreichten Punktzahl bei der PISA-Erweiterungsstudie. Beim fachspezifischen Interesse ist dagegen ein negativer Zusammenhang zu beobachten. Steigt das fachspezifische Interesse um einen Punkt, dann sinkt in Mathematik und Physik die durchschnittliche Leistung um sechs beziehungsweise acht Punkte, in Biologie und Chemie ist der Zusammenhang nicht signifikant. Bleibt der Zusammenhang bestehen, wenn für Lernstrategien kontrolliert wird?

	<b>Mathematik</b>	<b>Biologie</b>	<b>Chemie</b>	<b>Physik</b>
Jungen	9.21	1.73 <sup>ns</sup>	28.70 <sup>ns</sup>	28.98
Realschule	81.01	77.78	78.48	66.88
Gymnasium	147.51	158.70	141.28	119.73
IGS	38.29	50.55	26.39 <sup>ns</sup>	43.20
Realschule*Jungen	2.12 <sup>ns</sup>	-5.17 <sup>ns</sup>	-29.11 <sup>ns</sup>	-7.60 <sup>ns</sup>
Gymnasium*Jungen	10.46	11.51 <sup>ns</sup>	13.12 <sup>ns</sup>	16.28 <sup>ns</sup>
IGS*Jungen	7.54 <sup>ns</sup>	-16.05 <sup>ns</sup>	2.39 <sup>ns</sup>	-13.91 <sup>ns</sup>
Selbstwirksamkeit	6.73	17.27	15.61	9.04 <sup>ns</sup>
akademisches Konzept	13.00	21.09	32.22	19.99
fachspezifisches Konzept	21.31	21.25	9.14 <sup>ns</sup>	13.48
fachspezifisches Interesse	-6.82	-9.02 <sup>ns</sup>	-4.26 <sup>ns</sup>	-8.68 <sup>ns</sup>
Kontrolle	1.98 <sup>ns</sup>	1.46 <sup>ns</sup>	19.70	5.41 <sup>ns</sup>
Elaboration	6.66	2.56 <sup>ns</sup>	-1.84 <sup>ns</sup>	2.65 <sup>ns</sup>
Memorieren	-13.34	-9.23	-22.35	-14.22
Konstante	344.65	310.07	282.99	351.15
Fallzahl	14.085	633	621	632
r <sup>2</sup>	0.5047	0.5109	0.4049	0.4804

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 7-5: Regressionsmodell 4 (PISA-E-Studie)

Die Lernstrategien – Kontrolle, Elaboration und Memorieren – dienen in diesen Regressionsmodellen wieder als klassische Kontrollvariablen. Wie bei der internationalen PISA-Studie 2003 wurden die Schüler bei der nationalen PISA-Erweiterungsstudie 2000 über den Einsatz ihrer Lernstrategien befragt, die Repräsentation von Kenntnissen mit sich bringen, die unverarbeitet im Gedächtnis gespeichert werden (Memorieren), bei denen neue Informationen mit bereits Gelerntem verknüpft werden (Elaboration) und mit denen sie prüfen, was sie gelernt haben, und herauszufinden versuchen, was sie noch lernen müssen (Kontrolle). Es zeigt sich ein negativer Zusammenhang zwischen dem mechanischen Auswendiglernen und der erreichten Punktzahl in den einzelnen Bereichen. Die Elaborationsstrategien zeigen nur bei der erreichten Punktzahl im PISA-Mathematiktest einen signifikanten positiven Zusammenhang, in den naturwissenschaftlichen Bereichen ist der Zusammenhang jeweils nicht signifikant, in Chemie ist ein negativer (nicht signifikanter) Zusammenhang zu beobachten. Dagegen zeigt sich in Chemie ein stark positiver und signifikanter Zusammenhang zwischen den Kontrollstrategien und der erreichten Punktzahl, während die Zusammenhänge in Mathematik, Physik und Biologie jeweils nicht zufallskritisch abgesichert werden können.

Insgesamt hat sich das Bild auch nach Berücksichtigung der Kontrollvariablen nur geringfügig verändert. Die geschlechtsspezifischen Leistungsunterschiede haben sich minimal reduziert. In Mathematik ist der Vorsprung der Jungen von 14 auf neun Punkte gesunken, in Chemie und Physik beträgt der Leistungsvorsprung jeweils knapp 29 Punkte. Die Effekte der Bildungsgänge auf die geschlechtsspezifischen Leistungsunterschiede sind mit Ausnahme des Gymnasiums und der Mathematik nicht signifikant, was zumindest in den naturwissenschaftlichen Fächern auch auf die – im Vergleich zur Mathematik – geringeren Fallzahlen zurückzuführen ist. Im Gymnasium vergrößert sich der geschlechtsspezifische Leistungsvorteil in der Mathematik um rund zehn Punkte und beträgt insgesamt 19 Punkte, in der Hauptschule, in der Realschule und in der integrierten Gesamtschule bleibt es bei neun Punkten.

Bei der Selbstwirksamkeit, dem akademischen und dem fachspezifischen Selbstkonzept ist weiterhin ein positiver Zusammenhang mit der erreichten Punktzahl auszumachen. Je höher die Selbstwirksamkeit, je höher das akademische und je höher das fachspezifische Selbstkonzept, desto höher die erreichte Punktzahl bei den jeweiligen PISA-Tests. In Physik ist der Zusammenhang mit der Selbstwirksamkeit, in Chemie das fachspezifische Selbstkonzept jedoch nicht signifikant. Die Zusammenhänge zwischen dem fachspezifischen Interesse und der erzielten Punktzahl bleibt in allen Fächern negativ, in den naturwissenschaftlichen Fächern ist dieser Zusammenhang jedoch nicht signifikant.

Bei den Analysen mit der nationalen PISA-Erweiterungsstudie spielt das fachspezifische Interesse eine besondere Rolle. Der Stereotype-Threat-Effekt sollte sich insbesondere bei den Personen zeigen, die sich mit dem Bereich identifizieren, aber mit negativen stereotypen Erwartungen konfrontiert sind. Ausgehend von der Argumentation, dass Biologie eher als weibliche, Chemie und Physik eher als männliche Domänen angesehen werden, sind je nach Geschlecht unterschiedliche Zusammenhänge zwischen dem fachspezifischen Interesse (Indikator für Domainidentifikation) und dem jeweiligen Leistungsbereich zu erwarten. Für die Mädchen werden in Chemie und Physik negative Zusammenhänge zwischen dem Interesse und der erzielten Punktzahl, in Biologie jedoch ein positiver Zusammenhang zwischen Interesse und Leistung erwartet. Dagegen sollte sich bei den Jungen in Biologie ein negativer Zusammenhang zwischen dem fachspezifischen Interesse und der erzielten Punktzahl zeigen. In Mathematik wird – analog zur internationalen PISA-Studie 2003 – für die Mädchen ein negativer Zusammenhang zwischen dem mathematischen Interesse und der Leistung erwartet. Um diese Annahmen zu überprüfen, werden die nächsten Regressionsmodelle um die Interaktionsterme zwischen dem Geschlecht und den einzelnen Variablen erweitert.

	<b>Mathematik</b>	<b>Biologie</b>	<b>Chemie</b>	<b>Physik</b>
Jungen	6.31 <sup>ns</sup>	-38.59 <sup>ns</sup>	66.07 <sup>ns</sup>	35.09 <sup>ns</sup>
Realschule	80.17	81.03	81.23	68.19
Gymnasium	146.96	161.05	144.65	121.92
IGS	37.93	52.78	31.32 <sup>ns</sup>	46.50
Realschule*Jungen	3.23 <sup>ns</sup>	-10.67 <sup>ns</sup>	-34.00 <sup>ns</sup>	-9.45 <sup>ns</sup>
Gymnasium*Jungen	10.42	4.53 <sup>ns</sup>	10.40 <sup>ns</sup>	9.71 <sup>ns</sup>
IGS*Jungen	7.94 <sup>ns</sup>	-20.94 <sup>ns</sup>	-1.31 <sup>ns</sup>	-17.91 <sup>ns</sup>
Selbstwirksamkeit	2.10 <sup>ns</sup>	11.05 <sup>ns</sup>	4.47 <sup>ns</sup>	4.77 <sup>ns</sup>
Selbstwirksamkeit*Jungen	10.05	12.90 <sup>ns</sup>	19.37 <sup>ns</sup>	8.31 <sup>ns</sup>
akademisches Konzept	14.61	22.21	49.27	26.87
akademisches Konzept*Jungen	-3.67 <sup>ns</sup>	-1.51 <sup>ns</sup>	-29.80	-15.42 <sup>ns</sup>
fachspezifisches Konzept	23.87	10.03 <sup>ns</sup>	-5.60 <sup>ns</sup>	8.06 <sup>ns</sup>
fachspezifisches Konzept*Jungen	-5.12	23.69	25.13 <sup>ns</sup>	10.32 <sup>ns</sup>
fachspezifisches Interesse	-10.62	1.72 <sup>ns</sup>	-5.35 <sup>ns</sup>	-1.93 <sup>ns</sup>
fachspezifisches Interesse*Jungen	7.96	-23.48	3.84 <sup>ns</sup>	-13.36 <sup>ns</sup>
Kontrolle	1.70 <sup>ns</sup>	-4.79 <sup>ns</sup>	24.86	8.81 <sup>ns</sup>
Kontrolle*Jungen	.07 <sup>ns</sup>	11.63 <sup>ns</sup>	-8.33 <sup>ns</sup>	-3.03 <sup>ns</sup>
Elaboration	8.48	3.92 <sup>ns</sup>	7.71 <sup>ns</sup>	-5.58 <sup>ns</sup>
Elaboration*Jungen	-3.67 <sup>ns</sup>	-1.71 <sup>ns</sup>	-18.70 <sup>ns</sup>	16.27 <sup>ns</sup>
Memorieren	-11.32	-8.37 <sup>ns</sup>	-21.59	-14.64
Memorieren*Jungen	-4.23	-6.19 <sup>ns</sup>	.30 <sup>ns</sup>	-2.78 <sup>ns</sup>
Konstante	345.98	333.85	254.97	351.05
Fallzahl	14.085	633	621	632
r <sup>2</sup>	0.5060	0.5195	0.4205	0.4866

Anmerkung: ns = nicht signifikant bei einer Irrtumswahrscheinlichkeit von  $\alpha = 0.05$ .

Tabelle 7-6: Regressionsmodell 5 (PISA-E-Studie)

Die Ergebnisse sind bedeutsam und überraschend zugleich: In Mathematik kann der mit der internationalen PISA-Studie 2003 gezeigte geschlechtsspezifische Zusammenhang zwischen dem Mathematikinteresse und der erzielten Punktzahl repliziert werden. Auch in der PISA-Erweiterungsstudie 2000 zeigt sich ein negativer Zusammenhang zwischen dem Mathematikinteresse und der Leistung, der weit gehend auf die Mädchen begrenzt ist. Dieser Befund stützt den Verdacht von Davies und Spencer (2005), wonach die geschlechtsspezifischen Leistungsunterschiede in der Mathematik in erster Linie ein Resultat von Stereotype Threat darstellen.

Doch nicht nur in der Mathematik finden sich Hinweise, die mit der Stereotype-Threat-Theorie vereinbar sind. Insbesondere der geschlechtsspezifische Zusammenhang des Interesses und der Testleistung in Biologie ist aufschlussreich. Ausgehend von der Argumentation, dass in Biologie eher die Jungen mit negativen stereotypen Erwartungen konfrontiert sind, zeigt sich bei den Schülern ein negativer Zusammenhang zwischen dem Interesse und der erreichten Punktzahl bei dem im Rahmen der PISA-Erweiterungsstudie durchgeführten Biologietest. Entsprechend der theoretischen Annahmen ist Stereotype Threat nicht auf Frauen und Mathematik begrenzt, sondern bezieht sich auf alle Situationen, in denen negative stereotype Erwartungen auf das Verhalten anwendbar sind. Mit anderen Worten: Sind die Mädchen in Mathematik mit negativen stereotypen Erwartungen konfrontiert, so müssen sich die Jungen in Biologie mit negativen stereotypen Erwartungen auseinandersetzen. In beiden Fällen kann Stereotype Threat zu einer Leistungsverringerung führen.

In Chemie und Physik zeigen sich keine signifikanten negativen Zusammenhänge zwischen dem fachspezifischem Interesse und der erzielten Punktzahl. Während aber das Ergebnismuster in Chemie mit den theoretischen Erwartungen übereinstimmt – tendenziell ein negativer Zusammenhang zwischen dem Interesse und der Testleistung bei den Mädchen –, zeigt sich in Physik das gegenläufige Bild. Da Physik – analog zu Mathematik und Chemie – als männliche Domäne gilt, ist der insbesondere bei den Jungen (nicht signifikante) negative Zusammenhang zwischen dem Interesse und der erzielten Punktzahl überraschend. An dieser Stelle kann dieser Befund nicht aufgeschlüsselt werden. Möglicherweise wird Physik von den untersuchten Schülern nicht als männliche Domäne angesehen. Da bei PISA die wahrgenommene Stereotypisierung nicht erfasst wurde, bleibt es künftiger Forschung vorbehalten, den Zusammenhang genauer zu untersuchen.

Die Analysen in diesem Kapitel sollten die bisherigen Befunde zu geschlechtsspezifischen Leistungsunterschieden in der Mathematik durch eine Erweiterung der Perspektive auf Leistungsunterschiede zwischen Mädchen und Jungen in den naturwissenschaftlichen Bereichen ergänzen. Die Ergebnisse können dabei weit gehend im Sinne der Stereotype-Threat-Theorie interpretiert werden. Das folgende Kapitel fasst die zentralen Ergebnisse zusammen und bewertet sie vor den im theoretischen Teil berichteten Erklärungsansätzen.

## **8. Biologie, Sozialisation oder Situation?**

Geschlechterunterschiede in der mathematischen Leistung werden seit Jahrzehnten untersucht und haben unzählige Debatten ausgelöst. Fakt ist: Es gibt auch heute noch in zahlreichen Ländern einen statistisch signifikanten Unterschied in den mathematischen Testleistungen zwischen Mädchen und Jungen beziehungsweise Frauen und Männern. Im Allgemeinen zeigen sich ab einem Alter von 15 Jahren Leistungsvorteile zugunsten des männlichen Geschlechts. Die Schlüsselfrage besteht nun darin, warum das so ist. Sind diese geschlechtsspezifischen Leistungsunterschiede das Resultat von stabilen Faktoren biologischer und/oder psychosozialer Art oder das Ergebnis von situativen Merkmalen der Testsituation?

In experimentellen Untersuchungen konnte bereits mehrfach gezeigt werden, dass sich die geschlechtsspezifischen Leistungsunterschiede (nicht nur) in der Mathematik auf situative Merkmale der Testsituation zurückführen lassen und auf einem Phänomen beruhen, das als *Stereotype Threat* bezeichnet wird. Eine weit gehend offene Frage war es jedoch, ob die Ergebnisse der experimentellen Untersuchungen auch auf das tatsächliche Leben generalisiert werden können. Die empirischen Analysen in dieser Arbeit zielten deshalb darauf ab, zu prüfen, ob sich auch bei einer repräsentativen Stichprobe unter normalen Testbedingungen Indizien finden lassen, die mit den Annahmen der *Stereotype-Threat-Theorie* und der theoretischen Weiterentwicklung – das *MERF-Modell* – vereinbar sind. Die empirischen Ergebnisse lassen dabei den Schluss zu, dass das Phänomen *Stereotype Threat* nicht auf Laborbedingungen begrenzt ist.

Der in der ersten Hypothese für die Mädchen postulierte negative Zusammenhang zwischen dem Mathematikinteresse und der erreichten Punktzahl im PISA-Mathematiktest zeigt sich in zehn der 13 ausgewählten Länder. In zwei weiteren Ländern (Niederlande und Luxemburg) entsprechen die Zusammenhänge den theoretischen Erwartungen, sind aber nicht signifikant. Die Ergebnisse sind deshalb besonders bedeutsam, da bei dem PISA-Mathematiktest sowohl schwierige als auch leichte Aufgaben verwendet wurden. Es ist davon auszugehen, dass die durch *Stereotype Threat* ausgelösten Mechanismen nicht erst dann die Leistungsfähigkeit beeinflussen, wenn ausschließlich schwierige Aufgaben verwendet werden, sondern bereits bei einer Aufgabenzusammenstellung von schwierigen und leichten Aufgaben. Insofern stellten die Analysen mit dem PISA-Datensatz einen besonders strengen Test der *Stereotype-Threat-Theorie* dar, andererseits unterstreichen die Befunde die Bedeutung situativer Merkmale der Testsituation im Unterrichtsalltag.

Das *MERF-Modell* – die Einbettung der *Stereotype-Threat-Theorie* in einen größeren theoretischen Kontext – postuliert eine Beziehung zwischen der Art der Erwartung (positiv oder negativ) und der Zugänglichkeit zu einem regulatorischen Fokus (*Prevention* oder *Promotion*)

auf die Leistungsfähigkeit. Nach dem MERF-Modell führen eine negative Erwartung und die Zugänglichkeit zu einem Prevention Focus oder eine positive Erwartung und ein aktivierter Promotion Focus zu einem Leistungsabfall, der bei den Kombinationen positive Erwartung/Prevention Focus und negative Erwartung/Promotion Focus nicht zu beobachten ist.

In den Analysen diente die instrumentelle Motivation als Indikator für einen chronischen Prevention Focus, während das mathematische Selbstkonzept als Indikator für einen chronischen Promotion Focus herangezogen wurde. Ferner wurde angenommen, dass die Mädchen beim PISA-Test tendenziell mit einer negativen Erwartung, die Jungen mit einer positiven Erwartung konfrontiert waren. Für die Mädchen wurde deshalb ein negativer Zusammenhang zwischen der instrumentellen Motivation und der Leistung erwartet (Hypothese 2), während sich bei den Jungen ein negativer Zusammenhang zwischen dem mathematischen Selbstkonzept und der erreichten Punktzahl im PISA-Mathematiktest zeigen sollte (Hypothese 3).

In fünf der 13 Länder zeigten sich signifikante negative Zusammenhänge zwischen der instrumentellen Motivation und der Testleistung, die weitgehend auf die Mädchen beschränkt waren. In sieben Ländern waren diese Zusammenhänge nicht signifikant, in Norwegen war der Zusammenhang positiv. Bezüglich der dritten Hypothese zeigten sich für die Mädchen durchgehend positive Zusammenhänge zwischen dem mathematischen Selbstkonzept und der erzielten Punktzahl. Bei den Jungen waren diese Zusammenhänge tendenziell schwächer ausgeprägt; in der Schweiz, in Deutschland, in den Niederlanden und in Luxemburg waren die Zusammenhänge negativ. Insgesamt sind die empirischen Ergebnisse weitgehend mit dem MERF-Modell kompatibel.

Der PISA-Index Mathematikangst erfasst die emotionale Angespanntheit der Schüler beim Bearbeiten von Mathematikaufgaben. Je wichtiger eine Person die mathematische Fähigkeit einschätzt – erfasst über die Skala der instrumentellen Motivation –, desto stärker war die Person beim Lösen der Aufgaben emotional angespannt. Die vierte Hypothese postulierte einen negativen Zusammenhang zwischen instrumenteller Motivation/Mathematikangst und Testleistung. Solche Zusammenhänge zeigten sich in Luxemburg, Belgien, den Niederlanden, Spanien, Deutschland, Österreich, Dänemark und Norwegen. In der Schweiz zeigte sich der Zusammenhang nur für die Mädchen, in Italien nur für die Jungen.

Zwar sind in einigen Ländern keine signifikanten Zusammenhänge zwischen den regulatorischen Fokussen und der Punktzahl im PISA-Test nachzuweisen, dies ist aber möglicherweise auf die konzeptionell problematische Erfassung des Promotion Focus sowie des Prevention Focus und/oder das unterschiedliche Ausmaß der geschlechtsspezifischen Erwartungen in den einzelnen Ländern zurückzuführen. An dieser Stelle sind die Grenzen der internationalen PISA-Studie offensichtlich, bei der weder die wahrgenommene Stereotypisierung von Ma-

thematik als männliche Domäne noch die chronische Disposition zu einem regulatorischen Fokus konzeptionell zufrieden stellend erfasst wurde. Trotz dieser Mängel erlauben die empirischen Befunde eine Interpretation im Sinne der Stereotype-Threat-Theorie und des MERF-Modells.

Gegen eine biologische Verursachung der geschlechtsspezifischen Leistungsdifferenzen in der Mathematik sprechen mindestens zwei Ergebnisse. Einerseits zeigen sich zwischen den Ländern beachtliche Unterschiede in dem Ausmaß der Leistungsdifferenzen zwischen Mädchen und Jungen, andererseits sind diese Differenzen durch Unterschiede in den motivationalen Merkmalen erklärbar. Konkret zeigten die Mediationsanalysen, dass die Unterschiede in den geschlechtsspezifischen Leistungsdifferenzen zwischen der Schweiz und Österreich sowie Dänemark und Norwegen durch Unterschiede der Selbstwirksamkeit und der instrumentellen Motivation erklärt werden können. Eine partielle Mediation konnte bei der Analyse der Länderunterschiede zwischen Italien und Spanien nachgewiesen werden.

Die Länderunterschiede in den motivationalen Schülermerkmalen sind in erster Linie ein Resultat des allgemeinen und kulturellen Kontexts. Den Staaten gelingt es offenbar unterschiedlich gut, für Mädchen und Jungen gleichsam förderliche Lernbedingungen zu schaffen. Dabei müssen hohe Leistung und geringe Geschlechtsunterschiede – wie die Beispiele Niederlande und Finnland zeigen – kein Widerspruch sein. Möglicherweise sind die Entwicklungsvoraussetzungen – insbesondere im Bereich der räumlich-visuellen Wahrnehmung – für Mädchen und Jungen in der Tat bis zu einem gewissen Grad unterschiedlich gelagert. Die Varianz bei den geschlechtsspezifischen Leistungsunterschieden impliziert jedoch, dass die bestehenden Unterschiede nicht als unvermeidlich hingenommen werden müssen und dass mit Hilfe effektiver Maßnahmen ein Zustand überwunden werden kann, der lange Zeit als unabdingbare Konsequenz grundlegender Begabungsunterschiede zwischen Jungen und Mädchen betrachtet wurde.

Die Ergebnisse der Analysen der naturwissenschaftlichen Leistungsunterschiede zwischen Mädchen und Jungen können dahin gehend interpretiert werden, dass das Phänomen der Bedrohung durch Stereotype nicht auf die Mathematik beschränkt ist, sondern sich ebenso auf andere Leistungsfelder bezieht, in denen es stereotype Leistungserwartungen gibt. Besonders interessant ist dabei ein geschlechtsspezifischer Einfluss des differenziellen Entwicklungsmilieus auf die Leistung. Möglicherweise ist die geschlechtsspezifische Leistungserwartung je nach Bildungsgang unterschiedlich stark ausgeprägt. Künftiger Forschung bleibt es vorbehalten, den Einfluss differenzieller Entwicklungsmilieus auf die Leistung insbesondere vor dem Hintergrund stereotyper Erwartungen genauer zu untersuchen.

Fazit: Innerhalb der untersuchten Staaten finden sich Hinweise, die eine Erklärung der geschlechtsspezifischen Leistungsunterschiede durch Stereotype Threat und das MERF-Modell nahe legen. Das Ausmaß der geschlechtsspezifischen Differenzen zwischen den Staaten scheint primär auf Unterschiede in den motivationalen Merkmalen der Schüler zurückzuführen zu sein. Soziale und situative Erklärungsansätze schließen sich an dieser Stelle nicht aus, vielmehr stellen die situationalen Erklärungsmodelle eine Untergruppe der sozialen Erklärungsansätze dar. Nur durch die Stereotypisierung der Mathematik als männliche Domäne ist ein Stereotype-Threat-Effekt überhaupt denkbar. Die Stereotype-Threat-Theorie beziehungsweise das MERF-Modell spezifiziert die Randbedingungen, wann sich die stereotype Bedrohung in einer niedrigeren Leistung niederschlägt. Die Antwort auf die Frage, wie die Leistungsunterschiede zwischen Jungen und Mädchen in der Mathematik erklärt werden können, lautet also nicht Sozialisation oder Situation, sondern vielmehr Sozialisation *und* Situation.

Die Leistungen sowie die Motivation und die Einstellungen gegenüber Mathematik können den künftigen Bildungs- und Berufsweg erheblich beeinflussen, was sich wiederum auf die Karriere- und Verdienstaussichten auswirken kann. Stereotype Threat führt unter bestimmten Randbedingungen dabei nicht nur zu einer geringeren Leistungsfähigkeit, sondern führt langfristig dazu, dass die betroffenen Personen den jeweiligen Leistungsbereich von ihrem Selbstwertgefühl entkoppeln, um den erhöhten Druck zu vermeiden, was wiederum die Motivation reduziert, in diesem Bereich erfolgreich zu sein. Wie kann den negativen Konsequenzen von Stereotype Threat begegnet werden? Dieser Frage widmet sich das anschließende Kapitel.

## **9. Geschlechtsspezifische Leistungsunterschiede: Handlungsempfehlungen**

In den vorangegangenen Abschnitten wurde untersucht, welche Leistungsunterschiede zwischen Jungen und Mädchen in der Mathematik und verwandten Bereichen bestehen. Obwohl Frauen große Fortschritte bei der Verringerung ihres historischen Bildungsnachteils erzielt und das männliche Geschlecht in vielen Bereichen überrundet haben, bestehen in der Mathematik und verwandten Bereichen weiterhin Leistungsvorteile zugunsten der Jungen. Die Ergebnisse dieser Arbeit stützen die These, wonach die geschlechtsspezifischen Leistungsunterschiede in der Mathematik primär auf Stereotype Threat zurückzuführen sind. In der Bildungsforschung werden zahlreiche Ansatzpunkte diskutiert und umgesetzt, um die Genderlücke in Mathematik zu verringern (Beerman et al., 1992; Halpern, 2000; Srocke, 1989), die jedoch situationale Merkmale der Testsituation weit gehend ausblenden. Wie kann den geschlechtsspezifischen Leistungsunterschieden vor dem Hintergrund der Stereotype-Threat-Theorie und dem MERF-Modell begegnet werden?

Um die geschlechtsspezifischen Leistungsdifferenzen zu verringern, empfiehlt die OECD, das Mathematikinteresse von Mädchen zu fördern (OECD, 2004a). Zweifellos besitzt das Interesse innerhalb der Gruppe motivationaler Schülermerkmale eine herausragende Bedeutung (Schiefele, Krapp & Schreyer, 1993), da es als Motor für Lernprozesse die Auswahl von Lerngegenständen und die Intensität der Auseinandersetzung mit diesen Gegenständen beeinflussen kann. Die OECD-Empfehlung erscheint angesichts der berichteten Ergebnisse jedoch problematisch, da insbesondere Schülerinnen, die sich durch ein hohes mathematisches Interesse auszeichnen, von Stereotype Threat bedroht sind. Isolierte Maßnahmen, die lediglich darauf abzielen, das Interesse der Mädchen an der Mathematik zu fördern, sind deshalb als kontraproduktiv zu bezeichnen.

Vielmehr muss in der pädagogischen Praxis auf situative Merkmale geachtet werden, die hinsichtlich der Anwendbarkeit von stereotypen Erwartungen von Bedeutung sind. Die gesellschaftliche Stereotypisierung von Mathematik als männliche Domäne schafft für die Mädchen schlechtere Lernvoraussetzungen und erschwert eine vorurteilsfreie Berufs- und Studienwahl. Die Institution Schule ist sicherlich nicht alleinige Verursacherin einer Stereotypisierung von Mathematik als männliche Domäne, sie scheint dieser Stereotypisierung aber auch nicht entgegenzuwirken. Damit Schüler die Stereotypisierung von Schulfächern abbauen können, sollte diese im Fachunterricht thematisiert und hinterfragt werden. Lehrer müssen ihre eigene Stereotypisierung von Unterrichtsfächern erkennen und gegebenenfalls abbauen. Schülern gilt es schließlich zu vermitteln, dass Stereotype nicht als korrekte Wirklichkeitsbeschreibungen zu verstehen sind und Beurteilungen sich stets auf das individuelle Leistungsvermögen der Person und nicht auf die Zugehörigkeit zu einer sozialen Kategorie beziehen.

In einem Überblicksartikel nennt Steele (1997) weitere Handlungsempfehlungen, um die negativen Konsequenzen von Stereotype Threat zu reduzieren. Neben einer optimistischen Schüler-Lehrer-Beziehung, bei der das Feedback mit einer optimistischen Einschätzung bezüglich des Potenzials des Schülers gekoppelt wird, sollen herausfordernde Aufgaben verwendet werden, die die Personen nicht überfordern. Allgemein sollte der Unterricht Möglichkeiten bieten, dass die Schüler das durch das Stereotyp vermeintliche Limit überwinden können und Selbstwirksamkeit aufgebaut wird.

In diesem Zusammenhang wird häufig vorgeschlagen, in den mathematisch-naturwissenschaftlichen Fächern Identifikationsmöglichkeiten für beide Geschlechter zu schaffen. Positive Rollenmodelle sollten dabei nicht nur in Schulbüchern erscheinen, sondern auch ein lebendiger Bestandteil des Unterrichts sein, indem Frauen bei Unterrichtsbesuchen oder Gastvorträgen ihre Erfahrungen und Erfolge in männerdominierten Berufsfeldern mitteilen. In experimentellen Untersuchungen konnten Marx und Roman (2002) sowie McIntry et al. (2003)

durch den Einsatz von Rollenmodellen die leistungsreduzierende Wirkung von Stereotype Threat aufheben. Bei pädagogischen Interventionen mit Rollenmodellen sind jedoch mindestens zwei Aspekte zu berücksichtigen: Einerseits müssen sich die Schüler mit dem Rollenmodell identifizieren können, das heißt, sie müssen es – zumindest im begrenzten Maße – als ähnlich wahrnehmen, andererseits müssen die Schüler die Erfolge des Rollenmodells als für sich selbst erreichbare Ziele erkennen. Ansonsten ist mit Kontrasteffekten zu rechnen, da das Rollenmodell als Ausnahme der Regel angesehen und die potenzielle Bedrohung durch Stereotype Threat eher noch verstärkt wird (Keller, 2004).

Als Handlungsoption gegen Stereotype Threat wird in der Literatur vielfach auf ein Interventionsprogramm an der Universität von Michigan verwiesen (Davies & Spencer, 2005; Steele, 1997; Steele et al., 2002). Primäre Zielgruppe des Programms waren Afroamerikaner, die in den Vereinigten Staaten mit dem Stereotyp der intellektuellen Minderwertigkeit konfrontiert sind. Die Intervention zielte darauf ab, eine stereotypfreie Umwelt zu schaffen, in der die betroffenen Personen ihr Leistungspotenzial komplett ausschöpfen können. Neben den regulären Kursen waren die Studierenden eingeladen, an zusätzlichen Seminaren, Diskussionsgruppen und Workshops teilzunehmen, bei denen der Glaube an die eigenen Stärken gefördert wurde. Afroamerikaner, die an dem Programm teilgenommen hatten, zeigten bessere Leistungen als die Afroamerikaner in der Kontrollgruppe. Die Ergebnisse der Längsschnittstudie lassen den Schluss zu, dass durch die Teilnahme an dem Programm die wahrgenommene Stereotypisierung reduziert werden konnte.

Gegenüber dem Interventionsprogramm an der Universität von Michigan bleibt allerdings kritisch anzumerken, dass den negativen Konsequenzen von Stereotype Threat nicht erst im universitären Kontext begegnet werden sollte. Im Alter von 15 Jahren stehen Schüler an der wichtigen Schwelle des Übergangs von der Schule ins Arbeitsleben oder zu einer weiterführenden Bildung. Sowohl die Leistung als auch die Motivation und die Einstellung gegenüber Mathematik können den Bildungs- und Berufsweg erheblich beeinflussen, weshalb entsprechende Maßnahmen bei den Jugendlichen, ja sogar schon bei Kindern ansetzen sollten.

Die Überlegungen des Interventionsprogramms können jedoch grundsätzlich auch auf geschlechtsspezifische Leistungsunterschiede in der Mathematik übertragen werden. Sommercamps und Förderprogramme könnten mathematisch interessierte Mädchen zusammenführen, um ihnen die Möglichkeit zu geben, Inhalte aus der Mathematik oder den Naturwissenschaften in einer fördernden Atmosphäre zu lernen, ohne dabei mit den klassischen Stereotypen konfrontiert zu werden. Förderprogramme und Sommercamps sind jedoch nicht als Vorstufe zu einer Trennung der Schulklassen nach Geschlecht beziehungsweise einer Abkehr von der Koedukation zu verstehen. Aus historischer Perspektive war die organisatorische Abgrenzung

von Mädchen- zum Jungenunterricht mit einer qualitativen Unterscheidung des Unterrichts verbunden, wobei Mädchenunterricht vielfach als minderwertig angesehen wurde. Genau in diese Gefahr begibt man sich, wenn man die mathematisch-naturwissenschaftlichen Fächer getrennt für Mädchen und Jungen anbieten will. Derartiger Unterricht kann schnell als ein Unterricht mit geringeren Anforderungen – eben „nur für Mädchen“ – abqualifiziert werden und erzielt auf diese Weise genau das Gegenteil des gewünschten Effekts (Srocke, 1989).

Die bisher geschilderten Handlungsempfehlungen haben sich in erster Linie mit stereotypen Erwartungen auseinander gesetzt und zielten darauf ab, die Stereotypisierung von Mathematik als männliche Domäne zu reduzieren. Eine stereotypfreie Umwelt lässt sich jedoch nicht per Dienstvorschrift verordnen, sondern muss von den Beteiligten gewollt werden. Zudem entlasten Stereotype das kognitive System, indem sie die Komplexität der Welt reduzieren. Insofern erscheint es ein unrealistisches Ziel, stereotype Erwartungen in einer Gesellschaft völlig zu eliminieren. Die geschilderten Handlungsempfehlungen können den negativen Einfluss von stereotypen Erwartungen auf die Leistungsfähigkeit aber zumindest reduzieren.

Wenn sich die stereotypen Erwartungen als äußerst resistent gegenüber Veränderungen erweisen, so legen die berichteten Befunde nahe, die Zugänglichkeit zu einem regulatorischen Fokus bei der Diskussion der Handlungsempfehlungen zu berücksichtigen. Nach dem MERF-Modell sind zwei Kombinationen zwischen stereotyper Erwartung und Zugänglichkeit zu einem regulatorischen Fokus zu unterscheiden, die zu einer geringeren Leistung führen: Negative Erwartung und Prevention Focus sowie positive Erwartung und Promotion Focus. Nach Higgins hat jeder Mensch eine Prädisposition für einen der beiden Fokusse, die Zugänglichkeit kann aber auch von Situation zu Situation variieren (Higgins, 1998). Bei Personen, die mit negativen Erwartungen konfrontiert sind, sollte die Zugänglichkeit zu einem Promotion Focus gestärkt werden, Personen, die sich mit positiven Erwartungen auseinandersetzen müssen, sollten angeregt werden, sich mit eigenen Verantwortlichkeiten und Pflichten zu beschäftigen, die eine Zugänglichkeit zu einem Prevention Focus nahe legen.

Diese allgemeinen Gedanken lassen sich wieder auf geschlechtsspezifische Leistungsunterschiede in der Mathematik übertragen. So könnte bei Mädchen, die in den mathematisch-naturwissenschaftlichen Bereichen mit negativen Erwartungen konfrontiert sind, ein Promotion Focus aktiviert werden. Keller (2004) empfiehlt ein Promotiontraining. Was kann man sich unter einem solchen Training vorstellen?

„Such a promotion training program could consist of several procedures designed to strengthen the input factors related to the promotion focus, such as personal ideals, nurturance needs, and possible gains. For example, the teacher could try to strengthen the degree to which female students (a) perceive reaching competence in math as a

personal ideal (rather than an ought), (b) see learning math and training their math ability as a chance to satisfy nurturance needs, and (c) perceive reaching competence in math as a personal gain.” (Keller, 2004, S. 120)

Vor einer Mathematikarbeit könnten Mädchen beispielsweise angeregt werden, über ihre Ideale und Hoffnungen nachzudenken. Der dadurch aktivierte Promotion Focus würde eine bessere Leistung ermöglichen. An die Jungen, die in Mathematik eher mit positiven Erwartungen konfrontiert sind, sollten die Lehrer eher Worte richten, die eine Auseinandersetzung über Verantwortlichkeiten und Pflichten implizieren. Die chronische Disposition zu einem regulatorischen Fokus oder auch die wahrgenommene Erwartung könnten die Lehrer zu Beginn eines Schuljahres per Fragebogen erfassen, um die individuelle Förderung jedes einzelnen Schülers zu optimieren.

Vor dem Hintergrund der theoretischen Argumentation und der empirischen Ergebnisse bieten die an dieser Stelle skizzierten Hinweise für die pädagogische Praxis eine Fülle an Handlungsoptionen, um den geschlechtsspezifischen Leistungsunterschieden nicht nur in der Mathematik zu begegnen. Gesellschaftliche Aufgabe sollte es sein, allen Schülern – unabhängig von Geschlecht und Rasse – die Möglichkeit zu bieten, ihr Leistungspotenzial vollkommen entfalten zu können. Dieses Ziel kann jedoch nur erreicht werden, wenn durch eine Kombination der dargestellten Vorschläge auf verschiedene Seiten der Problematik Einfluss genommen wird.

## **10. Resümee**

Ziel der Arbeit war zu prüfen, ob sich der Stereotype-Threat-Effekt auch in repräsentativen Stichproben unter normalen Testbedingungen nachweisen lässt. Die Ergebnisse der Analysen sind mit der Stereotype-Threat-Theorie kompatibel. Zudem können die Ergebnisse dahin gehend interpretiert werden, dass je nach regulatorischem Fokus Leistungserwartungen einen unterschiedlichen Einfluss auf die Testleistung haben. Die Indizienkette erlaubt die Schlussfolgerung, wonach die geschlechtsspezifischen Leistungsunterschiede in der Mathematik weniger das Resultat von stabilen Faktoren biologischer und/oder psychosozialer Art sind, sondern vielmehr das Ergebnis von situativen Merkmalen der Testsituation. Dabei stellen die situativen Erklärungsansätze jedoch eine Untergruppe der sozialen Erklärungsansätze dar.

An dieser Stelle muss jedoch auch auf die Grenzen der empirischen Analysen eingegangen werden. Weder die stereotype Bedrohung noch die Zugänglichkeit zu einem regulatorischen Fokus wurde bei der PISA-Studie experimentell variiert. Zudem wurden theoretisch wichtige Konstrukte, beispielsweise die wahrgenommene Stereotypisierung von Mathematik als männ-

liche Domäne, bei PISA entweder gar nicht oder nur sehr begrenzt erfasst. Letzteres gilt insbesondere für die chronische Zugänglichkeit zu einem regulatorischen Fokus. PISA ist zudem als eine Querschnittsstudie angelegt, bei der die Wirkungsursachen nicht eindeutig bestimmt werden können.

Die Grenzen des internationalen PISA-Datensatzes bieten Ansatzpunkte für weitere Forschung. Bei künftigen Schulleistungstudien wäre es zweifelsohne wünschenswert, den sicherlich ohnehin schon umfangreichen Schülerfragebogen um weitere Fragen zu ergänzen, um die theoretisch relevanten Konstrukte besser erfassen zu können. Dies kann in einem ersten Schritt über Optionen für nationale Ergänzungen und Erweiterungen relativ kostenneutral umgesetzt werden. Um Fehlinterpretationen des Wechselspiels von Schülermerkmalen und Kompetenzerwerb vorzubeugen, bieten sich experimentelle Designs in ausgewählten Schulklassen und Längsschnittanalysen an. Mit Längsschnittanalysen könnte zudem der von der Stereotype-Threat-Theorie postulierte Prozess der Deidentifikation besser untersucht werden. Im Mittelpunkt dieser Arbeit standen die geschlechtsspezifischen Leistungsdifferenzen in der Mathematik. Das Phänomen Stereotype Threat ist aber weder auf das Geschlecht noch auf die Mathematik beschränkt. Bei den Analysen der naturwissenschaftlichen Fächer konnte ein ähnliches Ergebnismuster gefunden werden. Künftige Forschung sollte sich der Frage widmen, inwieweit die mit Stereotype Threat verbundenen Prozesse auch für geschlechtsspezifische Leistungsunterschiede im verbalen Bereich verantwortlich sind.

Die Analysen der geschlechtsspezifischen Leistungsdifferenzen in dieser Arbeit zielten darauf ab, Unterschiede zwischen Mädchen und Jungen beziehungsweise Frauen und Männern zu erklären. Dabei sollte jedoch nicht vergessen werden, dass die Gemeinsamkeiten der Geschlechter in nahezu allen Bereichen deutlich größer sind als die Differenzen. Bei den dargestellten Befunden handelte es sich um Unterschiede zwischen Durchschnittswerten für Gruppen von Mädchen und Jungen. Die Überlappungsbereiche der Verteilungen, die diesen Durchschnittswerten zu Grunde liegen, sind jedoch durchweg größer als die Anteile, die sich nicht überschneiden. So liegen beispielsweise die Ergebnisse der besten Mädchen deutlich über dem Mittelwert der Jungen und umgekehrt (Stanat & Kunter, 2003). Jüngste Forschung thematisiert deshalb auch stärker die Gemeinsamkeiten als die Unterschiede der Geschlechter (Hyde, 2005).

Das Bewusstmachen von Unterschieden, von Bevorzungen und Benachteiligungen ist jedoch ein unverzichtbarer Schritt auf dem Weg zu Veränderungen (Ulich, 2002). Die in dieser Arbeit erfolgte Darstellung und Analyse von geschlechtsspezifischen Leistungsunterschieden soll dazu beitragen, bestehende Benachteiligungen zu erkennen, und die Möglichkeit eröffnen, Veränderungen herbeizuführen. Die vorgestellten Handlungsempfehlungen können dabei ein erster Schritt sein, den Leistungsunterschieden von Mädchen und Jungen zu begegnen.

## 11. Literaturverzeichnis

- Alfermann, D. (1996). *Geschlechterrollen und geschlechtstypisches Verhalten*. Stuttgart: Kohlhammer.
- Aronson, J., Quinn, D. M. & Spencer, S. J. (1998). Stereotype threat and the academic underperformance of minorities and women. In J. K. Swim & C. Stangor (Hg.), *Prejudice: The target's perspective* (S. 83-103). San Diego, CA: Academic Press.
- Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M. & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology*, 35, 29-46.
- Aronson, E., Wilson, T. D. & Akert, R. M. (2004). *Sozialpsychologie*. 4. Aufl. München: Pearson Studium.
- Asendorpf, J. B. (1999). *Psychologie der Persönlichkeit*. 2. Aufl. Berlin: Springer.
- Bargel, T., Ramm, M. & Multrus, F. (2004). *Studiensituation und studentische Orientierungen. 8. Studierendensurvey an Universitäten und Fachhochschulen*. Bonn: Bundesministerium für Bildung und Forschung.
- Baron-Cohen, S. (2003). „Frauen denken anders.“ In *Spiegel Special – Die Entschlüsselung des Gehirns* (S. 32-35). Hamburg: Spiegel-Verlag Rudolf Augstein.
- Baron-Cohen, S. (2004). *Vom ersten Tag an anders. Das weibliche und das männliche Gehirn*. Düsseldorf: Walter.
- Baumert, J. (1992). Koedukation oder Geschlechtertrennung. *Zeitschrift für Pädagogik*, 38, 83-110.
- Baumert, J., Stanat, P. & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In Deutsches PISA-Konsortium (Hg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15-68). Opladen: Leske + Budrich.
- Baumert, J., Artelt, C., Carstensen, C. H., Sibberns, H. & Stanat, P. (2002). Untersuchungsgegenstand, Fragestellungen und technische Grundlagen der Studie. In Deutsches PISA-Konsortium (Hg.), *PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich* (S. 11-38). Opladen: Leske + Budrich.
- Baumert, J., Trautwein, U. & Artelt, C. (2003). Schulumwelten – institutionelle Bedingungen des Lehrens und Lernens. In Deutsches PISA-Konsortium (Hg.), *PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 261-331). Opladen: Leske + Budrich.
- Beerman, L., Heller, K. A. & Menacher, P. (1992). *Mathe: nichts für Mädchen? Begabung und Geschlecht am Beispiel von Mathematik, Naturwissenschaft und Technik*. Bern: Huber.
- Benbow, C. P. & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science*, 210, 1262-1264.

- Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences*, 11, 169-232.
- Berk, L. E. (2005). *Entwicklungspsychologie*. 3. Aufl. München: Pearson Studium.
- Bilden, H. (2002). Geschlechtsspezifische Sozialisation. In K. Hurrelmann & D. Ulich (Hg.), *Handbuch der Sozialisationsforschung* (S. 279-301). 6. Aufl. Weinheim: Beltz.
- Bischof-Köhler, D. (2002). *Von Natur aus anders: Die Psychologie der Geschlechtsunterschiede*. Stuttgart: Kohlhammer.
- Blascovich, J., Spencer, S. J., Quinn, D. & Steele, C. (2001). African Americans and high blood pressure: The role of stereotype threat. *Psychological Science*, 12, 225-229.
- Bock, R. D. & Kolakowski, D. (1973). Further evidence of sex-linked major-gene influence on human spatial visualizing ability. *American Journal of Human Genetics*, 25, 1-14.
- Bosson, J. K., Haymovitz, E. L. & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. *Journal of Experimental Social Psychology*, 40, 247-255.
- Brehm, S. S., Kassin, S. M. & Fein, S. (1999). *Social psychology*. 4. Aufl. Boston: Houghton Mifflin Company.
- Brown, R. P. & Josephs, R. A. (1999). A burden of proof: Stereotype relevance and gender differences in math performance. *Journal of Personality and Social Psychology*, 76, 246-257.
- Brown, R. P. & Pinel, E. C. (2003). Stigma on my mind: Individual differences in the experience of stereotype threat. *Journal of Experimental Social Psychology*, 39, 626-633.
- Bühner, M. (2004). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Buss, D. M. (2004). *Evolutionäre Psychologie*. 2. Aufl. München: Pearson Studium.
- Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L. & Latinotti, S. (2003). Stereotype threat: The effects of expectancy on performance. *European Journal of Social Psychology*, 33, 267-285.
- Carlson, N. R. (2004). *Physiologische Psychologie*. 8. Aufl. München: Pearson Studium.
- Carstensen, C. H., Knoll, S., Rost, J. & Prenzel, M. (2004). Technische Grundlagen. In PISA-Konsortium Deutschland (Hg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 371-387). Münster: Waxmann.
- Chipman, S. F. (2005). Research on the women and mathematics issue. A personal case history. In A. M. Gallagher & J. C. Kaufman (Hg.), *Gender differences in mathematics. An integrative psychological approach* (S. 1-24). Cambridge: University Press.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. 2. Aufl. Hillsdale, NJ: Lawrence Erlbaum.
- Colberg-Schrader, H. & von Derschau, D. (2002). Sozialisationsfeld Kindergarten. In K. Hurrelmann & D. Ulich (Hg.), *Handbuch der Sozialisationsforschung* (S. 335-353). 6. Aufl. Weinheim: Beltz.
- Cooper, S. E. & Robinson, D. A. G. (1989). Childhood play activities of women and men entering engineering and science careers. *The School Counselor*, 36, 338-342.
- Davies, P. G., Spencer, S. J., Quinn, D. M. & Gerhardstein, R. (2002). Consuming images: How television commercials that elicit stereotype threat can restrain women academically and professionally. *Personality and Social Psychology Bulletin*, 28, 1615-1628.
- Davies, P. G. & Spencer, S. J. (2005). The gender-gap artifact. Women's underperformance in quantitative domains through the lens of stereotype threat. In A. M. Gallagher & J. C. Kaufman (Hg.), *Gender differences in mathematics. An integrative psychological approach* (S. 172-188). Cambridge: University Press.
- Davies, P. G., Spencer, S. J. & Steele, C. M. (2005). Clearing the air: Identity safety moderates the effects of stereotype threat on women's leadership aspirations. *Journal of Personality and Social Psychology*, 88, 276-287.
- Davison, G. C. & Neale, J. M. (2002). *Klinische Psychologie*. 6. Aufl. Weinheim: Beltz.
- Deutsches PISA-Konsortium (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Deutsches PISA-Konsortium (2002). *PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich*. Opladen: Leske + Budrich.
- Deutsches PISA-Konsortium (2003). *PISA 2000 – Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland*. Opladen: Leske + Budrich.
- „Die Welt“ vom 27. August 2005.
- Eccles, J. S. (1987). Gender roles and women's achievement-related decisions. *Psychology of Women Quarterly*, 11, 135-172.
- Eccles, J. S. (1994). Understanding women's educational and occupational choices. Applying the Eccles et al. model of achievement-related choices. *Psychology of Women Quarterly*, 18, 585-609.
- Esser, H. (1999a). *Soziologie. Allgemeine Grundlagen*. 2. Aufl. Frankfurt: Campus.
- Esser, H. (1999b). *Soziologie. Spezielle Grundlagen. Band 1: Situationslogik und Handeln*. Frankfurt: Campus.
- Esser, H. (2001). *Soziologie. Spezielle Grundlagen. Band 6: Sinn und Kultur*. Frankfurt: Campus.

- Faulstich-Wieland, H. (2001). Mädchen: besser in der Schule ... aber benachteiligt bei der Ausbildung. *Erziehung und Wissenschaft. Zeitschrift der Bildungsgewerkschaft GEW*, 29 (2), 14-15.
- Fennema, E. & Peterson, P. (1985). Autonomous learning behavior: A possible explanation of gender-related differences in mathematics. In L. C. Wilkinson & C. B. Marrett (Hg.), *Gender influences in classroom interaction* (S. 17-35). Orlando: Academic Press.
- Ford, T. E., Ferguson, M. A., Brooks, J. L. & Hagadone, K. M. (2004). Coping sense of humor reduces effects of stereotype threat on women's math performance. *Personality and Social Psychology Bulletin*, 30, 643-653.
- Fox, J. (1999). *Applied regression analysis, linear models, and related methods*. Thousand Oaks: Sage.
- Gallagher, A. M. & Kaufman, J. C. (2005). Gender differences in mathematics. What we know and what we need to know. In A. M. Gallagher & J. C. Kaufman (Hg.), *Gender differences in mathematics. An integrative psychological approach* (S. 316-331). Cambridge: University Press.
- Geißler, R. (2004). Geschlechtsspezifische Ungleichheit. In Bundeszentrale für politische Bildung (Hg.), *Sozialer Wandel in Deutschland. Informationen zur politischen Bildung, Nr. 269* (S. 54-61). München: Franzis' print & media.
- Geschwind, N. & Galaburda, A. M. (1987). *Cerebral lateralization. Biological mechanisms, associations, and pathology*. Cambridge: MIT Press.
- Gonzales, P. M., Blanton, H. & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. *Personality and Social Psychology Bulletin*, 28, S. 659-670.
- Halpern, D. F. (2000). *Sex differences in cognitive abilities*. 3. Aufl. Mahwah, NJ: Lawrence Erlbaum Associates.
- Halpern, D. F. & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review*, 12, 229-246.
- Halpern, D. F., Wai, J. & Saw, A. (2005). A psychobiosocial model. Why females are sometimes greater than sometimes less than males in math achievement. In A. M. Gallagher & J. C. Kaufman (Hg.), *Gender differences in mathematics. An integrative psychological approach* (S. 48-72). Cambridge: University Press.
- Hannover, B. (1991). Zur Unterrepräsentanz von Mädchen in Naturwissenschaften und Technik: Psychologische Prädiktoren der Fach- und Berufswahl. *Zeitschrift für Pädagogische Psychologie*, 5, 169-186.
- Hannover, B. (1992). Geschlecht und Interessenentwicklung. Mädchen in geschlechtsuntypischen Berufen. Eine quasiexperimentelle Studie zur Förderung des Interesses Jugendlicher an Naturwissenschaft und Technik. *Zeitschrift für Sozialpsychologie*, 23, 36-45.
- Henecka, H. P. (1997). *Grundkurs Soziologie*. 6. Aufl. Opladen: Leske + Budrich.

- Higgins, E. T. (1997). Beyond pleasure and pain. *American Psychologist*, 52, 1280-1300.
- Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. *Advances in Experimental Social Psychology*, 30, 1-46.
- Higgins, E. T. & Silberman, I. (1998). Development of regulatory Focus: Promotion and prevention as ways of living. In J. Heckhausen & C. S. Dweck (Hg.), *Motivation and self-regulation across the life span* (S. 78-113). Cambridge: University Press.
- Hillmann, K.-H. (1994). *Wörterbuch der Soziologie*. 4. Aufl. Stuttgart: Kröner.
- Horstkemper, M. (1992). Koedukation in mathematisch-naturwissenschaftlichen Fächern – zweifelhafter Gewinn für die Mädchen oder Entwicklungschance für alle? In A. Grabosch & A. Zwölfer (Hg.), *Frauen und Mathematik. Die allmähliche Rückeroberung der Normalität?* Tübingen: Attempto.
- Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using  $\omega^2$  and  $d$ . *American Psychologist*, 36, 892-901.
- Hyde, J. S., Fennema, E. & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107, 139-155.
- Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A. & Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect. A meta-analysis. *Psychology of Women Quarterly*, 14, 299-324.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, 60, 581-592.
- Inzlicht, M. & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science*, 11, 365-371.
- Jacobs, J. E., Davies-Kean, P., Bleeker, M., Eccles J. S. & Malanchuk, O. (2005). „I can, but I don't want to.“ The impact of parents, interests, and activities on gender differences in math. In A. M. Gallagher & J. C. Kaufman (Hg.), *Gender differences in mathematics. An integrative psychological approach* (S. 246-263). Cambridge: University Press.
- James, K. & Greenberg, J. (1997). Beliefs about self and about gender groups: Interactive effects on the spatial performance of women. *Basic and Applied Social Psychology*, 19, 411-425.
- Jensen, A. R. (1998). *The g factor. The science of mental ability*. Westport, CT: Praeger.
- Judkins, D. R. (1990). Fay's method for variance estimation. *Journal of Official Statistics*, 6, 233-239.
- Kasten, H. (2001). Geschlechtsunterschiede. In D. H. Rost (Hg.), *Handwörterbuch Pädagogische Psychologie* (S. 212-219). 2. Aufl. Weinheim: Beltz.
- Keller, C. (1998). *Geschlechterdifferenzen in der Mathematik: Prüfung von Erklärungsansätzen. Eine mehrebenenanalytische Untersuchung im Rahmen der Third International Mathematics and Science Study*. Zürich: Zentralstelle der Studentenschaft.

- Keller, J. (2002). Blatant stereotype threat and women's math performance: Self-handicapping as a strategic means to cope with obtrusive negative performance expectations. *Sex Roles, 47*, 193-198.
- Keller, J. & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women's math performance. *Personality and Social Psychology Bulletin, 29*, 371-381.
- Keller, J. (2004). *Expectancy effects in performance situations*. Lengerich: Pabst Science Publishers.
- Kimball, M. M. (1989). A new perspective on women's mathe achievement. *Psychological Bulletin, 105*, 198-214.
- Kimura, D. (1992). Weibliches und männliches Gehirn. *Spektrum der Wissenschaft, 15 (11)*, 104-113.
- Klieme, E. (1986). Bildliches Denken als Mediator für Geschlechterunterschiede beim Lösen mathematischer Probleme. In H.-G. Steiner (Hg.), *Grundfragen der Entwicklung mathematischer Fähigkeiten* (S. 133-151). Köln: Aulis-Verlag Deubner.
- Köller, O. & Klieme, E. (2000). Geschlechtsdifferenzen in den mathematisch-naturwissenschaftlichen Leistungen. In J. Baumert, W. Bos & R. Lehmann (Hg.), *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 2. Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe* (S. 373-404). Opladen: Leske + Budrich.
- Kohler, U. & Kreuter, F. (2001). *Datenanalyse mit Stata. Allgemeine Konzepte der Datenanalyse und ihre praktische Anwendung*. München: Oldenbourg.
- Kunter, M., Schümer, G., Artelt, C., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K-J. & Weiß, M. (2002). *PISA 2000: Dokumentation der Erhebungsinstrumente*. Berlin: Max-Planck-Institut für Bildungsforschung.
- Lewontin, R. C., Rose, S. & Kamin, L. J. (1988). *Die Gene sind es nicht ...: Biologie, Ideologie und menschliche Natur*. München: Psychologie Verlags Union.
- Lind, D. & Knoche, N. (2004). Testtheoretische Modelle und Verfahren bei PISA-2000-Mathematik. In M. Neubrand (Hg.), *Mathematische Kompetenzen von Schülerinnen und Schülern in Deutschland. Vertiefende Analysen im Rahmen von PISA 2000* (S. 51-69). Wiesbaden: Verlag für Sozialwissenschaften.
- Linn, M. C. & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479-1498.
- Maccoby, E. E. & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CF: Stanford University Press.
- Maier, P. H. (1994). *Räumliches Vorstellungsvermögen – Komponenten, geschlechtsspezifische Differenzen, Relevanz, Entwicklung und Realisierung in der Realschule*. Frankfurt: Lang.

- Maier, P. H. (1996). Geschlechtsspezifische Differenzen im räumlichen Vorstellungsvermögen. *Psychologie in Erziehung und Unterricht*, 43, 245-265.
- Marx, D. M. & Roman, J. S. (2002). Female role models: Protecting women's math test performance. *Personality and Social Psychology Bulletin*, 28, 1183-1193.
- McIntyre, R. B., Paulson, R. M. & Lord, C. G. (2003). Alleviating women's mathematics stereotype threat through salience of group achievements. *Journal of Experimental Social Psychology*, 39, 83-90.
- Muller, D., Judd, C. M. & Yzerbyt, V. Y. (in Druck). *When moderation is mediated and mediation is moderated*.
- Menacher, P. (1994). Erklärungsansätze für geschlechtsspezifische Interessen- und Leistungsunterschiede in Mathematik, Naturwissenschaften und Technik. *Zentralblatt für Didaktik der Mathematik*, 26, 1-11.
- Nuttall, R. L., Casey, M. B. & Pezaris, E. (2005). Spatial ability as a mediator of gender differences on mathematics tests. A biological-environmental framework. In A. M. Gallagher & J. C. Kaufman (Hg.), *Gender differences in mathematics. An integrative psychological approach* (S. 121-142). Cambridge: University Press.
- O'Brien, L. T. & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. *Personality and Social Psychology Bulletin*, 29, 782-789.
- OECD (2001). *Lernen für das Leben. Erste Ergebnisse von PISA 2000*. Paris: Organisation für wirtschaftliche Zusammenarbeit und Entwicklung.
- OECD (2004a). *Lernen für die Welt von morgen. Erste Ergebnisse von PISA 2003*. Paris: Organisation für wirtschaftliche Zusammenarbeit und Entwicklung.
- OECD (2004b). *Bildung auf einen Blick. OECD-Indikatoren 2004*. Paris: Organisation für wirtschaftliche Zusammenarbeit und Entwicklung.
- Oerter, R. (1998). Kindheit. In R. Oerter & L. Montada (Hg.), *Entwicklungspsychologie* (S. 249-309). 4. Aufl. Weinheim: Beltz.
- Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology*, 26, 291-310.
- Oswald, D. L. & Harvey, R. D. (2000/01). Hostile environments, stereotype threat, and math performance among undergraduate women. *Current Psychology: Developmental, Learning, Personality, Social*, 19, 338-356.
- Oyserman, D., Harrison, K. & Bybee, D. (2001). Can racial identity be promotive of academic efficacy. *International Journal of Behavioral Development*, 25, 379-385.
- Paglin, M. & Rufolo, A. M. (1990). Heterogeneous human capital, occupational choice, and male-female earnings differences. *Journal of Labor Economics*, 8, 123-144.

- Pekrun, R. & Zirngibl, A. (2004). Schülermerkmale im Fach Mathematik. In PISA-Konsortium Deutschland (Hg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 191-210). Münster: Waxmann.
- Pinel, E. C. (1999). Stigma consciousness: The psychological legacy of social stereotypes. *Journal of Personality and Social Psychology*, 76, 114-128.
- PISA-Konsortium Deutschland (2004). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.
- Prenzel, M., Rost, J., Senkbeil, M., Häußler, P. & Klopp, A. (2001). Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse. In Deutsches PISA-Konsortium (Hg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 191-248). Opladen: Leske + Budrich.
- Prenzel, M., Carstensen, C. H., Rost, J. & Senkbeil, M. (2002). Naturwissenschaftliche Grundbildung im Ländervergleich. In Deutsches PISA-Konsortium (Hg.), *PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich* (S. 129-158). Opladen: Leske + Budrich.
- Prenzel, M., Drechsel, B., Carstensen, C. H. & Ramm, G. (2004). PISA 2003 – eine Einführung. In PISA-Konsortium Deutschland (Hg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 13-46). Münster: Waxmann.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rost, J. & Schiefele, U. (2005). *Vorinformation zu PISA 2003: Zentrale Ergebnisse des zweiten Vergleichs der Länder in Deutschland*. Verfügbar unter: [http://pisa.ipn.uni-kiel.de/Vorinformation\\_E.pdf](http://pisa.ipn.uni-kiel.de/Vorinformation_E.pdf) (27. September 2005).
- Quinn, D. M. & Spencer, S. J. (2001). The interference of stereotype threat with women's generation of mathematical problem-solving strategies. *Journal of Social Issues*, 57, 55-71.
- Rahhal, T. A., Hasher, L. & Colcombe, S. J. (2001). Instructional manipulations and age differences in memory: Now you see them, now you don't. *Psychology and Aging*, 16, 697-706.
- Richter, S. (1996). *Unterschiede in den Schulleistungen von Mädchen und Jungen. Geschlechtsspezifische Aspekte des Schriftspracherwerbs und ihre Berücksichtigung im Unterricht*. Regensburg: Roderer.
- Rost, J. (1999). Was ist aus dem Rasch-Modell geworden? *Psychologische Rundschau*, 50, 140-156.
- Sackett, P. R., Schmitt, N., Ellingson, J. E. & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education. Prospects in a post-affirmative-action World. *American Psychologist*, 56, 302-318.

- Schiefele, U., Krapp, A. & Schreyer, I. (1993). Metaanalyse des Zusammenhangs von Interesse und schulischer Leistung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 25, 120-148.
- Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. *Journal of Experimental Social Psychology*, 38, 194-201.
- Schmader, T. & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85, 440-452.
- Schnell, R., Hill, P. B. & Esser, E. (1999). *Methoden der empirischen Sozialforschung*. 6. Aufl. München: Oldenbourg.
- Sharps, M. J., Welton, A. L. & Price, J. L. (1993). Gender and task in the determination of spatial cognitive performance. *Psychology of Women Quarterly*, 17, 71-83.
- Sharps, M. J., Price, J. L. & Williams, J. K. (1994). Spatial cognition and gender. Instructional and stimulus influences on mental image rotation performance. *Psychology of Women Quarterly*, 18, 413-425.
- Sibberns, H. & Baumert, J. (2001). Stichprobenziehung und Stichprobengewichtung. In Deutsches PISA-Konsortium (Hg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 511-517). Opladen: Leske + Budrich.
- Silverman, I. & Eals, M. (1992). Sex differences in spatial abilities: Evolutionary theory and data. In J. H. Barkow, L. Cosmides & J. Tooby (Hg.), *The adapted mind. Evolutionary psychology and the generation of culture* (S. 533-549). New York: Oxford University Press.
- Silverman, I. & Philips, K. (1998). The evolutionary psychology of spatial sex differences. In C. Crawford & D. L. Krebs (Hg.), *Handbook of evolutionary psychology. Ideas, issues, and applications* (S. 595-612). Mahwah, NJ: Lawrence Erlbaum.
- Smith, J. L. & White, P. H. (2002). An examination of implicitly activated, explicitly activated, and nullified stereotypes on mathematical performance: It's not just a woman's issue. *Sex Roles*, 47, 179-191.
- Spencer, S. J., Steele, C. M. & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4-28.
- Springer, S. P. & Deutsch, G. (1998). *Linkes – rechtes Gehirn*. 4. Aufl. Heidelberg: Spektrum Akademischer Verlag.
- Srocke, B. (1989). *Mädchen und Mathematik. Historisch-systematische Untersuchung der unterschiedlichen Bedingungen des Mathematiklernens von Mädchen und Jungen*. Wiesbaden: Deutscher Universitäts-Verlag.
- Stanat, P. & Kunter, M. (2001). Geschlechterunterschiede in Basiskompetenzen. In Deutsches PISA-Konsortium (Hg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 249-269). Opladen: Leske + Budrich.

- Stanat, P. & Kunter, M. (2002). Geschlechterspezifische Leistungsunterschiede bei Fünfzehnjährigen im internationalen Vergleich. *Zeitschrift für Erziehungswissenschaft*, 4, 28-48.
- Stanat, P. & Kunter, M. (2003). Kompetenzerwerb, Bildungsbeteiligung und Schullaufbahn von Mädchen und Jungen im Ländervergleich. In Deutsches PISA-Konsortium (Hg.), *PISA 2000. Ein differenzierter Blick auf die Länder der Bundesrepublik Deutschland* (S. 211-242). Opladen: Leske + Budrich.
- Steele, C. M. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69, 797-811.
- Steele, C. M. (1997). A threat in the air. How stereotypes shape intellectual identity and performance. *American Psychologist*, 52, 613-629.
- Steele, C. M., Spencer, S. J. & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology*, 34, 379-440.
- Stern, E. (1997). Mathematik. In F. E. Weinert (Hg.), *Enzyklopädie der Psychologie. Band 3: Psychologie des Unterrichts und der Schule* (S. 397-426). Göttingen: Hogrefe.
- Stone, J., Lynch, C. I., Sjomeling, M. & Darley, J. M. (1999). Stereotype threat effects on black and white athletic performance. *Journal of Personality and Social Psychology*, 77, 1213-1227.
- Trautner, H. M. (1994). Geschlechtsspezifische Erziehung und Sozialisation. In K. A. Schneewind (Hg.), *Enzyklopädie der Psychologie. Band 1: Psychologie der Erziehung und Sozialisation* (S. 167-195). Göttingen: Hogrefe.
- Ulich, K. (2002). Schulische Sozialisation. In K. Hurrelmann & D. Ulich (Hg.), *Handbuch der Sozialisationsforschung* (S. 378-396). 6. Aufl. Weinheim: Beltz.
- Walsh, M., Hickey, C. & Duffy, J. (1999). Influence of item content and stereotype situation on gender differences in mathematical problem solving. *Sex Roles*, 41, 219-240.
- Werth, L., Denzler, M. & Förster, J. (2002). Was motiviert wen? Worauf der Fokus liegt, entscheidet über den Erfolg. *Wirtschaftspsychologie*, 3, 5-12.
- Whaley, A. L. (1998). Issues of validity in empirical tests of stereotype threat theory. *American Psychologist*, 53, 679-680.
- Ziegler, A., Kuhn, C. & Heller, K. A. (1998). Implizite Theorien von gymnasialen Mathematik- und Physiklehrkräften zu geschlechtsspezifischer Begabung und Motivation. *Psychologische Beiträge*, 40, 271-287.
- Zimbardo, P. G. & Gerrig, R. J. (1999). *Psychologie*. 7. Aufl. Berlin: Springer.
- Zimmer, K., Burba, D. & Rost, J. (2004). Kompetenzen von Jungen und Mädchen. In PISA-Konsortium Deutschland (Hg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 211-223). Münster: Waxmann.

## 12. Anhang: Abbildungsverzeichnis

Abbildung 3-1:	Geschlechtsspezifische Leistungsunterschiede in der PISA-Studie 2003 in Punkten (nach OECD, 2004a)	Seite 11
Abbildung 4-1:	Geschlechtsspezifische Unterschiede bei Schülermerkmalen in der PISA-Studie 2003 (nach OECD, 2004a)	Seite 31
Abbildung 4-2:	Modell des autonomen Lernverhaltens (nach Fennema & Peterson, 1985)	Seite 32
Abbildung 4-3:	Modell zur Vorhersage von leistungsbezogenen Entscheidungen (nach Eccles, 1987)	Seite 34
Abbildung 4-4:	Psychobiosoziales Modell (nach Halpern, 2000)	Seite 36
Abbildung 4-5:	Testleistung in Abhängigkeit von Testbeschreibung und Geschlecht (nach Spencer et al., 1999)	Seite 42
Abbildung 4-6:	Testleistung in Abhängigkeit von Testschwierigkeit und Geschlecht (nach Spencer et al., 1999)	Seite 43
Abbildung 4-7:	Promotion Focus (nach Higgins, 1998)	Seite 54
Abbildung 4-8:	Prevention Focus (nach Higgins, 1998)	Seite 55
Abbildung 4-9:	„Moderation of Expectancy Effects by Regulatory Focus“-Modell (nach Keller, 2004)	Seite 57
Abbildung 4-10:	Testleistung in Abhängigkeit von Testbeschreibung und Geschlecht (nach Brown & Josephs, 1999)	Seite 60

## 12. Anhang: Tabellenverzeichnis

Tabelle 3-1:	Geschlechtsunterschiede in mathematischen Fähigkeiten (nach Hyde et al., 1990)	Seite 9
Tabelle 4-1:	Meinungsumfrage zum „Leben in Mannheim“	Seite 14
Tabelle 4-2:	Experimentelle Arbeiten zum Stereotype-Threat-Effekt bei mathematischen Leistungstests	Seite 49
Tabelle 4-3:	Übersicht über potenzielle Mediatoren des Stereotype-Threat-Effekts	Seite 50
Tabelle 6-1:	Regressionsmodell 1 (Deutschland)	Seite 73
Tabelle 6-2:	Regressionsmodell 2 (Deutschland)	Seite 74
Tabelle 6-3:	Regressionsmodell 3 (Deutschland)	Seite 75
Tabelle 6-4:	Regressionsmodell 4 (Deutschland)	Seite 75
Tabelle 6-5:	Regressionsmodell 5 (Deutschland)	Seite 76
Tabelle 6-6:	Regressionsmodell 6 (Deutschland)	Seite 76
Tabelle 6-7:	Regressionsmodell 7 (Deutschland)	Seite 77
Tabelle 6-8:	Regressionsmodell 8 (Deutschland)	Seite 78
Tabelle 6-9:	Regressionsmodell 9 (Deutschland)	Seite 81
Tabelle 6-10:	Regressionsmodell 10 (Deutschland)	Seite 82
Tabelle 6-11:	Mathematiknote nach Geschlecht	Seite 83
Tabelle 6-12:	Mathematiknote und PISA-Punktzahl im Vergleich (standardisierte Regressionskoeffizienten)	Seite 84
Tabelle 6-13:	Untersuchungsländer	Seite 85
Tabelle 6-14:	Regressionsmodell Deutschland, Österreich und Schweiz 1	Seite 86
Tabelle 6-15:	Regressionsmodell Deutschland, Österreich und Schweiz 2	Seite 87
Tabelle 6-16:	Regressionsmodell Dänemark, Finnland, Norwegen und Schweden 1	Seite 89
Tabelle 6-17:	Regressionsmodell Dänemark, Finnland, Norwegen und Schweden 2	Seite 90
Tabelle 6-18:	Regressionsmodell Spanien, Italien und Portugal 1	Seite 91

Tabelle 6-19:	Regressionsmodell Spanien, Italien und Portugal 2	Seite 93
Tabelle 6-20:	Regressionsmodell Niederlande, Belgien und Luxemburg 1	Seite 94
Tabelle 6-21:	Regressionsmodell Niederlande, Belgien und Luxemburg 2	Seite 95
Tabelle 6-22:	Geschlechtsspezifische Leistungsunterschiede in den Ländern (sortiert nach d-Werten)	Seite 97
Tabelle 6-23:	Vergleich Österreich und Schweiz 1	Seite 99
Tabelle 6-24:	Vergleich Österreich und Schweiz 2	Seite 99
Tabelle 6-25:	Vergleich Österreich und Schweiz 3	Seite 100
Tabelle 6-26:	Vergleich Dänemark und Norwegen 1	Seite 101
Tabelle 6-27:	Vergleich Dänemark und Norwegen 2	Seite 101
Tabelle 6-28:	Vergleich Dänemark und Norwegen 3	Seite 101
Tabelle 6-29:	Vergleich Italien und Spanien 1	Seite 102
Tabelle 6-30:	Vergleich Italien und Spanien 2	Seite 102
Tabelle 6-31:	Vergleich Italien und Spanien 3	Seite 103
Tabelle 7-1:	Verwendete Messinstrumente bei den Analysen mit dem PISA-E-Datensatz 2000	Seite 107
Tabelle 7-2:	Regressionsmodell 1 (PISA-E-Studie)	Seite 108
Tabelle 7-3:	Regressionsmodell 2 (PISA-E-Studie)	Seite 109
Tabelle 7-4:	Regressionsmodell 3 (PISA-E-Studie)	Seite 110
Tabelle 7-5:	Regressionsmodell 4 (PISA-E-Studie)	Seite 111
Tabelle 7-6:	Regressionsmodell 5 (PISA-E-Studie)	Seite 113

## **12. Anhang: Hinweis zur DVD**

Auf der beiliegenden DVD befinden sich die verwendeten PISA-Datensätze. Im Ordner „PISA 2003“ befindet sich der internationale PISA-Datensatz 2003 als txt-Datei mit zugehöriger SPSS-Syntax sowie als SPSS- und Stata-Datei. Im Ordner „PISA 2000“ befindet sich der deutsche Erweiterungsdatensatz der PISA-Studie 2000 als SPSS- und als Stata-Datei. Im Ordner „Do-Files“ befinden sich die Do-Files, im Ordner „Log-Files“ die vom Statistikprogramm Stata ausgegebenen Ergebnisse. Im Hauptverzeichnis befindet sich diese Arbeit als pdf-Dokument.

Das Cover der DVD zeigt eine schriftliche Abiturprüfung an der Albertus-Magnus-Schule (AMS) in Viernheim. Das Foto wurde freundlicherweise von Bernhard Kreutzer zur Verfügung gestellt.

## **12. Anhang: Persönliche Erklärung**

Ich versichere, dass ich diese Diplomarbeit ohne Hilfe Dritter und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt und die den benutzten Quellen wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Viernheim, 5. Oktober 2005

*Markus Tausendpfund*