



universität  
wien

# Diplomarbeit

Titel der Arbeit

Lehrererwartungen – Schülerleistungen:  
Gibt es einen Zusammenhang?

Verfasserin

Madeleine Scholze

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag<sup>a</sup>. rer. nat.)

Wien, im März 2010

Studienkennzahl: 298  
Studienrichtung: Psychologie  
Betreuer: Ao. Univ.-Prof. Dr. Alfred Schabmann

## **Danksagung**

Vorab möchte ich mich bei all jenen bedanken, die mich beim Verfassen dieser Arbeit und während meines Studiums unterstützten.

Zu allererst bedanke ich mich bei meinem Lebensgefährten für seine emotionale Unterstützung während des gesamten Studiums, für seine unermüdliche Geduld und dafür, dass er immer an mich geglaubt hat.

Meinen Eltern danke ich für ihre finanzielle und emotionale Unterstützung und dafür, dass sie mich auf meinem Weg immer begleitet und unterstützt haben.

Außerdem möchte ich meinem Bruder und all meinen Freunden und Kollegen für ihre tatkräftige Unterstützung beim Schreiben dieser Arbeit danken. Ihre Fragen, Gedanken und Ratschläge haben mich immer wieder weitergebracht, v.a. wenn es gerade schwierig war.

Herrn Professor Schabmann danke ich für seine methodische wie auch praktische Unterstützung während der Betreuung meiner Diplomarbeit.

## **Erklärung**

In der vorliegenden Arbeit verzichte ich auf eine gendersensible Sprache, da mir diese Schreibweise vertrauter ist und aufgrund besserer Lesbarkeit. Es wird jedoch immer sowohl das weibliche als auch das männliche Geschlecht angesprochen.

# 1 Inhaltsverzeichnis

1	Inhaltsverzeichnis .....	3
2	Einleitung.....	5
3	Begriffserklärung .....	7
4	Erwartungseffekt in der Schule.....	10
4.1	Die Anfänge des Erwartungseffekts im Unterricht.....	10
4.2	Schwächen des „Oak School“ Experiments .....	13
4.3	Aktueller Forschungsstand zum Erwartungseffekt in der Schule .....	16
4.3.1	Erwartungseffekt und „benachteiligte“ Schülergruppen .....	17
4.3.2	Galatea- oder Golem-Effekt?.....	19
4.3.3	Gibt es differenziertes Lehrerverhalten gegenüber Schülern? .....	20
5	Wie genau sind Lehrerurteile? .....	24
5.1	Welche Schülermerkmale moderieren Lehrerurteile? .....	29
5.1.1	Verhalten .....	30
5.1.2	Geschlecht .....	32
5.1.3	Leistungsniveau und Intelligenz der Schüler .....	34
6	Zielsetzung der Untersuchung.....	37
7	Methode.....	40
7.1	Stichprobe .....	40
7.2	Erhebungsinstrumente .....	40
7.2.1	Lehrereinschätzungen .....	40
7.2.2	Lese- und Rechtschreibfähigkeit .....	43
8	Statistische Analyse .....	45
8.1	Akkuratheit der Lehrerurteile.....	45
8.2	Parallelisierung .....	46
9	Ergebnisse.....	50
9.1	Lesen .....	50
9.2	Rechtschreiben .....	52

10	Zusammenfassung und Diskussion .....	56
11	Abbildungsverzeichnis .....	59
12	Literatur.....	60

## 2 Einleitung

Seit vielen Jahren beschäftigt sich die Wissenschaft mit dem Thema des Erwartungseffekts, da dieser in vielen zwischenmenschlichen Beziehungen eine bedeutende Rolle spielt und daher große Praxisrelevanz hat.

Der amerikanische Soziologe Thomas meinte 1928 „if men define situations as real, they are real in their consequences“ (zitiert nach Merton, 1957, S. 475) und begründete mit diesem Theorem die literarischen Anfänge zu dem Thema des Erwartungseffekts. Rosenthal und Fode (1963) und Rosenthal und Jacobson (1968) haben in den 1960er Jahren begonnen das Theorem von Thomas auf Erwartungseffekte von Versuchsleitern und Lehrern zu übertragen (zitiert nach Berka & Westhoff, 1981).

Wie sich Lehrererwartungen auf die schulischen Leistungen von Volksschulkindern auswirken, untersuchten Rosenthal und Jacobson 1968 in den USA in ihrer weltbekannten Studie „Pygmalion im Unterricht“. Durch die Ergebnisse ihrer Arbeit wurde das Interesse vieler Wissenschaftler für diesen Forschungsbereich geweckt. Bis heute gibt es zahlreiche Untersuchungen zum Thema Lehrererwartungseffekt.

Die vorliegende Untersuchung beschäftigt sich mit der Frage, ob die Erwartungen von Seiten der Lehrer bezüglich der Lese- und Rechtschreibleistungen ihrer Schüler einen Einfluss auf deren spätere Leistungen in diesen Bereichen haben. Wenn sich in der vorliegenden Studie ein Erwartungseffekt zeigt, ist außerdem zu prüfen, ob das Überschätzen oder das Unterschätzen von schulischen Leistungen einen stärkeren Einfluss auf die späteren Leistungen der Schüler hat. Dieses Thema ist von besonderer Praxisrelevanz, v.a. wenn sich zeigen sollte, dass das Unterschätzen von Seiten der Lehrer dazu führt, dass die Leistungen der Schüler auch tatsächlich schlechter werden. Im Zusammenhang mit dem Erwartungseffekt ist auch immer die Frage zu klären, wie genau Lehrer ihre Schüler beurteilen können, da für das Auftreten eines Erwartungseffekts fehlerhafte Lehrereinschätzungen eine Voraussetzung sind. Daher wird in dieser Untersuchung auch überprüft wie akkurat Lehrereinschätzungen sind und ob

Lehrer bei der Leistungsbeurteilung auch leistungsferne Schülermerkmale, wie beispielsweise Verhalten und Geschlecht, mitberücksichtigen.

Da sich diese Arbeit im Wesentlichen mit der Frage beschäftigt, ob sich durch fehlerhafte Lehrererwartungen, wie das Über- und Unterschätzen von schulischen Leistungen, ein Erwartungseffekt zeigt, wird im Folgenden zunächst näher darauf eingegangen, wie sich dieses Forschungsgebiet über die Jahre hinweg entwickelt hat. Zuerst werden die Anfänge dieses Wissenschaftsgebiets beleuchtet und anschließend der heutige Wissensstand genauer dargestellt. Wie bereits erwähnt, ist im Zusammenhang mit diesem Untersuchungsfeld auch die Frage interessant, wie akkurat Lehrereinschätzungen sind. Aus diesem Grund beschäftigt sich die vorliegende Arbeit auch mit dieser Thematik.

Für ein besseres Verständnis dieser Arbeit ist es wichtig zu Beginn einige Begriffe genauer zu betrachten, daher werden vorerst die wichtigsten Begriffe erklärt.

### 3 Begriffserklärung

Zuerst soll der Begriff des Erwartungseffekts definiert werden, da sich, wie bereits erwähnt, diese Untersuchung v.a. mit diesem beschäftigt.

Robert Rosenthal (2000, S. 294) definiert den Begriff folgendermaßen: „Expectancy effects refer to the situation in which expectations we hold for others come to serve as self-fulfilling prophecies [...]“ Ludwig (1991, S. 45) unterteilt den spezifischen Erwartungseffekt in zwei Effekte mit unterschiedlicher Auswirkungsrichtung: Selbsterfüllungseffekt (self fulfilling prophecy) und Anti-Selbsterfüllungseffekt, wobei für diese Arbeit nur der erste Effekt von Bedeutung ist. Er definiert diesen folgendermaßen: „Beim **Selbsterfüllungseffekt** [...] steuert die Erwartung den Verlauf des erwarteten Ereignisses in Richtung der Erwartung. Die Erwartung selbst führt zu ihrer eigenen *Verifikation*.“

Weiters ist zwischen einem interpersonalen und einem intrapersonalen Erwartungseffekt zu unterscheiden. Ersterer meint Erwartungen die sich auf das Verhalten einer anderen Person beziehen, wohingegen der Zweite jene Erwartungen betrifft welche das Verhalten der eigenen Person betreffen (Ludwig, 2006). Die vorliegende Arbeit beschäftigt sich ausschließlich mit dem interpersonalen Erwartungseffekt.

Der bereits oben erwähnte Begriff der Self-fulfilling-prophecy (SFP) wird sowohl in dieser Arbeit als auch in vielen anderen Arbeiten synonym mit dem Begriff des Erwartungseffekts verwendet. Im deutschen Sprachraum spricht man von einer Sich-selbst-erfüllenden-Prophezeiung, welche, laut Lexikon der Psychologie (2001, S. 170), als ein psychischer Mechanismus definiert wird, dem eine spezifische Erwartungshaltung bzw. Attribution und vorteilvolles, diskriminierendes Verhalten gegenüber einer anderen Person oder sozialen Gruppe zugrunde liegt. Durch die Zuschreibung von bestimmten Verhaltensweisen wird ein Prozess in Gang gesetzt, der bei diesen Personen oder Gruppen schließlich das vermutete Verhalten nach sich zieht, das die Erwartung bestätigt. Bei dieser Definition wird das Verhalten, das durch die Erwartung einer Person einer anderen gegenüber hervorgerufen wird, in den Prozess der Erwartungserfüllung mit einbezogen. Die Begriffserklärung meint: Das Verhalten,

das durch eine bestimmte Erwartungshaltung einer Person gegenüber einer anderen entsteht, verursacht das Eintreten des gewünschten Verhaltens bei der anderen Person.

Das Konzept der SFP wurde 1948 von Merton in die Sozialwissenschaften eingeführt und bildet heute die theoretische Basis für einen Großteil der Untersuchungen zum Thema Erwartungseffekt (z.B. Rosenthal, 2000; Ludwig, 2006; Casparis, 1980). Laut der Definition von Merton (1948, zitiert nach Jussim & Harber, 2005) müssen die Erwartungen der Lehrer ihren Schülern gegenüber fehlerhaft sein, damit man von einer SFP sprechen kann. Merton verwendete das theoretische Konzept der SFP zur Erklärung gesellschaftlicher und ökonomischer Phänomene (Ludwig, 2006).

Eine weitere Unterteilung des Phänomens der SFP bezieht sich auf die Unterscheidung zwischen einem negativen und einem positiven Erwartungseffekt. Ein negativer Erwartungseffekt wird in der Literatur auch als Golem-Effekt bezeichnet. Mit diesem Effekt ist gemeint, dass, z.B. bezogen auf die Schule, das Unterschätzen von Schülern von Seiten der Lehrer zu einer Verschlechterung der Schülerleistungen führt. Der Zweite, der positive Erwartungseffekt, wird Galatea-Effekt genannt und kennzeichnet den gegenteiligen Effekt, dass sich durch eine Überschätzung die Schülerleistungen verbessern (z.B. Babad, Inbar & Rosenthal, 1982; Rubie-Davies, Hattie & Hamilton, 2006).

Ein Spezialfall der SFP/des Erwartungseffekts ist der Pygmalioneffekt (Berka & Westhoff, 1981). Rosenthal verwendete für seine erste Studie zum Thema Erwartungseffekt in der Schule den Namen des zyprischen Bildhauers Pygmalion, nach Ovid, aus der antiken Mythologie (Werner, 1999). Dadurch etablierte sich der Ausdruck Pygmalioneffekt in der Bedeutung eines interpersonalen Erwartungseffekts im Schulunterricht. Da dieser Begriff durch Rosenthal geprägt wurde, wird er in der Literatur auch immer wieder als Rosenthal-Effekt bezeichnet. Im Allgemeinen bedeuten diese beiden Begriffe, laut Berka und Westhoff (1981, S. 2), dass sich ein Schüler gemäß den Erwartungen

seines Lehrers verhält, weil der Lehrer diese hat. In der wissenschaftlichen Literatur werden die beiden Begriffe auch häufig im nicht-schulischen Bereich verwendet (Ludwig, 2006).

In der vorliegenden Arbeit werden die vier soeben aufgezählten Begriffe (Erwartungseffekt, SFP, Pygmalioneffekt und Rosenthal-Effekt) synonym verwendet.

Abschließend soll nun noch etwas genauer beschrieben werden, welche Voraussetzungen für das Auftreten eines Erwartungseffekts in der Schule erfüllt sein müssen.

Brophy und Good (1970) haben in ihrer Arbeit ein Modell dargestellt, welches die Faktoren beschreibt, die vorhanden sein müssen, damit Lehrererwartungen in Form einer SFP wirksam werden können. Das Modell der Autoren verläuft in 6 Schritten:

1. Ein Lehrer entwickelt verschiedene Erwartungen den Leistungen seiner Schüler gegenüber.
2. Er beginnt dadurch seine Schüler, entsprechend seiner Erwartungen, unterschiedlich zu behandeln.
3. Durch die unterschiedliche Umgangsweise seitens des Lehrers reagieren auch die Schüler unterschiedlich auf den Lehrer.
4. Die Schüler reagieren eher mit einem Verhalten, das die Erwartungen des Lehrers noch verstärkt.
5. Das Ergebnis ist dann, dass sich die allgemeine Schulleistung mancher Kinder verbessert und die anderer hingegen verschlechtert, wobei die Richtung der Leistungsveränderung der Richtung der Lehrererwartung entspricht.
6. Diese Effekte zeigen sich dann in den Leistungstests am Jahresende und sie unterstützen die Auffassung der SFP.

Nur wenn alle Elemente dieses Modells vorhanden sind kann ein Lehrererwartungseffekt auftreten (Brophy & Good, 1970, S. 365).

### 4 Erwartungseffekt in der Schule

In diesem Kapitel sollen sowohl die Anfänge der Untersuchungen zum Erwartungseffekt in der Schule dargestellt, als auch der aktuelle Stand der Forschung zu diesem Thema näher beleuchtet werden. Es soll in groben Zügen gezeigt werden, was sich auf diesem Forschungsgebiet in den letzten 40 Jahren verändert hat und auf welchem Wissensstand wir uns heute befinden.

Die erste umfangreiche Studie zum Thema Erwartungseffekt in der Schule wurde, wie schon weiter oben erwähnt, 1964 von Rosenthal und Jacobson durchgeführt, daher wird diese im Folgenden kurz dargestellt.

#### 4.1 Die Anfänge des Erwartungseffekts im Unterricht

Rosenthal und Jacobson haben 1964 zum ersten Mal den Erwartungseffekt in einer Volksschule südlich von San Francisco (USA) untersucht. Die Autoren gingen der Frage nach, ob sich die positiven Erwartungen der Lehrer bezüglich der zukünftigen Intelligenzentwicklung ihrer Schüler auch tatsächlich positiv auf den IQ der Kinder auswirkten. Die Volksschule, in der sie ihre Untersuchung durchführten, nannten Rosenthal und Jacobson „Oak School“, um die Anonymität der 20 Lehrer und 650 Schüler zu wahren. Die Schüler der „Oak School“ stammten zum Großteil aus Familien die der Unterschicht angehörten und ca. 30% der Kinder stammten aus mexikanischen Familien (Rosenthal & Jacobson, 1974).

Die Kinder der „Oak School“ wurden v.a. anhand ihrer Leseleistungen, die von ihren Lehrern beurteilt wurden, und den Ergebnissen eines Leistungstests in drei unterschiedliche Leistungsklassen/Leistungszüge eingeteilt. Jede Klassenstufe von 1 bis 6 hatte also drei Leistungsklassen: den schnellen Zug (sehr gute Leser), den mittleren und den langsamen Zug (schwache Leser). Es war auch möglich in eine leistungshöhere oder -niedrigere Gruppe überwiesen zu werden. Die Lehrer gaben am Ende jedes Schuljahres an, ob ein Kind auf- oder abgestuft werden oder im selben Leistungszug bleiben sollte (Rosenthal & Jacobson, 1974).

Die erste Erhebung der Studie fand im Frühjahr 1964 statt. Alle Kinder, die voraussichtlich im darauf folgenden Herbst weiterhin die „Oak School“ besuchen würden, nahmen an dem „Harvard Test der flektierten Aneignung“ teil. Es handelte sich dabei aber in Wirklichkeit um einen standardisierten, überwiegend nonverbalen Intelligenztest (Test of General Ability - TOGA) von Flanagan (1960, zitiert nach Rosenthal & Jacobson, 1974). Der TOGA setzte sich aus zwei voneinander relativ unabhängigen Subtests zusammen, wovon der eine die verbalen Fähigkeiten und der andere das schlussfolgernde Denken (Reasoning) erfasste. Den Lehrern teilte man jedoch mit, dass man mit Hilfe dieses Tests ein geistiges „Aufblühen“ oder eine gesteigerte „Schnellentwicklung“ vorhersagen könne. Insgesamt gab es vier Testdurchführungen: Im Mai 1964 (Vortest), Januar 1965 (Wiederholungstest), Mai 1965 (Nachtest) und im Mai 1966 (Abschlusstest); wobei die Lehrer nur über die ersten drei Testzeitpunkte informiert wurden. Mit Hilfe der Differenz zwischen dem Vortest-IQ und dem Nachtest-IQ wurde beurteilt, ob sich bei einem Kind eine Intelligenzsteigerung gezeigt hatte. Die Hypothese von Rosenthal und Jacobson war, dass jene Kinder, hinsichtlich derer die Lehrer besonders positive Erwartungen hegten, aufgrund der Erwartungsmanipulation eine größere Intelligenzsteigerung zeigen würden als die Kontrollgruppenkinder (jene Kinder für die es keine Erwartungsmanipulation gab). Die Erwartungsmanipulation wurde am Ende des Sommers 1964 vorgenommen. Insgesamt wurden 20% der Schüler, welche angeblich die besten Testergebnisse im „Harvard Test der flektierten Aneignung“ zeigten, zu „Schnellentwicklern“ erklärt und jeder Lehrer bekam die Namensliste der Schüler, die mit Hilfe des Vortests als „Aufblüher“ identifiziert wurden. Die Lehrer wurden ersucht die Ergebnisse nicht mit ihren Schülern und deren Eltern zu besprechen. Die „besonderen“ Kinder wurden nach dem Zufallsprinzip ausgewählt und hatten somit rein gar nichts mit den realen Testergebnissen zu tun. Aus ethischen Gründen wurde auf eine negative Erwartungsmanipulation verzichtet (Rosenthal & Jacobson, 1974).

Die wichtigsten Ergebnisse des Experiments, die Rosenthal und Jacobson in ihrer Untersuchung fanden, lassen sich in ein paar Sätzen zusammenfassen:

## Erwartungseffekt in der Schule

Es zeigte sich ein signifikanter Unterschied zwischen den Kindern die als „Aufblüher“ bezeichnet wurden und den Kontrollgruppenkindern hinsichtlich der Intelligenzentwicklung in dem Jahr, in dem das Experiment durchgeführt wurde. Wenn man jedoch jede Klassenstufe separat betrachtet, zeigt sich, dass die Gruppe der „Aufblüher“ nur in der 1. und 2. Klasse einen signifikanten Erwartungsvorteil hatten. In den anderen Klassenstufen gab es keinen signifikanten Unterschied zwischen Kontroll- und Versuchsgruppe.

Obwohl sich nur bei zwei Klassenstufen ein signifikanter Unterschied zwischen der Experimentalgruppe und der Kontrollgruppe zeigte, meinten die Autoren ganz allgemein, dass ihre Ergebnisse einen weiteren Beweis dafür liefern würden, dass die Erwartungen, die eine Person einer anderen gegenüber hat, zu einer SFP führen können. Dafür, dass sich nur in den ersten beiden Klassenstufen, d.h. nur bei 19 Schülern, die Intelligenzentwicklung der „Aufblüher“ signifikant von der Kontrollgruppe unterschied, hatten Rosenthal und Jacobson mehrere Erklärungsansätze. Eine Erklärungsmöglichkeit lieferte ihnen Scott, der im Allgemeinen davon ausgeht, dass jüngere Kinder formbarer, weniger festgelegt und dadurch beeinflussbarer sind (1962, zitiert nach Rosenthal & Jacobson, 1974). Eine andere Möglichkeit dieses Ergebnis zu interpretieren war, dass jüngere Schüler sich in der Schule noch nicht so etabliert haben wie die Älteren und es daher einfacher ist die Erwartungen der Lehrer den jüngeren Schülern gegenüber zu manipulieren (Rosenthal & Jacobson, 1974). Bei Raudenbush (1984) zeigten sich die stärksten SFPs bei Schülern der 1., 2. und 7. Klassenstufe. Dieses Ergebnis passt zu der Überlegung von Rosenthal und Jacobson, nämlich, dass jene Kinder formbarer sind, die noch jünger sind und sich daher in der Schule noch nicht so stark etabliert haben. Raudenbush (1984) berücksichtigte bei seiner Studie auch Kinder die bereits die High School besuchen und dabei zeigte sich, dass auch in der High School die „Jüngsten“, d.h. die Kinder der 7. Klasse, am stärksten von den Erwartungen der Lehrer beeinflusst werden. Auch Jussim, Eccles und Madon (1996) nehmen an, dass Menschen dann anfälliger für SFPs sind, wenn sie in neue, unbekannte Situationen kommen. Diese Überlegung

stimmt auch mit den Befunden von McNatt (2000, zitiert nach Jussim & Harber, 2005) überein.

Die Studie von Rosenthal und Jacobson zum Pygmalioneffekt weckte, wie auch schon weiter oben erwähnt, das Interesse vieler Wissenschaftler. Die Autoren Brophy und Good z.B. sprangen schnell auf den Zug dieses Wissenschaftsgebiets auf. Sie fanden in ihrer Untersuchung im Jahr 1970 heraus, dass Schüler der 1. Klasse Grundschule mehrere Verhaltensweisen der Lehrer als Hinweise dafür sahen, dass diese sie im Vergleich zu ihren Mitschülern als weniger fähig einstufen. Die Schüler sahen sich beispielsweise als schlechtere Schüler, wenn sie vom Lehrer weniger Antwortmöglichkeiten erhielten und häufiger für falsche Antworten kritisiert wurden. Weitere Indizien dafür, dass sie von den Lehrern als weniger begabt gesehen wurden, waren Verhaltensweisen wie z.B., dass sie für richtige Antworten ihrer Meinung nach seltener gelobt wurden und sich Lehrer schneller von ihnen abwandten, wenn sie die Antwort auf eine Frage nicht wussten oder eine falsche Antwort gaben. Passend zu diesem Befund zeigte sich in der Studie von Good, dass Lehrer der 1. Klasse Volksschule jenen Kindern mehr Antwortmöglichkeiten einräumten, die sie als begabter ansahen (1969, zitiert nach Good & Nichols, 2001).

Das „Oak School“ Experiment wurde von vielen Seiten bewundert und regte zu etlichen weiteren Untersuchungen auf diesem Forschungsgebiet an, es wurde jedoch auch mit kritischen Augen betrachtet.

### **4.2 Schwächen des „Oak School“ Experiments**

Elashoff und Snow (1972) haben sich sehr intensiv mit der Untersuchung von Rosenthal und Jacobson auseinandergesetzt und sind, wie auch andere Wissenschaftler, auf einige methodologische Mängel gestoßen, wodurch die Glaubwürdigkeit der Ergebnisse in Frage gestellt werden kann.

Einige Forscher bemängelten z.B. die Reliabilität des verwendeten Intelligenztests und meinten, dass dadurch die Ergebnisse der Studie bedeutungslos seien (z.B. Roth, 1995; Thorndike, 1968, zitiert nach Jussim & Harber, 2005). Andere Autoren

hingegen meinten, dass der signifikante Unterschied zwischen Kontrollgruppe und Versuchsgruppe in den ersten beiden Klassenstufen durch die mangelnde Reliabilität des Erhebungsinstruments eher noch verstärkt wird, da es schwer ist, mit einem Messinstrument mit geringer Reliabilität, Unterschiede zwischen zwei Gruppen zu finden; dennoch waren diese bei der Untersuchung von Rosenthal und Jacobson zu sehen (Jussim & Harber, 2005).

Ein weiterer Kritikpunkt, angeführt von Elashoff und Snow (1972), bezieht sich auf das unterschiedliche Ausgangsniveau der Versuchs- und Kontrollgruppe hinsichtlich der IQ-Werte im Vortest, wodurch die beiden Gruppen, laut Autoren, nicht vergleichbar sind. Außerdem kritisierten die beiden, dass der Erfolg der Erwartungsmanipulation zu hinterfragen sei, da nicht berücksichtigt wurde, dass die Namen der „Aufblüher“ von den Lehrern teilweise vergessen wurden. Rosenthal und Jacobson (1974) äußerten in ihren Darstellungen zwar den Verdacht, dass die Lehrer die Namen der Kinder die als „Aufblüher“ definiert wurden teilweise vergessen hatten. Sie gingen in weiterer Folge jedoch nicht darauf ein, was das für Auswirkungen auf die Interpretation der Ergebnisse haben würde, falls die Erwartungsmanipulation nur teilweise gegriffen hätte. Der Pygmalioneffekt beruht darauf, dass die Erwartungen der Lehrer einen Einfluss auf die zukünftigen Leistungen der Schüler haben. Wenn die Lehrer nun aber die Namen einiger Schüler nicht mehr wissen, die in den Untersuchungen von Rosenthal und Jacobson als „Aufblüher“ ausgewiesen wurden, kann man nicht sagen, welche Erwartungen die Lehrer diesen Kindern gegenüber hatten, da die subjektiven Lehrererwartungen nicht erhoben wurden. All das macht es schwierig, die Ergebnisse ausschließlich auf die Erwartungsmanipulation zurückzuführen, da nicht bekannt ist, welche Erwartungen wie lange wirksam waren. Eine Möglichkeit dieses Problem zu umgehen wäre gewesen, die Erwartungsmanipulation im Laufe der Untersuchung immer wieder zu überprüfen. Laut Fleming und Anttonen (1971, zitiert nach Berka & Westhoff, 1981, S. 9) wird nur selten überprüft, ob eine Erwartungsinduktion auch wirklich gelungen ist. Bezüglich der Auswirkungen eines fraglichen Manipulationserfolges auf die Ergebnisse solcher Studien, bei denen die Lehrererwartung mit Hilfe von Informationen über die künftige Entwicklung der

Schüler beeinflusst werden soll, geht Ludwig (1994) in seiner Arbeit sogar so weit zu meinen, dass der fragliche Erfolg einer Erwartungsmanipulation auch ein Grund dafür sei, dass bei Metaanalysen nur 40% der erfassten Studien zu diesem Thema signifikante Unterschiede aufwiesen. Das Problem ist laut Ludwig (1994) auch die Sicherstellung des Manipulationserfolges, da es, ihm zu Folge, keine Verfahren gibt mit denen man den Erfolg einer Erwartungsmanipulation zweifellos feststellen kann. Auch für eine nachträgliche Überprüfung des Ausmaßes des Manipulationserfolges gibt es, seiner Meinung nach, keine verlässlichen Messinstrumente (Ludwig, 1994, S. 101). Um dieses Problem, des fraglichen Manipulationserfolges der Lehrererwartungen, auszuschalten, kamen bald auch Untersuchungen auf, die mit den tatsächlichen Erwartungen der Lehrer den Rosenthal-Effekt untersuchten (Palardy, 1969; Seaver, 1971, zitiert nach Elashoff & Snow, 1972).

Auch heute wird bei vielen Felduntersuchungen mit den persönlich Erwartungen der Lehrer ihren Schülern gegenüber gearbeitet (z.B. Rubie-Davies et al., 2006; Good & Nichols, 2001; Hinnant, O'Brien & Ghazarian, 2009). Demzufolge verwendet auch die vorliegende Untersuchung die individuellen Lehrererwartungen.

Die Untersuchung von Rosenthal und Jacobson wurde von Elashoff und Snow (1972) und auch von anderen Autoren (z.B. Thorndike, 1972; Berka & Westhoff, 1981) noch viel ausführlicher kritisiert. Darauf soll an dieser Stelle jedoch nicht weiter eingegangen werden, da dies für die vorliegende Arbeit nicht von Bedeutung ist.

Trotz zahlreicher Kritik am „Oak School“ Experiment von Rosenthal und Jacobson sollte nicht übersehen werden, dass diese Untersuchung den Startschuss für viele weitere Studien auf diesem Gebiet gab und sich bei den ersten beiden Klassenstufen ein signifikanter Unterschied zwischen der Versuchsgruppe und der Kontrollgruppe zeigte, obwohl sich dieses Ergebnis nur auf 19 Versuchspersonen stützt. Dieses Ergebnis lässt sich, trotz aller Untersuchungsmängel, als Hinweis dafür sehen, dass die positiven Erwartungen von Lehrern ihren Schülern

gegenüber einen positiven Einfluss auf deren zukünftige schulische Leistung haben, auch wenn dieser Effekt sich nicht bei allen Klassenstufen gezeigt hat. Außerdem machen gerade die Kritikpunkte an einer Untersuchung weitere Untersuchungen notwendig und geben Hinweise auf offen gebliebene Fragen. Wenn Rosenthal und Jacobson das Phänomen des Pygmalioneffekts mit ihrer ersten Studie bereits zur Gänze erklärt hätten, was bei so einem komplexen Untersuchungsgegenstand unmöglich ist, wären weitere Untersuchungen zu diesem Thema gegenstandslos bzw. keine Bereicherung mehr gewesen. Im nächsten Abschnitt werden einige nachfolgende Forschungsarbeiten zu diesem Thema näher beleuchtet, um die Entwicklung in diesem Bereich und den heutigen Forschungsstand aufzuzeigen.

### 4.3 Aktueller Forschungsstand zum Erwartungseffekt in der Schule

Jussim und Eccles (1995) und Jussim und Harber (2005) zeigen in ihren Arbeiten zum einen, dass der Effekt der SFP in experimentellen Studien schon mehrfach nachgewiesen werden konnte, zum anderen, dass auch in nicht-experimentellen Untersuchungen Belege für den Lehrererwartungseffekt zu finden sind, auch wenn diese eher klein bis moderat sind.

Sowohl in Reviews (z.B. Jussim & Eccles, 1995; Jussim & Harber, 2005) als auch in Metaanalysen (Raudenbush, 1984; Rosenthal & Rubin, 1978) konnte die Existenz des Lehrererwartungseffekts nachgewiesen werden. Trotzdem zeigen einzelne Arbeiten zu diesem Thema teilweise auch sehr widersprüchliche Ergebnisse auf. Zum einen gibt es Studien die zeigen, dass Lehrererwartungen einen sehr starken Einfluss auf Schüler haben (z.B. Gilbert, 1995). Zum anderen gibt es auch Literatur zu diesem Thema, welche die Existenz des Lehrererwartungseffekts rigoros bestreitet (z.B. Roth, 1995; Rowe, 1995, zitiert nach Jussim & Harber, 2005). Darüber hinaus finden sich auch noch Arbeiten, die keinen der soeben genannten extremen Standpunkte einnehmen. Sie sprechen zwar von der Existenz des Lehrererwartungseffekts, finden jedoch nur minimale Effekte (z.B. Spitz, 1999).

Im nächsten Kapitel soll der aktuelle Forschungsstand in diesem Bereich skizziert werden. Da die vorliegende Arbeit eine Felduntersuchung ist, werden im Folgenden v.a. Ergebnisse vergleichbarer Studien näher betrachtet.

### 4.3.1 Erwartungseffekt und „benachteiligte“ Schülergruppen

Eine sehr aktuelle Felduntersuchung wurde von Hinnant et al. (2009) durchgeführt. Die Autoren untersuchten in ihrer Studie, welche Auswirkungen das Über- und Unterschätzen der schulischen Leistungen von Grundschulern im Lesen und in Mathematik auf deren zukünftiges Leistungsniveau (Untersuchungszeitraum 1. - 5. Klasse Grundschule) hat. Neben den Lehrereinschätzungen und den tatsächlichen Lese- und Mathematikleistungen berücksichtigten Hinnant et al. (2009) in ihrer Arbeit auch verschiedene Schülercharakteristika. Sie untersuchten zum einen, ob diese als Moderatorvariablen zwischen Lehrereinschätzungen und späteren Schülerleistungen eine Rolle spielen und zum anderen, ob sie einen Einfluss auf Lehrereinschätzungen haben, worauf aber im folgenden Kapitel noch genauer eingegangen werden wird.

Die Ergebnisse von Hinnant et al. (2009) zeigten, dass Lehrererwartungen einen größeren Einfluss auf bestimmte Schülergruppen haben. Im Lesen war nur bei Buben, die einer ethnischen Minderheit angehören, ein schwacher positiver Lehrererwartungseffekt zu sehen. Dies bedeutet, dass nur bei diesen Kindern die Lehrereinschätzung in der 1. Klasse einen leichten Einfluss auf die zukünftigen Leseleistungen hatte. Auch Jussim und Harber (2005) finden in ihrem Review Hinweise dafür, dass Lehrereinschätzungen dann stärker mit späteren Leistungen ihrer Schüler in Verbindung stehen, wenn die Schüler von ihren Lehrern als eher benachteiligt oder als Außenseiter in der Klasse gesehen werden. Ähnliche Ergebnisse zeigen sich bei Jussim et al. (1996). Lehrererwartungen haben nur bei Schülern mit einem niedrigen sozioökonomischen Status einen Einfluss auf deren zukünftige Leistungen und zwar mit einer Effektstärke von .2 - .3, welche einer geringen bis mittleren Effektstärke entspricht. Wenn jene Schüler noch dazu eher schlechte Leistungen in der Schule erbringen, zeigt sich sogar eine Effektstärke von .4 - .6, welche einer mittleren bis hohen Effektstärke entspricht. Auch bei der

## Erwartungseffekt in der Schule

Untersuchung von Madon, Jussim und Eccles (1997) zeigte sich, dass durch Lehrererwartungen die Leistungen schwacher Schüler besser vorhergesagt werden können als die guter Schüler. Das heißt, auch in diesen Untersuchungen ist zu sehen, dass Lehrererwartungen wirkungsvoller für benachteiligte Schüler sind.

Bei Rubie-Davies et al. (2006) zeigte sich hingegen diesbezüglich kein eindeutiges Ergebnis. Die Autoren untersuchten in ihrer Arbeit, ob die ethnische Herkunft eines Schülers als Mediatorvariable auf den Zusammenhang zwischen Lehrererwartung und Schülerleistung wirkt. Sie fanden in ihrer Studie einen positiven Erwartungseffekt bezüglich der Leseleistung von neuseeländischen Schülern europäischer Abstammung, welche die vorherrschende ethnische Gruppe bei dieser Untersuchung darstellten, als auch bei asiatischen Schülern, welche eine der drei Minderheitengruppen in dieser Untersuchung waren. Auch Kinder von den pazifischen Inseln, die ebenfalls eine ethnische Minderheit in Neuseeland sind, wurden am Anfang des Schuljahrs in ihren Leseleistungen überschätzt, hatten am Jahresende jedoch keine signifikant besseren Leseleistungen.

Hinnant et al. (2009) zeigten bezüglich Rechnen, dass die Lehrereinschätzungen der 1. Klasse einen Einfluss auf die Schülerleistungen in der 3. Klasse haben, jedoch nur bei Kindern aus Familien mit niedrigem oder durchschnittlichem Einkommen. Die Mathematikleistungen der 5. Klasse konnten hingegen direkt durch die Lehrereinschätzungen der 1. und 3. Klasse vorhergesagt werden, unabhängig vom Einkommen der Eltern. Das bedeutet, wenn ein Lehrer der 1. und 3. Grundschulklasse positive Erwartungen in die Leistungsfähigkeit eines Schülers hatte, zeigte dieser in der 5. Klasse eher bessere Mathematikleistungen und umgekehrt.

In der Arbeit von Hinnant et al. (2009) wurde außerdem untersucht, ob positive oder negative Erwartungen wirkungsvoller sind. Im Folgenden soll auf Ergebnisse dieses Forschungsbereichs näher eingegangen werden.

### 4.3.2 Galatea- oder Golem-Effekt?

Dieser Teilbereich des Wissenschaftsgebiets um den Pygmalioneffekt hat große Praxisrelevanz. Wenn negative Lehrererwartungen wirksamer wären als positive, ist es umso wichtiger, dass Lehrer sich ihrer Erwartungen, ihren Schülern gegenüber, im Klaren sind und ihnen deren Auswirkungen bewusst sind. Jedoch auch in diesem speziellen Bereich der Erwartungseffektforschung finden sich gegensätzliche Ergebnisse. So zeigten sich z.B. bei Madon et al. (1997) Hinweise dafür, dass das Überschätzen von Schülern der 6. Klassenstufe einen signifikant stärkeren Einfluss auf die zukünftigen Leistungen hat als das Unterschätzen von Schülern. Dieses Ergebnis könnte ein Indiz dafür sein, dass positive Erwartungen sich stärker auf schulische Leistungen auswirken als negative, was für die höhere Wirksamkeit des Galatea-Effekts sprechen würde. Bei anderen Autoren hingegen zeigten sich Ergebnisse, die für eine stärkere Wirksamkeit von negativen Lehrererwartungen sprechen (z.B. Sutherland & Goldschmid, 1974; Alvidrez & Weinstein, 1999). Bei genauerer Betrachtung der Befunde von Sutherland und Goldschmid (1974) zeigte sich jedoch, dass die negativen Erwartungen in dieser Untersuchung extremer waren als die positiven, was bedeutet, dass die negativen Erwartungen weiter von den tatsächlichen Leistungen der Schüler entfernt waren als die positiven. Daher ist nicht klar, ob das Ergebnis, zu welchem die Autoren in ihrer Arbeit kommen, nur bedeutet, dass je weiter die Erwartungen von den tatsächlichen Leistungen entfernt sind es eher zu einer SFP kommt oder ob es für die größere Wirksamkeit von negativen Erwartungen spricht. Madon et al. (1997) achteten bei ihrer Untersuchung hingegen darauf, dass das Ausmaß der Überschätzung vergleichbar mit dem der Unterschätzung ist, um herauszufinden, ob positive oder negative Lehrererwartungen wirkungsvoller sind. Wie oben schon erwähnt, zeigte sich der gegenteilige Effekt, verglichen mit den Ergebnissen von Sutherland und Goldschmid (1974) und Alvidrez und Weinstein (1999). Hinnant et al. (2009) hingegen konnten weder für den Galatea- noch für den Golem-Effekt eine höhere Einflusskraft nachweisen. Auch Babad et al. (1982) fanden diesbezüglich keine eindeutigen Ergebnisse, es zeigten sich sowohl Belege für positive als auch für negative SFPs.

### 4.3.3 Gibt es differenziertes Lehrerverhalten gegenüber Schülern?

Im Zusammenhang mit der Erforschung des Lehrererwartungseffekts wurde immer wieder untersucht, wie sich Lehrer gegenüber guten und schlechten Schülern oder gegenüber Schülern für die sie hohe oder niedrige Erwartungen haben verhalten und ob die Schüler ein unterschiedliches Verhalten ihrer Lehrer wahrnehmen.

Weinstein und Middlestadt (1979) zeigten diesbezüglich in ihrer Untersuchung, dass Schüler sehr bewusst unterschiedliche Behandlungsweisen seitens der Lehrer wahrnehmen. Die Schüler gaben an, dass schwächere Kollegen weniger Hilfe bei selbstständig, am Platz zu erledigenden Aufgaben erhielten und die Lehrer auch die Arbeiten oft absammelten bevor die Schüler damit fertig waren. Brattesani, Weinstein und Marshall (1984) fanden in ihrer Studie mit Hilfe von Schülerberichten Anzeichen dafür, dass es in manchen Klassen eine stärkere Verhaltensdifferenzierung seitens der Lehrer gibt. In einigen Klassen berichteten Schüler, in die die Lehrer hohe Erwartungen hatten, von anderen Verhaltensweisen ihnen gegenüber als jene Schüler, in die die Lehrer niedrige Erwartungen hatten. Die Autoren zeigten, dass in jenen Klassen, von denen die Schüler angaben, dass eine höhere Verhaltensdifferenzierung wahrzunehmen sei, die Erwartungen seitens der Lehrer eine stärkere Auswirkung auf die schulische Leistung am Jahresende hatten als in jenen Klassen mit geringerer Verhaltensdifferenziertheit seitens der Lehrer. Die Effektstärke für die Vorhersagekraft von Schülerleistungen durch Lehrererwartungen betrug in Klassen, in denen die Schüler kaum Verhaltensunterschiede wahrnahmen, zwischen 0 - .1, hingegen zeigten sich in Klassen bei denen die Schüler eine hohe Verhaltensdifferenziertheit angaben Effektstärken von .3 - .4 (mittlere Effektstärken) (Brattesani et al., 1984). Diese Befunde stimmen mit dem Modell von Brophy und Good (1970) überein, welches weiter oben schon beschrieben wurde. Laut ihrem Modell ist eine Voraussetzung für das Auftreten eines Lehrererwartungseffekts, dass die Lehrer ihre Schüler, entsprechend ihrer Erwartungen, unterschiedlich behandeln. Auch Jussim und Harber (2005) stellen Überlegungen dahingehend an, dass womöglich gerade aufgrund dieser

unterschiedlichen Verhaltensweisen seitens der Lehrer ihren Schülern gegenüber, für die sie hohe oder niedrige Erwartungen hegen, SFPs auftreten.

Ergänzend dazu wurde in mehreren Studien untersucht, in welcher Form Lehrer sich ihren Schülern gegenüber, für die sie hohe oder niedrige Erwartungen haben, unterschiedlich verhalten. Lehrer zeigen jenen Kindern gegenüber mehr emotionale Wärme und unterstützen diese auch mehr, in die sie hohe Erwartungen haben. Außerdem bekommen diese Kinder meist positiveres und klareres Feedback von ihren Lehrern und haben häufiger die Möglichkeit sich unter Beweis zu stellen, da ihnen mehr und schwerere Inhalte beigebracht werden als ihren Kollegen, an welche die Lehrer nicht so hohe Erwartungen stellen (Brophy, 1983; Jussim, 1986; Rosenthal, 1974; Harris & Rosenthal, 1985, zitiert nach Jussim & Harber, 2005). Wie schon weiter oben erwähnt fand Good (1969, zitiert nach Good & Nichols, 2001) passend dazu heraus, dass Lehrer jenen Kindern der 1. Klasse Grundschule mehr Antwortmöglichkeiten einräumen, die von ihnen als fähiger angesehen werden, und Brophy und Good (1970) zeigten Hinweise dafür, dass Kinder der 1. Klasse sich selbst als schwächere Schüler einschätzten, wenn sie vom Lehrer weniger Antwortmöglichkeiten bekamen, bei falschen Antworten häufiger kritisiert und bei richtigen seltener gelobt wurden. Ebenfalls war zu sehen, dass diese Kinder häufiger „fallen gelassen“ wurden wenn sie auf eine Frage nicht antworten konnten oder eine falsche Antwort gaben.

Auch Eder (1981) konnte in seiner Studie differenzierte Verhaltensweisen seitens der Lehrer gegenüber schlechteren und besseren Schülern nachweisen. Lehrer verbringen, laut seinen Ergebnissen, weniger Zeit mit erzieherischen/lehrreichen Aktivitäten bei der schwächeren Lesegruppe als bei der besseren. Die Leistungsunterschiede der Schüler sind laut Eder größtenteils auf diese Instruktionsunterschiede zurückzuführen. Dadurch wurden Schüler, die der niedrigeren Lesegruppe zugeordnet wurden, mit der Zeit auch schlechtere Leser als Schüler mit ähnlichen Leseleistungen, die jedoch in die bessere Lesegruppe kamen, da die Anleitungen weniger unterstützend und aussagekräftig waren. Auch Weinstein (1976) zeigte in ihrer Arbeit, dass durch das Zuordnen der Schüler zu einer besseren oder schlechteren Lesegruppe innerhalb einer Klasse seitens der

## Erwartungseffekt in der Schule

Lehrer, die Leseleistungen besser vorhersagbar sind. Jene Schüler die von ihren Lehrern als bessere Leser beurteilt wurden hatten später auch bessere Leseleistungen und das Leseniveau der Kinder die als schlechtere Leser eingestuft wurden verschlechterte sich auch tatsächlich.

Durch unterschiedliche Behandlung der Schüler seitens der Lehrer kommt es laut dem Modell von Brophy und Good (1970) auch zu verschiedenen Reaktionen von Seiten der Schüler. Den Überlegungen der Autoren zu Folge verstärken die Reaktionen der Kinder die Erwartungen der Lehrer. Brophy und Good (1970) überlegten diesbezüglich in ihrer Arbeit, ob vielleicht manche Schüler gegenüber Lehrererwartungen widerstandsfähiger sind als andere. Auch Weinstein und McKown (1998) zeigten in ihrer Arbeit, dass manche Schüler durch die Erwartungen ihrer Lehrer leichter zu beeinflussen sind als andere. Ergebnisse, die schon weiter oben erwähnt wurden, wie z.B., dass Erwartungseffekte eher bei leistungsschwachen Schülern oder bei Schülern die zu verschiedenen Randgruppen gehören auftreten, sprechen dafür, dass manche Schüler leichter von ihren Lehrern zu beeinflussen sind als andere. Johnson (1970, zitiert nach Good & Nichols, 2001) meinte in seiner Arbeit, dass v.a. Schüler die sich stark an Erwachsenen orientieren, unselbstständig und daher generell eher fremdbestimmt sind besonders anfällig für Erwartungseffekte sind. Auch die Ergebnisse von Persell (1977, zitiert nach Good & Nichols, 2001) und West und Anderson (1976, zitiert nach Good & Nichols, 2001) deuteten darauf hin, dass abhängigere und jüngere Kinder eher durch die Erwartungen ihrer Lehrer beeinflussbar sind.

Wie einführend schon erwähnt, gibt es also eine Reihe von Studien die die Existenz des Lehrererwartungseffekts nachweisen (z.B. Jussim & Eccles, 1995; Jussim & Harber, 2005; Raudenbush, 1984; Brophy, 1983; Rosenthal & Rubin, 1978). Es zeigt sich in diesen Metaanalysen und Reviews, dass SFPs in der Schule zwar auftreten, dass diese durchschnittlich jedoch eher klein sind ( $r = .1 - .2$ ). Sowohl Brophy (1983) als auch Rosenthal (1984, zitiert nach Jussim & Harber, 2005) weisen in ihren Arbeiten darauf hin, dass Lehrererwartungen durchschnittlich nur bei 5 - 10% der Kinder einen Pygmalioneffekt bewirken. Jussim und Harber (2005) finden in ihrer Arbeit jedoch auch Belege dafür, dass es

neben den kleinen Effekten auch immer wieder zu großen Lehrer-erwartungseffekten kommt. Diese werden, wie bereits erwähnt, v.a. für Personen gefunden die einer Risikogruppe angehören, wie z.B. für schwache Schüler, Kinder aus niedrigeren sozioökonomischen Schichten oder für Schüler die einer ethnischen Minderheit angehören. Dann erreichen sie auch Effektstärken zwischen .4 - .6, welche einer mittleren bis hohen Effektstärke entspricht. Die Autoren stellen in ihrer Arbeit bezüglich der Frage, warum Lehrer-erwartungseffekte im Durchschnitt nicht besonders groß sind, auch die These auf, dass die Erwartungen der Lehrer generell eher genau sind, d.h. mit den tatsächlichen Leistungen der Schüler übereinstimmen. Wie schon weiter oben erwähnt, sind für das Auftreten einer SFP jedoch fehlerhafte Lehrer-erwartungen notwendig. Gelegentlich zeigt sich auch, dass Lehrer-erwartungen sehr ungenau sind, d.h., dass die Schüler von ihren Lehrern über- oder unterschätzt werden. Das nächste Kapitel beschäftigt sich daher damit, unter welchen Bedingungen Lehrer-erwartungen akkurat sind und welche Faktoren dazu führen können, dass Lehrer ihre Schüler hinsichtlich ihrer schulischen Leistungen falsch einschätzen.

### 5 Wie genau sind Lehrerurteile?

Wie eben schon erwähnt weisen einige Studien zum Thema Akkuratheit von Lehrerurteilen, der letzten 25 Jahre, darauf hin, dass Lehrer durchaus in der Lage sind ihre Schüler akkurat einzuschätzen (z.B. Hinnant et al., 2009; Brophy, 1983; Hoge & Butcher, 1984; Jussim & Eccles, 1992).

Egan und Archer (1985) gehen davon aus, dass Lehrer im Allgemeinen gute Diagnostiker sind und ihre Urteile kaum Verzerrungen unterliegen. Auch Schrader (1989) fand diesbezüglich Korrelationen zwischen Lehrereinschätzungen und der tatsächlichen Schülerleistung in der Höhe von .6 - .7. In der Metaanalyse von Hoge und Coladarci (1989), die insgesamt 16 Studien umfasst, zeigte sich ebenfalls eine durchschnittliche Korrelation von  $r = .66$  zwischen Lehrerurteil und Schülerleistung. Die Werte variierten jedoch zwischen .28 - .92. Diese Variabilität könnte ein Hinweis dafür sein, dass entweder zufällige Abweichungen oder auch systematische Faktoren manchmal vorherrschend sind und das Ergebnis somit verzerren (Bennett, Gottesman, Rock & Cerullo, 1993). Passend dazu finden sich in der Literatur auch immer wieder Belege dafür, dass Lehrer nicht in der Lage sind ihre Schüler akkurat zu beurteilen.

Obwohl Brophy (1983) zwar Hinweise für akkurate Lehrereinschätzungen in seiner Arbeit fand, war seine Hypothese, dass sich Lehrer oft mit nur wenig Information ein relativ stabiles Urteil über die Leistungen und das Verhalten ihrer Schüler bilden, welches jedoch oft wenig zutreffend ist. In einigen jüngeren Studien konnte nachgewiesen werden, dass Lehrer dazu neigen, die Leistungen ihrer Schüler zu unter- oder überschätzen (z.B. Bates & Nettlebeck, 2001; Feinberg & Shapiro, 2003). Auch ein Ergebnis der PISA Studie im Jahr 2001 deutet darauf hin, dass Lehrerurteile wenig akkurat sind. Es zeigte sich, dass Hauptschullehrer erheblich weniger Schüler identifizierten die aufgrund schlechter Lesefähigkeiten erhebliche Schwierigkeiten beim Übergang in das Berufsleben haben dürften, als aufgrund der PISA-Tests vorhergesagt wurde (Deutsches PISA-Konsortium, 2001, S. 119 f., zitiert nach Spinath, 2005, S. 86). Durch solche Befunde wird die Akkuratheit von Lehrerurteilen immer wieder in Frage gestellt.

## Wie genau sind Lehrerurteile?

Laut Spinath (2005, S. 86) kann ein Teil der widersprüchlichen Aussagen, bezüglich der Güte von Lehrerurteilen, auf den Einsatz unterschiedlicher Maße zurückgeführt werden, die zur Überprüfung der Genauigkeit von Lehrer-einschätzungen verwendet werden. Verschiedene Arten von Akkuratheitsmaßen weisen natürlich auch unterschiedliche Einschätzungsfehler auf. Schrader und Helmke (1987) unterscheiden diesbezüglich drei Akkuratheitsmaße /Urteilkomponenten. Dabei handelt es sich um die Differenzierungs-, die Rang- und die Niveauelemente. Die Differenzierungskomponente bezieht sich auf die Streuung der Merkmalsausprägungen der Schüler, die von den Lehrern eingeschätzt werden sollen; daher werden in diesem Fall Varianzen miteinander verglichen (Varianzkomponente). Mittels einer Rangkorrelation oder einer Produkt-Moment-Korrelation kann die Rangkomponente erhoben werden, d.h., wie genau Lehrer die Rangordnung der Schülerleistungen beurteilen können. Die Niveauelemente beschreibt hingegen die Tendenz der Lehrer, die Merkmalsausprägung der Klasse im Vergleich zu den tatsächlichen Schülermerkmalen zu über- oder zu unterschätzen. Um diese Komponente beurteilen zu können werden Mittelwerte miteinander verglichen (t-Test) (Schrader & Helmke, 1987; Spinath, 2005).

Um die Akkuratheit von Lehrerurteilen zu bewerten werden häufig Korrelationsstudien durchgeführt. Bei Studien dieser Art wird die Lehrer-einschätzung mit der tatsächlichen Leistung des Schülers, die dieser in einem Leistungstest (meist standardisiert) erzielt, korreliert. Hierbei ist zu bedenken, dass eine Rangkorrelation oder Produkt-Moment-Korrelation das absolute Niveau der Schülerleistungen vernachlässigt. Das heißt, dass in einer solchen Studie Lehrer auch dann als gute Diagnostiker erscheinen, wenn sie die Leistungen ihrer Schüler systematisch über- oder unterschätzen.

Eine andere Vorgehensweise, die immer wieder eingesetzt wird, um zu ermitteln wie gut Lehrer die Leistungen oder auch andere Merkmale ihrer Schüler einschätzen können, ist, dass das Ausmaß der mittleren Abweichung zwischen dem Lehrerurteil und der tatsächlichen Leistung des Schülers berichtet wird

## Wie genau sind Lehrerurteile?

(Spinath, 2005). Diese Methode wird in der vorliegenden Untersuchung verwendet, um die Akkuratheit der Lehrereinschätzungen beurteilen zu können.

Auch in Hinblick auf die Erhebungsmethode der Lehrerurteile unterscheiden sich die Studien die sich mit der Genauigkeit von Lehrereinschätzungen beschäftigen, was ebenfalls Grund für die teilweise sehr unterschiedlichen Ergebnisse sein könnte. Es werden sowohl direkte als auch indirekte Methoden verwendet. Erstere beziehen sich auf die Leistung und das Verhalten der Schüler im Schulalltag. Hierbei wird beispielsweise die Lesegeschwindigkeit oder das Leseverständnis erfasst, indem der Lehrer ein Leistungsurteil darüber abgibt, wie gut ein Schüler bei einem konkreten Leistungstest abschneiden wird. Indirekte Methoden zur Erfassung der Leistungen wären beispielsweise Rating-Skalen, bei denen Lehrer, unabhängig von einem bestimmten vorgegebenen Leistungstest, auf einer Skala beurteilen sollen, wie gut das Kind z.B. bereits gelernt hat Buchstaben zu erkennen. Außerdem kann zwischen peerunabhängigen und peerabhängigen Erhebungsmethoden unterschieden werden. Bei peerunabhängigen Methoden muss der Lehrer die Leistung eines Schülers ohne Berücksichtigung der Leistungen der anderen Schüler einschätzen (Begeny, Eckert, Montarello & Storie, 2008).

Die meisten Untersuchungen bezüglich der Genauigkeit von Lehrerurteilen beschäftigen sich mit Leistungseinschätzungen von Lehrern (Spinath, 2005). Es gibt jedoch auch Studien die sich mit der Einschätzung von Schülermerkmalen wie dem Selbstkonzept, der Leistungsängstlichkeit, Motivation, Fähigkeits-selbstwahrnehmung und einigen anderen beschäftigen (z.B. Spinath, 2005; Schrader & Helmke, 1990). Da für diese Untersuchung jedoch v.a. Ergebnisse von Studien die sich mit der Akkuratheit von Leistungseinschätzungen seitens der Lehrer beschäftigen relevant sind, wird im Folgenden v.a. auf Ergebnisse solcher Untersuchungen eingegangen.

Die Akkuratheit von Lehrerurteilen hinsichtlich der Lesefähigkeit ihrer Schüler wurde z.B. von Begeny et al. (2008) untersucht. Die Autoren verwendeten zur

## Wie genau sind Lehrerurteile?

Vorhersage der Lesefähigkeit sowohl indirekte als auch direkte Erhebungsmethoden. Außerdem verglichen sie die Lehrereinschätzung mit dem CBM - Curriculum-Based Measurement (Deno, 1985; Deno & Mirkin, 1977, zitiert nach Begeny et al., 2001). Es handelt sich dabei um ein Erhebungsverfahren, mit dem man direkt und peerunabhängig die Fähigkeiten der Schüler erfassen kann. Die Ergebnisse der Studie zeigen, dass sowohl die direkten als auch die indirekten Methoden eine moderate bis sogar hohe Korrelation mit der tatsächlichen Leistung aufzeigen. Bezogen auf bestimmte Schülergruppen zeigt sich jedoch, dass das Lehrerurteil nicht besonders akkurat ist. Obwohl die untersuchten Lehrer die pro Minute korrekt gelesenen Wörter von schwachen Lesern sehr gut einschätzen konnten, fiel es ihnen schwer schlechte Leser zu identifizieren, da sie diese im Allgemeinen eher überschätzten. Auch Feinberg und Shapiro (2003) verwendeten den CBM um die Akkuratheit der Lehrerurteile zu untersuchen. Bei den Autoren zeigte sich ebenfalls, dass Lehrer vor allem Schwierigkeiten haben schwache Schüler zu identifizieren, da sie dazu neigen deren Fähigkeiten zu überschätzen. Graney (2008) interessierte sich in ihrer Arbeit vor allem dafür, ob Lehrereinschätzungen bezüglich der Leistungsentwicklung von Schülern akkurat sind. In diesem Zusammenhang ist laut Graney (2008) zu bedenken, dass eine systematische Über- oder Unterschätzung der derzeitigen Leistung des Schülers nicht zwangsläufig bedeuten muss, dass der Lehrer den Progress des Schülers nicht richtig vorhersagen kann. Das heißt aber auch, dass ein Lehrer, der die derzeitige Leistung seines Schülers richtig vorhersagt, nicht zwangsläufig dessen zukünftige Verbesserung oder Verschlechterung richtig vorhersagen muss. Auch hier zeigte sich, dass die Urteile von Lehrern hinsichtlich leistungsschwacher Schüler nicht akkurat sind. Der Lernfortschritt der Schüler wurde überschätzt. Interessant war, dass Experten (Schulpsychologen), im Umgang mit dem CBM, durchaus zu einer akkuraten Einschätzung kamen. Bei Bates und Nettlebeck (2001), die in ihrer Arbeit unter anderem die Akkuratheit von Lehrerurteilen bezüglich der Leseleistung von 108 Schulkindern im Alter von 6-8 Jahren untersuchten, zeigte sich ebenfalls eine hoch signifikante Tendenz dahingehend, dass schwächere Leser von ihren Lehrern überschätzt, während gute Schüler in ihrer Lesekompetenz eher unterschätzt wurden. Bei einigen Studien war jedoch zu

## Wie genau sind Lehrerurteile?

sehen, dass die Leistungseinschätzungen von Lehrern sehr wohl akkurat sind (z.B. Demaray & Elliot, 1998; Hoge & Butcher, 1984). Auch in der Untersuchung von Feinberg und Shapiro (2003) zeigte sich bei der Korrelation von Lehrereinschätzungen der Leseflüssigkeit von Schülern der 3. - 5. Klasse und deren tatsächlicher Leseflüssigkeit, dass Lehrer die Leseflüssigkeit ihrer Schüler grundsätzlich akkurat einschätzen können. Daher ist zu sagen, dass das Lehrerurteil hinsichtlich der Leseflüssigkeit ein guter Prädiktor für die aktuelle Leseflüssigkeit eines Schülers ist. Bei genauerer Betrachtung der Ergebnisse ist jedoch zu sehen, dass es zwischen den Schulstufen große Unterschiede in den Effektstärken gibt (3. Klasse: 1.04; 4. Klasse: .49). Wie auch schon weiter oben erwähnt, ist die Korrelation von Lehrereinschätzung und tatsächlicher Schülerleistung ein eher ungenaues Akkuratheitsmaß, daher ist es sinnvoll auch die Effektstärken zu berücksichtigen. Aufgrund der Effektstärken zeigt sich also, dass Lehrer die Leseflüssigkeit von Viert- und Fünftklässlern besser einschätzen als jene der Drittklässler. Zu dem Ergebnis, dass Lehrer die Fähigkeiten älterer Schüler besser einzuschätzen vermögen, kommen auch Schrader und Helmke (1990) in ihrer Untersuchung. Bei Schrader und Helmke (1990) wurden insgesamt 690 Schüler von ihren 32 Lehrern in Bezug auf ihre Mathematikleistungen, ihre Intelligenz und ihre Fähigkeitsselbstwahrnehmung eingeschätzt. Die Autoren gingen in ihrer Arbeit unter anderem der Frage nach, ob Lehrerurteile in Hinblick auf schulische Leistungen, sowohl auf Klassen- als auch auf Schülerebene, genau sind und zeigen diesbezüglich Folgendes:

**Klassenebene:** Bei der Verwendung einer direkten Leistungseinschätzung, d.h. in diesem Fall, dass die Lehrer die Mathematikleistungen ihrer Schüler in einem Mathematiktest einschätzten, zeigte sich kein Zusammenhang zwischen dem generellen Urteilsniveau des Lehrers und dem tatsächlichen Leistungsstand der Klasse. Aufgrund dieses Ergebnisses ist davon auszugehen, dass die Lehrer nicht im Stande waren das absolute Leistungsniveau ihrer Klasse akkurat einzuschätzen. Laut den Autoren ist jedoch zu berücksichtigen, dass keine direkte Abfrage des Leistungsniveaus der Klasse stattgefunden hat, sondern von den klassenweise gemittelten Individualeinschätzungen nur indirekte Rückschlüsse auf die Einschätzung des Klassenniveaus gezogen wurden.

**Schülerebene:** Die Korrelationen zwischen direkter Leistungseinschätzung und Testleistung und indirekter Lehrerbeurteilung (Notengebung) und der Mathematiktestleistung unterscheiden sich kaum ( $r = .67$  und  $r = .70$ ). Dieses Ergebnis deutet darauf hin, dass die Benotung Ergebnis einer umfangreichen und fundierten Urteilsbildung ist, in einer dem Lehrer vertrauten Form. Die direkte Einschätzung ist hingegen ungewohnt, wodurch der Lehrer deren Vorteile (unmittelbare Korrespondenz zur Testleistung, Verfügbarkeit einer differenzierten Urteilsskala) nicht vollständig ausschöpfen kann (Schrader & Helmke, 1990, S. 321).

Die Autoren untersuchten in ihrer Studie jedoch auch, inwiefern noch andere Schülermerkmale, außer der zu beurteilenden Leistung, in die Lehrer-einschätzungen mit einbezogen werden. Auch der Einfluss von Moderatorvariablen auf das Lehrerurteil kann ein Grund für entsprechend heterogene Ergebnisse sein, wie sie in der Literatur zu finden sind. Daher soll im folgenden Kapitel auf die Frage eingegangen werden, welche Variablen, außer der schulischen Leistung, einen Einfluss auf die Leistungseinschätzungen von Lehrern haben und diese somit verzerren.

### 5.1 Welche Schülermerkmale moderieren Lehrerurteile?

Durch die Berücksichtigung leistungsunabhängiger Merkmale der Schüler bei der Beurteilung ihrer Leistungen kommt es zu systematischen Fehlern bei den Leistungseinschätzungen durch Lehrer (Schrader & Helmke, 1990). Das heißt, dass durch den Einfluss von leistungsfremden Variablen bei der Einschätzung von schulischen Leistungen seitens der Lehrer ein Schüler entweder systematisch über- oder unterschätzt werden kann. Dass dieses Über- oder Unterschätzen Auswirkungen auf die zukünftigen schulischen Leistungen der Schüler haben kann, wurde bereits weiter oben, im Zusammenhang mit dem Pygmalioneffekt, besprochen. Wenn durch das Überschätzen der schulischen Leistung eines Schülers dieser in seiner Leistung tatsächlich besser wird, ist das für alle Beteiligten wahrscheinlich nur von Vorteil, solange das Kind dadurch nicht unter enormen Druck gesetzt wird. Wenn das Kind jedoch unterschätzt wird und sich

## Wie genau sind Lehrerurteile?

das genauso auswirkt wie das Überschätzen eines Schülers, nur in die entgegengesetzte Richtung, ist das für die Schullaufbahn dieses Schülers von großer praktischer Relevanz, da dieser über die Jahre hinweg, in denen er unterschätzt wird, immer schlechter wird und dadurch seine positive Schulkarriere gefährdet werden könnte. Aus diesem Grund ist es von großer Bedeutung herauszufinden, welche Schülermerkmale, abgesehen von Leistungsvariablen, einen Einfluss auf das Lehrerurteil haben können und dieses somit verzerren. Es ist wichtig die Erkenntnisse in diesem Bereich auch den Lehrern zugänglich zu machen und ihnen damit eine Hilfestellung dahingehend anzubieten, besser darüber Bescheid zu wissen worauf sie bei der Leistungsbeurteilung ihrer Schüler achten oder besser gesagt welche Variablen sie bei dieser nicht beachten sollten, da diese die Lehrereinschätzung verfälschen könnten. Es zeigte sich bereits in mehreren Untersuchungen, dass z.B. Variablen wie das Verhalten der Schüler und damit in Verbindung stehende Variablen wie die Anstrengung, die Motivation, die Anpassbarkeit und die Compliance eines Schülers das Lehrerurteil beeinflussen können (Bennett et al., 1993). Auch Faktoren wie das Geschlecht, das Alter, die Intelligenz, das Leistungsniveau, das Selbstkonzept und einige andere Variablen können einen signifikanten Einfluss auf die Lehrereinschätzung haben (z.B. Schabmann & Schmidt, 2009; Schrader & Helmke 1990; Hecht & Greenfield, 2002; Bennet et al., 1993; Feinberg & Shapiro, 2003). Im Folgenden wird jedoch nur auf die Variablen Verhalten, Geschlecht, Leistungsniveau und Intelligenz eingegangen, da diese für die vorliegende Untersuchung von Bedeutung sind.

### 5.1.1 Verhalten

Mit der Thematik, inwiefern das Verhalten der Schüler bei der Leistungseinschätzung durch den Lehrer eine Rolle spielt, haben sich z.B. Schabmann und Schmidt (2009) in ihrer Arbeit beschäftigt. Die Autoren verwenden in ihrer Studie, unter anderem, Lehrereinschätzungen von 293 Schülern der 1. Klasse Volksschule bezüglich deren Verhalten, Lese- und Rechtschreibleistungen. Die Variable Verhalten setzt sich dabei aus 6 Verhaltensweisen (Hyperaktivität, oppositionelles/störendes Verhalten, Aggression, sozialer Rückzug, Depressivität

und Viktimisierung) zusammen. Das Verhalten, als auch die Leistungen der Schüler im Lesen und Rechtschreiben wurden zu Beginn der 1. Klasse (Nov.) von den Lehrern beurteilt. Bezüglich der Verhaltensvariablen Depressivität und Viktimisierung zeigte sich, dass diese keinen signifikanten Effekt auf die Lehrereinschätzungen im Lesen und Rechtschreiben hatten, daher wurden diese in den weiteren Analysen von den Autoren nicht berücksichtigt. Die Ergebnisse der Untersuchung weisen darauf hin, dass das Verhalten einen signifikanten Einfluss auf das Lehrerurteil hat. Dabei ist zu sehen, dass die Leistungseinschätzungen der Lehrer am besten durch die Verhaltensvariable Hyperaktivität vorhergesagt werden können. Auch die anderen Verhaltensweisen (störendes Verhalten, Aggression und sozialer Rückzug) beeinflussen das Lehrerurteil signifikant, jedoch nicht so stark wie die Variable Hyperaktivität. Hinsichtlich des Faktors störendes Verhalten zeigte sich ein besonders interessantes Ergebnis. Dieser hatte nämlich einen positiven Einfluss auf die Leistungseinschätzung der Lehrer im Rechtschreiben. Bei Bennett et al. (1993) war zu sehen, dass sich aus dem Verhalten der Kinder in der Klasse und aus ihrer getesteten schulischen Leistung die Einschätzungen ihrer Lehrer gut ableiten lassen. Außerdem zeigte sich, dass das Verhalten der Schüler das Lehrerurteil signifikant beeinflusste, wenn das Schülersgeschlecht und die erhobene schulische Leistung statistisch kontrolliert wurden. Die Ergebnisse der Autoren weisen darauf hin, dass jene Schüler, die von ihren Lehrern eine schlechte Betragensnote am Jahresende erhielten, auch in ihren schulischen Leistungen schlechter eingeschätzt wurden als ihre Kollegen mit besseren Verhaltensbeurteilungen, unabhängig von deren tatsächlichen Leistungen und deren Geschlecht. Tournaki und Podell (2005) untersuchten in ihrer Studie, welche Schülercharakteristika von fiktiven Schülern einen Einfluss auf das Lehrerurteil hinsichtlich des Schulerfolgs haben. Die Autoren gaben 384 Lehrern jeweils eine von 32 Vignetten vor, in denen sie das Geschlecht, die Leseleistung, das Sozialverhalten und die Aufmerksamkeit eines Schülers experimentell manipulierten. Es zeigte sich, dass Lehrer durchschnittlich gute Leser besser beurteilen wenn sie freundlich sind als wenn sie aggressiv sind. Bei einem schlechten Schüler im Lesen berücksichtigen Lehrer jedoch nicht das Verhalten des Schülers um dessen Leseleistung

## Wie genau sind Lehrerurteile?

einzuschätzen. Das heißt, dass Lehrer nur bei der Beurteilung von durchschnittlich guten Schülern deren Verhalten mit einbeziehen. Die schlechteste Schulerfolgsvorhersage bekamen Schüler die schlechter als der Durchschnitt lasen, aggressiv und unaufmerksam waren. Außerdem zeigte sich bei Tournaki und Podell (2005), dass durchschnittlich gute Leser die aggressiv sind besser beurteilt wurden als freundliche schlechte Leser, was ein Hinweis dafür ist, dass Lehrer sich hauptsächlich an der Leseleistung der Schüler orientieren, um diese zu beurteilen. Zu einem ähnlichen Ergebnis kamen Hecht und Greenfield (2002). Sie untersuchten in einer Längsschnittstudie diejenigen Faktoren, die es Lehrern ermöglicht die künftige Lesefähigkeit ihrer Schüler vorherzusagen. Die Stichprobe bestand aus 170 unterprivilegierten Schülern von der 1. bis zur 3. Klasse Grundschule aus insgesamt vier Schulen. Es war zu sehen, dass die Einschätzung der Lehrer dabei zu gleichen Teilen auf vorhandenen Fähigkeiten und Verhalten im Unterricht basierte. Außerdem zeigte sich bei der Einschätzung der Lesefähigkeit der Schüler der 1. Klasse, dass das Verhalten der Schüler das Lehrerurteil beeinflusst. Als Prädiktor der zukünftigen Lesefähigkeit zeigte sich der Faktor Verhalten jedoch redundant. Die Lesefähigkeit eines Schülers in der 1. Klasse konnte den größten Anteil an Varianz der späteren Lesefähigkeit erklären, d.h., dass diese das Lehrerurteil am stärksten beeinflusste.

### 5.1.2 Geschlecht

Aus älteren Studien, wie bei Reynolds (1989), geht hervor, dass es geschlechtsspezifische Unterschiede bei der Lesefähigkeit von Schülern gibt. Beobachtet wurde, dass Mädchen im Durchschnitt bei Tests zur Lesefähigkeit besser abschneiden als Buben. Auch eine aktuellere schwedische Studie von Nycander (2006, zitiert nach Lekholm & Cliffordson, 2008, S. 184) hat gezeigt, dass Mädchen bessere Schulnoten bekommen als Buben, im Verhältnis zu ihren Leistungen bei nationalen Tests. Demnach wird vermutet, dass das Geschlecht auch einen Einfluss auf das Lehrerurteil hat, nämlich, dass Mädchen tendenziell besser beurteilt werden als Buben. Auch in jüngeren Studien wurde immer wieder der Frage nachgegangen, ob das Schülergeschlecht einen Einfluss auf die Leistungseinschätzung von Lehrern hat.

## Wie genau sind Lehrerurteile?

Tournaki und Podell (2005) beschäftigten sich in ihrer Untersuchung, unter anderem mit dieser Fragestellung und fanden Hinweise dafür, dass Lehrerurteile für Mädchen signifikant besser ausfallen als für Buben. Die Ergebnisse anderer Autoren, wie beispielsweise von Schrader und Helmke (1990), weisen jedoch darauf hin, dass das Schülergeschlecht unbedeutend für die Leistungseinschätzung von Lehrern ist. Dieses Ergebnis zeigte sich auch in der Studie von Hecht und Greenfield (2002). Das Geschlecht der Schüler zeigte weder bei der Leistungseinschätzung in der 1. Klasse, noch bei der Vorhersage der Leseleistung in der 3. Klasse einen bedeutsamen Einfluss auf das Lehrerurteil. Schabmann und Schmidt (2009) fanden in ihrer Arbeit Hinweise für beide Annahmen. Es zeigte sich, dass Lehrer die Rechtschreibkompetenzen von Mädchen der 1. Klasse Grundschule signifikant besser beurteilten als die der Buben. Außerdem war zu sehen, dass zurückgezogenes Verhalten bei Mädchen zu einem signifikant schlechteren Lehrerurteil im Rechtschreiben führte als bei zurückgezogenen Buben. Bezüglich der Lesefähigkeit zeigten sich bei Schabmann und Schmidt (2009) jedoch auch keine Geschlechtsunterschiede. Hinnant et al. (2009) untersuchten den Geschlechtseinfluss von Schülern der 1., 3. und 5. Klasse auf die Lehrereinschätzungen bezüglich Lesen und Rechnen. Die Leseleistungen der Mädchen wurden in der 1., 3. und 5. Klasse, im Vergleich zu ihren Testergebnissen, überschätzt und die der Buben in der 1. und 5. Klasse unterschätzt. Bezüglich der Lehrereinschätzungen in Mathematik zeigte sich in der 5. Klasse, dass das Geschlecht einen signifikanten Einfluss auf diese hat. Auch hier wurden Mädchen besser eingeschätzt als durch ihre Testwerte indiziert. Bennett et al. (1993) sprechen in ihrer Arbeit von einem indirekten Einfluss des Schülergeschlechts auf das Lehrerurteil. Es zeigte sich, dass Lehrer das Verhalten von Buben als signifikant weniger adäquat beurteilten als das der Mädchen. Im Zusammenhang mit dem bereits weiter oben erwähnten Ergebnis von Bennett et al. (1993), dass schlechte Verhaltensnoten mit negativeren Leistungsurteilen einhergehen, führt die Betrachtung dieser beiden Ergebnisse dazu, dass Buben schlechtere Lehrerurteile erhalten als Mädchen.

### 5.1.3 Leistungsniveau und Intelligenz der Schüler

Bei der Einschätzung der Schülerleistungen von Lehrern, scheint das Leistungsniveau der Schüler eine nicht zu vernachlässigende Rolle zu spielen. Lehrern scheint es leichter zu fallen Kinder mit guten Fähigkeiten im Lesen, Rechtschreiben oder Rechnen zu identifizieren als Kinder mit Leistungsproblemen. In der Studie von Begeny et al. (2008) wurde bei Schülern der 2. und 3. Klasse Volksschule die Lesefähigkeit mittels 10 standardisierter Leseproben unterschiedlicher Schwierigkeit erhoben. Damit konnte das individuelle Leistungsniveau erfasst werden, ob sich der Schüler für seine Klassenstufe auf dem „frustrational“ (schlechte Lesefähigkeit), „instructional“ (durchschnittlich) oder „mastery“ (sehr gut) Level im Lesen befindet. Die Ergebnisse zeigten, dass Lehrer Schüler die sich laut Leseproben auf dem „mastery“ Level befanden sehr gut identifizieren konnten. Hingegen konnten sie weniger als die Hälfte der Schüler die dem „instructional“ oder „frustrational“ Level zugeordnet wurden richtig bestimmen. Auch in einer anderen Studie zeigte sich, dass Lehrer Schwierigkeiten haben leistungsschwächere Kinder und die drei Kinder mit den schlechtesten Leseleistungen in ihrer Klasse zu erkennen (Madelaine & Wheldall, 2005, zitiert nach Graney, 2008). In der Längsschnittstudie von Graney (2008) war außerdem zu sehen, dass Lehrer Probleme haben den Lernfortschritt leistungsschwacher Schüler akkurat einzuschätzen. Dabei wurde der Leistungsfortschritt von schwachen Schülern der 2. Schulstufe über sechs Wochen hinweg wöchentlich erfasst, welcher danach von den Lehrern eingeschätzt werden musste. Hier zeigte sich, dass Lehrer gerade im niedrigen Leistungsbereich ihre Schüler und deren Lernfortschritt tendenziell überschätzen. Einerseits scheinen Lehrer Schwierigkeiten mit der Identifikation von leseschwachen Schülern zu haben, andererseits zeigte sich in der Untersuchung von Begeny et al. (2008) auch, dass sie die pro Minute korrekt gelesenen Wörter bei diesen am genauesten einschätzten. Auch bei Bates und Nettlebeck (2001) zeigte sich, dass schwächere Leser von ihren Lehrern überschätzt werden und Schüler mit guten Leseleistungen unterschätzt werden. Gegensätzliche Ergebnisse fanden Tournaki und Podell (2005) in ihrer Arbeit. Schwache Leser wurden bei dieser Untersuchung von den Lehrern auch schlechter eingeschätzt als durchschnittlich

## Wie genau sind Lehrerurteile?

gute Leser. Außerdem zeigten die Autoren in ihrer Studie Hinweise dafür, dass wenn es für die Lese Probleme eines unterdurchschnittlich lesenden Kindes keinen ersichtlichen Grund gibt, dass dieses vom Lehrer besser beurteilt wird als wenn es klare Erklärungen für die schlechte Leseleistung gibt. Weiters zeigte sich, dass Lehrer bei der Leistungseinschätzung von schwachen Lesern deren Verhalten nicht berücksichtigen, bei Schülern mit durchschnittlich guten Leseleistungen hingegen schon.

Auch die Intelligenz eines Schülers kann das Lehrerurteil beeinflussen. In der Untersuchung von Schrader und Helmke (1990) zeigte sich diesbezüglich, dass die Intelligenz eines Schülers über die Leistung hinaus (in diesem Fall die Mathematikleistung) signifikant zur Vorhersage der Lehrereinschätzungen beitragen konnte. Das heißt, dass die Leistungseinschätzung auch bei objektiv gleicher Schülerleistung umso günstiger ausfällt, je höher die Intelligenz des Schülers ausgeprägt ist. Für die Autoren ist es wenig überraschend, dass die Intelligenz eines Schülers in die Leistungsbeurteilung mit einfließt, da Intelligenz und Leistung nicht nur empirisch zusammenhängen, sondern auch vermutlich vom Beurteiler nur schwer unterschieden werden können.

Zusammenfassend ist zu sagen, dass es sowohl Indizien für die Akkuratheit von Lehrereinschätzungen gibt, als auch Hinweise die diesen widersprechen (z.B. Hinnant et al., 2009; Brophy, 1983; Hoge & Butcher, 1984; Jussim & Eccles, 1992; Bates & Nettlebeck, 2001; Feinberg & Shapiro, 2003). Jene Studien bei denen sich gezeigt hat, dass Lehrereinschätzungen eher ungenau sind, haben auch gezeigt, dass leistungsfremde Variablen, wie z.B. Verhalten, Geschlecht, Leistungsniveau und noch einige andere Faktoren, einen Einfluss auf das Lehrerurteil haben und dieses verzerren (z.B. Tournaki & Podell, 2005; Schabmann & Schmidt, 2009; Hinnant et al., 2009; Schrader & Helmke, 1990). Hinsichtlich des Leistungsniveaus ist auffallend, dass Lehrereinschätzungen v.a. bei schwächeren Schülern ungenau sind (z.B. Begeny et al., 2008; Graney, 2008; Feinberg & Shapiro, 2003). Bezüglich der Variable Verhalten ist zu sehen, dass dieses gemeinsam mit der tatsächlichen Schülerleistung das Lehrerurteil am besten vorhersagen kann (z.B. Hecht & Greenfield, 2002; Bennett et al., 1993).

## Wie genau sind Lehrerurteile?

Bei der Moderatorvariable Geschlecht zeigen sich sehr widersprüchliche Ergebnisse. Zum einen ist zu sehen, dass die schulischen Leistungen der Buben im Vergleich zu ihren weiblichen Kolleginnen von Lehrern schlechter eingeschätzt werden (z.B. Bennett et al, 1993; Tournaki & Podell, 2005). Zum anderen scheint das Schülersgeschlecht jedoch unbedeutend für die Lehrereinschätzung zu sein (z.B. Schrader & Helmke, 1990; Hecht & Greenfield, 2002).

In der folgenden empirischen Untersuchung wird auf den Einfluss von Moderatorvariablen auf die Lehrereinschätzungen nicht weiter eingegangen, da diese Fragestellung bereits von Schabmann und Schmidt (2009) mit denselben Daten untersucht wurde die auch für diese Analyse verwendet wurden. In ihrer Untersuchung kamen die Autoren zu dem Schluss, dass einige Verhaltensweisen der Schüler das Lehrerurteil beeinflussen. Geschlechtsunterschiede zeigten sich nur beim Rechtschreiben.

Die vorliegende Studie beschäftigt sich v.a. mit der Frage, welchen Einfluss Lehrererwartungen auf die zukünftigen Leistungen von Schülern haben. Da für die Wirksamkeit des Erwartungseffekts fehlerhafte Lehrereinschätzungen Voraussetzung sind, müssen diese zu Beginn der Untersuchung auf ihre Akkuratheit überprüft werden.

### 6 Zielsetzung der Untersuchung

Ziel der vorliegenden Untersuchung ist es, die Wirksamkeit des Erwartungseffekts in einer Längsschnittstudie zu untersuchen. In dieser Arbeit wird in erster Linie der Frage nachgegangen, inwiefern sich fehlerhafte Lehrereinschätzungen in der 1. Klasse Volksschule im Lesen und Rechtschreiben auf die Leistungen der Kinder, betrachtet über einen Zeitraum von 4 Jahren, auswirken. Die Bereiche Lesen und Rechtschreiben werden dabei getrennt voneinander untersucht.

Seit den Anfängen in den 1960er Jahren wurde der Frage, wie sich das Über- oder Unterschätzen von Schulkindern auf deren spätere schulische Leistungen auswirkt, schon in vielen Untersuchungen nachgegangen, es gibt aber nur wenig Längsschnittstudien die sich mit diesem Forschungsbereich beschäftigen. Längsschnittstudien haben den Vorteil, dass sowohl zu sehen ist wie gut Lehrer die aktuellen Leistungen ihrer Schüler einschätzen können, als auch wie sich diese frühen Lehrereinschätzungen auf die späteren schulischen Leistungen der Kinder auswirken (Hinnant et al., 2009).

In dieser Studie wird auch untersucht, ob sich das Überschätzen von Schülern stärker auf deren weitere Leistungen auswirkt oder das Unterschätzen.

Wie bereits erwähnt, geht aus mehreren Studien hervor, dass das Leistungsniveau der Schüler einen Einfluss auf die Wirksamkeit des Erwartungseffekts haben kann. Auch auf das Lehrerurteil hatte der aktuelle Leistungsstand der Schüler einen Einfluss. Es hat sich in einigen Untersuchungen gezeigt, dass v.a. bei leistungsschwachen Schülern größere Erwartungseffekte auftreten (vgl. Hinnant et al., 2009; Jussim & Harber, 2005) und auch, dass diese Kinder von ihren Lehrern nur schwer identifiziert werden können und somit oft falsch eingeschätzt, meist überschätzt, werden (vgl. Begeny et al., 2008; Graney, 2008). Um diesen Einfluss des Leistungsniveaus der Kinder zu kontrollieren, wurden die Schüler hinsichtlich ihrer aktuellen Leistungen parallelisiert, d.h. es wurden Paare gebildet, jeweils mit einem über- und einem unterschätzten Kind, die sich in ihrer tatsächlichen schulischen Leistung im Lesen oder Rechtschreiben nicht unterscheiden. Das gleiche wurde mit dem Verhalten der Schüler gemacht,

welches durch die Lehrer eingeschätzt wurde, um auch den Einfluss des Schülerverhaltens auf die Lehrereinschätzungen kontrollieren zu können, da sich, wie ebenfalls bereits dargestellt, auch dieses auf das Lehrerurteil auswirken kann (vgl. Schabmann & Schmidt, 2009; Hecht & Greenfield, 2002; Bennett et al., 1993).

In der bisherigen Literatur gibt es keine Untersuchungen, die mit Hilfe von Parallelisierung den Einfluss von verschiedenen Variablen kontrolliert haben.

Im Detail sollen in der vorliegenden Längsschnittuntersuchung folgende Fragestellungen überprüft werden:

### **zentrale Fragestellung für Lesen**

- Unterscheiden sich die Leistungsverläufe der über- und unterschätzten Kinder über einen Zeitraum von 4 Jahren hinweg voneinander?

Falls sich keine signifikante Wechselwirkung zwischen den Erhebungszeitpunkten und den beiden Schülergruppen (über- und unterschätzte) zeigt, werden zusätzlich auch die Haupteffekte betrachtet; das heißt, dann wird auch auf die beiden Fragestellungen eingegangen, die sich auf den Haupteffekt Zeit und Schülergruppe beziehen:

- Unterscheiden sich die Leseleistungen aller Schüler gemeinsam betrachtet über einen Zeitraum von 4 Jahren hinweg voneinander?
- Unterscheiden sich über- und unterschätzte Kinder hinsichtlich ihrer Leseleistungen voneinander?

### **zentrale Fragestellung für Rechtschreiben**

- Unterscheiden sich die Leistungsverläufe der über- und unterschätzten Kinder über einen Zeitraum von 4 Jahren hinweg voneinander?

Auch hier gilt, dass wenn sich keine signifikante Wechselwirkung zwischen den Erhebungszeitpunkten und den beiden Schülergruppen (über- und unterschätzte) zeigt, zusätzlich auch die Haupteffekte Zeit und Schülergruppe betrachtet werden:

## Zielsetzung der Untersuchung

- Unterscheiden sich die Rechtschreibleistungen aller Schüler gemeinsam betrachtet über einen Zeitraum von 4 Jahren hinweg voneinander?
- Unterscheiden sich über- und unterschätzte Kinder hinsichtlich ihrer Rechtschreibleistungen voneinander?

Falls sich beim Lesen und/oder Rechtschreiben zeigt, dass sich die Leistungsverläufe der Über- und Unterschätzten signifikant voneinander unterscheiden, d.h. falls ein Pygmalioneffekt zu sehen ist, soll außerdem folgende Fragestellung untersucht werden:

- Wirkt sich das Über- oder Unterschätzen von Schülern stärker auf deren zukünftige Lese- oder Rechtschreibleistungen aus?

## 7 Methoden

### 7.1 Stichprobe

Die vorliegende Untersuchung bezieht sich auf eine Längsschnittstudie des Instituts für Psychologie der Universität Wien. Die Erhebung wurde an 15 Grundschulen in Niederösterreich durchgeführt. Die hier verwendeten Daten sind in zwei aufeinander folgende Kohorten gegliedert. Die erste Kohorte bezieht sich auf Kinder die im Schuljahr 1998/99 eingeschult wurden und die zweite Kohorte auf Schüler die 1999/00 in die Schule kamen. Die Längsschnittuntersuchung zieht sich in beiden Fällen über 4 Jahre, d.h. die Erhebungen der zwei Kohorten umfassen Daten von der 1. bis zur 4. Klasse Volksschule. An der Untersuchung nahmen insgesamt 293 Schüler teil, wobei etwas mehr Buben ( $n=163$ ) als Mädchen ( $n=130$ ) an der Studie teilnahmen (56,6% zu 44,4%). In beiden Kohorten gab es vier Erhebungszeitpunkte. Die erste Datenerhebung fand im Dezember nach der Einschulung statt. Die drei weiteren Erhebungen waren dann jeweils am Ende der 1., 2. und 4. Klasse Volksschule.

### 7.2 Erhebungsinstrumente

#### 7.2.1 Lehrereinschätzungen

Es wurden Lehrereinschätzungen bezüglich der Lese- und Rechtschreibleistung der Schüler im Dezember der 1. Klasse Volksschule, mit Hilfe eines Fragebogens, erhoben. Auch das Verhalten der Kinder sollten die Lehrer anhand eines Fragebogens zu diesem Zeitpunkt einschätzen. Mit diesen Fragebögen wurden indirekte und peerunabhängige Lehrereinschätzungen erhoben. Das heißt, die Items bezogen sich auf die bisherigen schulischen Leistungen im Lesen und Rechtschreiben im Allgemeinen und nicht auf spezifische Leistungstests. Die Lehrer sollten jedes Kind einzeln beurteilen, unabhängig von seinen Schulkameraden. Dasselbe gilt auch für die Verhaltenseinschätzungen der Lehrer (vgl. S. 26).

### 7.2.1.1 Leistung

Die Lehrereinschätzung hinsichtlich der Leseleistung der Mädchen und Buben wurde anhand von 6 Items erhoben. Um die Schüler einzuschätzen, hatten die Lehrer bei jedem Item eine 5-stufige Rating Skala von 0 (stimmt nicht) bis 4 (stimmt genau) zur Verfügung. Auch die Lehrereinschätzungen bezüglich der Rechtschreibleistung wurden anhand von 5-stufigen Rating-Skalen mit 6 Items erfasst.

**Tabelle 1: Faktorenanalyse Lesen**

	Faktor
	1
1. Das Kind kann die Position eines Lautes in einem Wort erkennen.	.863
2. Das Kind kann kurze Wörter in Laute aufgliedern.	.911
3. Das Kind kann zwei Laute zusammenschleifen.	.847
4. Das Kind kann alle durchgenommenen Buchstaben lesen.	.804
5. Das Kind kann alle bisher im Unterricht durchgenommenen Wörter sicher lesen.	.848
6. Das Kind kann auch bisher im Unterricht noch nicht durchgenommene Wörter, die aus bekannten Buchstaben zusammengesetzt sind, erlesen.	.839

**Tabelle 2: Faktorenanalyse Rechtschreiben**

	Faktor
	1
1. Das Kind hat mit der Linienführung beim Schreiben keine Probleme (schreibt sehr sauber).	.589
2. Das Kind kann alle durchgenommenen Buchstaben auf Ansage schreiben.	.761
3. Das Kind kann Wörter, die wir im Unterricht bereits geschrieben haben, bei nochmaliger Vorgabe richtig schreiben.	.866
4. Das Kind kann Wörter, die im Unterricht nur gelesen wurden, richtig schreiben.	.885
5. Das Kind kann auch kurze Wörter, die wir noch nicht gelesen haben, aber deren Buchstaben wir schon gelernt haben, richtig schreiben.	.874
6. Das Kind hat größere Leistungsschwankungen, d.h. es kann Wörter, die es schon lesen bzw. schreiben konnte, am nächsten Tag nicht mehr lesen oder schreiben.	-.628

Eine Faktorenanalyse mit den Items zur Einschätzung der Leseleistung der Kinder zeigt, dass die 6 Items zu einer Variable zusammengefasst werden können. Auch bei der Reliabilitätsanalyse war zu sehen, dass diese Variable auch eine sehr hohe Reliabilität hat ( $\alpha = .919$ ). Bezüglich der Items zur Einschätzung der Rechtschreibleistung der Schüler hat sich sowohl bei der Faktorenanalyse als auch bei der Reliabilitätsprüfung gezeigt, dass nur die 3 Items 3, 4 und 5 zur Bildung der Lehrereinschätzungsvariable für Rechtschreiben sinnvoll sind ( $\alpha = .907$ ). Auch inhaltliche Überlegungen sprechen dafür, nur die eben genannten Items für die Lehrereinschätzungsvariable Rechtschreiben heranzuziehen.

**7.2.1.2 Verhalten**

Wie schon erwähnt, sollten die Lehrer ihre Schüler am Anfang der 1. Klasse (Dez.) anhand eines Fragebogens auch im Verhalten einschätzen. Für diese Verhaltenseinschätzung hatten die Lehrer 26 Items mit jeweils einer 5-stufigen Rating Skala, 0 (stimmt nicht) bis 4 (stimmt genau), zur Verfügung. Von diesen Items wurden 17 ausgewählt und zu den Verhaltensdimensionen Hyperaktivität, oppositionelles Verhalten, depressives Verhalten, Aggression, sozialer Rückzug und Viktimisierung zusammengefasst (vgl. Schabmann & Schmidt, 2009). In der vorliegenden Untersuchung werden nur die Variablen Hyperaktivität, oppositionelles Verhalten, Aggression und sozialer Rückzug berücksichtigt, da sich, wie bereits erwähnt, bei Schabmann und Schmidt (2009) gezeigt hat, dass bei der hier verwendeten Längsschnittstudie nur diese Variablen einen Einfluss auf das Lehrerurteil haben.

**Tabelle 3: Faktorenanalyse Verhalten**

	Faktoren	
	1	2
oppositionelles Verhalten	.891	-.117
sozialer Rückzug	.062	.979
Aggression	.758	-.162
Hyperaktivität	.827	.201

Die Faktorenanalyse der Verhaltensvariablen und auch eine Reliabilitätsanalyse mit den Variablen oppositionelles Verhalten, Aggression und Hyperaktivität ( $\alpha = .767$ ) zeigen, dass jene Variablen die im 1. Faktor hoch laden, zu einer Variable (external) zusammengefasst werden können.

### **7.2.2 Lese- und Rechtschreibfähigkeit**

Neben der Lehrereinschätzung bezüglich Lesen und Rechtschreiben, wurden ca. 3 Monate nach Schulbeginn auch die tatsächlichen Lese- und Rechtschreibfähigkeiten der Kinder erfasst. Die tatsächlichen Leistungen, sowohl im Lesen als auch im Rechtschreiben, wurden jedoch auch am Ende der 1., 2. und 4. Klasse erhoben.

#### **7.2.2.1 Lesefähigkeit**

Erster Erhebungszeitpunkt (A1, November/Dezember 1.Klasse):

Die tatsächliche Leseleistung der Schüler wurde am Anfang der 1. Klasse mit dem Wiener Früherkennungstest (Klicpera, Humer, Gasteiger-Klicpera & Schabmann, 2008) erhoben. Die Kinder mussten sowohl bekannte Buchstaben und Wörter vorlesen als auch unbekannte Wörter und Pseudowörter (sinnfreie Wörter). Die Buchstaben und Wörter, die von den Mädchen und Buben vorgelesen werden sollten, waren auf das jeweilige Lesebuch, welches von ihrem Lehrer benutzt wurde, abgestimmt. Für die weitere Auswertung waren die richtig gelesenen Wörter relevant (T-Wert für alle richtig gelesenen Wörter).

Zweiter (E1), Dritter (E2) und Vierter (E4) Erhebungszeitpunkt (im Juni der 1., 2. und 4. Klasse):

Am Ende der 1., 2. und 4. Klasse wurde die Lesefähigkeit mit Hilfe des „Individuellen Lesetests“ (Klicpera & Gasteiger-Klicpera, 1994) erfasst. Die Kinder sollten 90 Wörter vorlesen, wobei der Test aus häufigen, seltenen und Pseudowörtern bestand, welche sich wiederum in kurze und lange Wörter unterteilen ließen. Bei dieser Testung wurde nicht nur darauf geachtet, ob die Wörter richtig gelesen wurden, sondern es wurde auch erhoben, wie viele Wörter

pro Minute richtig gelesen wurden. Auch hier wurden nur die richtig gelesenen Wörter ausgewertet (T-Wert für alle richtig gelesenen Wörter). Hinsichtlich der pro Minute richtig gelesenen Wörter wurden die richtig gelesenen, bekannten, neuen und Pseudowörter, zu allen 3 Zeitpunkten, zu einer Variable zusammengefasst.

### 7.2.2.2 Rechtschreibfähigkeit

Erster Erhebungszeitpunkt (A1, November/Dezember 1. Klasse):

Zur Erhebung der Rechtschreibfähigkeit mussten die Kinder bekannte Buchstaben und Wörter, die ihnen diktiert wurden, schreiben. Außerdem bekamen sie eine Leselernaufgabe vorgegeben, die insgesamt 8 Pseudowörter umfasste. Abschließend wurden ihnen auch diese Wörter diktiert. Bei der Auswertung wurden die richtig geschriebenen Wörter in Prozent angegeben

Zweiter Erhebungszeitpunkt (E1, Juni 1. Klasse):

Die Rechtschreibleistung der Schüler wurde auch am Ende der 1. Klasse überprüft. Es wurden den Kindern bekannte, neue und Pseudowörter diktiert. Ausgewertet wurden jene Wörter, die richtig geschrieben wurden (T-Wert für alle richtig geschriebenen Wörter).

Dritter und Vierter Erhebungszeitpunkt (E2 und E4, Juni 2. und 4. Klasse):

Am Ende der 2. und 4. Klasse wurden die Rechtschreibfertigkeiten der Buben und Mädchen mit Hilfe des Rechtschreibsubtests des Salzburger Lese Rechtschreib Tests erfasst. Auch bei diesem Lesetest wurden nur die richtig geschriebenen Wörter ausgewertet (T-Wert für die richtig geschriebenen Wörter).

## 8 Statistische Analyse

Um der Fragestellung nachgehen zu können, ob die Lehrereinschätzungen am Anfang der 1. Klasse Volksschule einen Einfluss auf die spätere Lese- und Rechtschreibleistung der Schüler haben, muss zunächst die Akkuratheit der Lehrereinschätzungen untersucht werden, da laut Merton (1948, zitiert nach Jussim & Harber, 2005) für das Auftreten eines Erwartungseffekts fehlerhafte Lehrererwartungen notwendig sind. Daher soll im Folgenden zuerst untersucht werden, wie genau die Lehrereinschätzungen in der vorliegenden Untersuchung sind.

### 8.1 Akkuratheit der Lehrerurteile

Um feststellen zu können, wie gut die untersuchten Lehrer ihre Schüler der 1. Klasse im Lesen und Rechtschreiben einschätzen konnten, wurden zunächst die tatsächlichen Schülerleistungen standardisiert und auch die Lehrereinschätzungen standardisiert, um diese vergleichbar zu machen. Danach wurde dann jeweils die Schülerleistung, welche mit den bereits weiter oben beschriebenen standardisierten Leistungstests erhoben wurde, von der jeweiligen Lehrereinschätzung subtrahiert. Das heißt beispielsweise, dass das Ergebnis im Lesetest eines Kindes, das es beim ersten Erhebungszeitpunkt (Nov./Dez.) erzielt hat, von der Leseinschätzung des Lehrers für dieses Kind, die zum selben Zeitpunkt erfasst wurde, subtrahiert wird. Es handelt sich dabei um die mittlere Abweichung zwischen dem Lehrerurteil und der tatsächlichen Leistung des Schülers. In der vorliegenden Arbeit wird von einer akkuraten Lehrereinschätzung gesprochen wenn die mittlere Abweichung einen Wert zwischen 0.9999 und -0.9999 hat. Ab einem Wert von -1 wird von einem Unterschätzen gesprochen und ab einem Wert von 1 wird von einem Überschätzen gesprochen.

Beim Lesen hat sich gezeigt, dass von 293 Schülern insgesamt 212 (119 Buben und 93 Mädchen) Kinder richtig eingeschätzt wurden. Von den 81 falsch eingeschätzten Kindern wurden 44 Schüler (23 Buben und 21 Mädchen) in ihrer Leseleistung überschätzt und 37 Kinder (21 Buben und 16 Mädchen) von ihren Lehrern unterschätzt. Auch beim Rechtschreiben wurden die meisten Kinder

richtig beurteilt (206, 105 Buben und 101 Mädchen) und weniger als ein Drittel der Schüler wurde falsch eingeschätzt (87). Davon wurden 45 Kinder überschätzt (29 Buben und 16 Mädchen) und 42 Kinder unterschätzt (29 Buben und 13 Mädchen). Für das weitere Vorgehen sind daher nur jene Schüler interessant, die in ihrer Schreib- und/oder Leseleistung entweder über- oder unterschätzt wurden. Das heißt für die folgende Parallelisierung ergibt sich eine Stichprobe von 168 Kindern.

### 8.2 Parallelisierung

Wie bereits im Theorieteil der vorliegenden Arbeit schon erwähnt wurde, zeigen sich in der Literatur widersprüchliche Ergebnisse hinsichtlich der größeren Wirksamkeit des Galatea- oder Golem-Effekts. Es gibt Studien die Hinweise finden, dass der Galatea-Effekt eine stärkere Wirkungskraft hat (z.B. Madon et al., 1997). Andererseits kommt es immer wieder zu Ergebnissen, die den gegenteiligen Effekt befürworten (z.B. Sutherland & Goldschmid, 1974; Alvidrez & Weinstein, 1999). Wieder andere Untersuchungen finden hingegen weder für den einen noch für den anderen Effekt eine höhere Einflusskraft (z.B. Hinnant et al., 2009; Babad et al., 1982).

In der vorliegenden Untersuchung soll ebenfalls der Frage nachgegangen werden, ob der Erwartungseffekt beim Über- oder Unterschätzen von Seiten der Lehrer wirkungsvoller ist. Um diese Fragestellung ohne Verzerrungen beantworten zu können, ist es notwendig, das Leistungsniveau der Schüler zu kontrollieren, da, wie bereits weiter oben erwähnt wurde, frühere Studien immer wieder zeigten, dass der Pygmalioneffekt bei leistungsschwachen/benachteiligten Schülern wirkungsvoller ist als bei durchschnittlich bis gut begabten/nicht benachteiligten Kindern (z.B. Hinnant et al., 2009; Jussim & Harber, 2005; Jussim et al., 1996; Madon et al., 1997). Außerdem ist in der Literatur zum Thema Akkuratheit von Lehrereinschätzungen zu sehen, dass das Leistungsniveau der Schüler auch einen Einfluss auf die Lehrereinschätzungen hat (z.B. Begeny et al., 2008; Graney, 2008; Bates & Nettlebeck, 2001). Auch diese Befunde sprechen dafür, dass es sinnvoll ist, das aktuelle Leistungsniveau der Schüler für die weitere Analyse zu kontrollieren.

Weiters wird das Verhalten der Schüler am Anfang der 1. Klasse Grundschule, das bei der vorliegenden Untersuchung mit Hilfe von Lehrerbefragungen erhoben wurde, konstant gehalten, da sich auch diesbezüglich immer wieder gezeigt hat, dass dieses einen Einfluss auf das Lehrerurteil haben kann (z.B. Schabmann & Schmidt, 2009; Bennett et al., 1993; Tournaki & Podell, 2005). In der hier durchgeführten Untersuchung wird somit der Einfluss der Verhaltensweisen oppositionelles Verhalten, Aggression und Hyperaktivität, zusammengefasst als externale Verhaltensvariable, durch die Parallelisierung kontrolliert, als auch die Verhaltensvariable sozialer Rückzug, da Schabmann und Schmidt (2009) zeigten, dass diese Variablen in den vorliegenden Daten einen Einfluss auf das Lehrerurteil haben.

Für die Parallelisierung werden daher Schülerpaare gebildet mit je einem über- und einem unterschätzten Kind, die sich hinsichtlich der Variablen tatsächliches Leistungsniveau (Lesen und Rechtschreiben), externes Verhalten (oppositionelles Verhalten, Aggression und Hyperaktivität) und sozialer Rückzug statistisch nicht signifikant voneinander unterscheiden ( $\alpha = .05$ ).

Um zunächst ein grobes Bild davon zu bekommen, wie gut eine Parallelisierung hinsichtlich des Leistungsniveaus der über- und unterschätzten Schüler funktionieren kann, werden in Tabelle 4 die Mittelwerte, Standardabweichungen und der kleinste und größte vorkommende Wert der Leistungsvariablen getrennt für Überschätzte und Unterschätzte dargestellt.

**Tabelle 4: Deskriptivstatistik**

	Gruppe	n	Mw	s
Leseleistung				
	über	44: 23m; 21w	35.473	8.992
	unter	37: 21m; 16w	52.383	7.024
Schreibleistung				
	über	45: 29m; 16w	37.830	9.801
	unter	42: 29m; 13w	54.804	7.194

m: männlich, w: weiblich; Mittelwerte (Mw) und Standardabweichungen (s) der über- und unterschätzten Schüler sind in T-Werten angegeben

Bereits bei der Deskriptivstatistik ist zu sehen, dass die T-Werte der tatsächlichen Lese- und Rechtschreibleistung der über- und unterschätzten Kinder recht weit auseinander liegen und es daher schwierig sein wird dahingehend Paare zu finden. Es zeigt sich, dass Schüler mit unterdurchschnittlichen Leistungen sowohl im Rechtschreiben als auch im Lesen eher überschätzt werden und Kinder, die in diesen Fächern durchschnittliche Leistungen zeigen, eher unterschätzt werden. Um Schülerpaare zu finden, die sich in ihrer aktuellen Leistung im Rechtschreiben oder Lesen nicht signifikant voneinander unterscheiden, wurde daher das Leistungsniveau der Über- und Unterschätzten direkt in den Daten miteinander verglichen, d.h. es wurden immer zwei Kinder gesucht (ein über- und ein unterschätztes), die sich in ihrer tatsächlichen Lese- oder Rechtschreibleistung beim 1. Erhebungszeitpunkt statistisch nicht voneinander unterscheiden, bei einem Signifikanzniveau von .05. Bei diesem Vergleich konnte das Schülergeschlecht nicht berücksichtigt werden, da es ohnedies schon sehr schwer war, immer zwei Kinder zu finden die in ihrer Lese- oder Rechtschreibleistung ähnlich sind, jedoch eines über- und das andere unterschätzt wurde. Mit dieser Methode der Datenanalyse konnten zumindest sowohl für Rechtschreiben als auch für Lesen je 11 Schülerpaare gefunden werden, die sich in ihrem aktuellen Leistungsniveau sehr ähnlich sind. Mit Hilfe von t-Tests für abhängige Stichproben soll nun auch statistisch überprüft werden, ob sich diese Paare hinsichtlich ihres Leistungsniveaus und ihres Verhaltens (external und sozialer Rückzug) nicht voneinander unterscheiden.

**Tabelle 5: Überprüfung der Parallelisierung mittels t-Tests für Lesen**

Parallelisierungsvariablen	Gruppe	<i>MW</i>	<i>s</i>	<i>t</i>	<i>df</i>	<i>p</i>
Leseleistung	über	42.832	5.362	-1.979	10	.076
	unter	44.017	6.681			
external	über	.032	.909	-1.711	9	.121
	unter	.640	1.131			
sozialer Rückzug	über	2.086	.337	-1.922	10	.083
	unter	2.314	.555			

**Tabelle 6: Überprüfung der Parallelisierung mittels t-Tests für Rechtschreiben**

Parallelisierungsvariablen	Gruppen	<i>MW</i>	<i>s</i>	<i>t</i>	<i>df</i>	<i>p</i>
Rechtschreibleistung	über	44.068	6.957	-1.936	10	.082
	unter	45.275	7.937			
external	über	-.426	.673	-2.196	8	.059
	unter	.383	.849			
sozialer Rückzug	über	2.199	.531	-1.899	10	.087
	unter	2.430	.453			

Die Ergebnisse der t-Tests zeigen, dass sich, sowohl beim Lesen als auch beim Rechtschreiben, die beiden Gruppen, Über- und Unterschätzte, in den Variablen tatsächliche Leistung und Verhalten nicht voneinander unterscheiden. Das bedeutet, dass die Parallelisierung gelungen ist und diese Variablen für die weitere statistische Analyse kontrolliert werden. Die folgenden Ergebnisse können daher durch diese nicht mehr verzerrt werden.

Für die weitere Auswertung ist zu berücksichtigen, dass durch die sehr kleine Stichprobe, sowohl bezüglich Lesen als auch hinsichtlich Rechtschreiben, nur sehr schwer signifikante Ergebnisse zu finden sein werden. Andererseits ist davon auszugehen, dass wenn sich signifikante Ergebnisse zeigen, diese nicht nur statistische sondern auch praktische Relevanz haben.

### 9 Ergebnisse

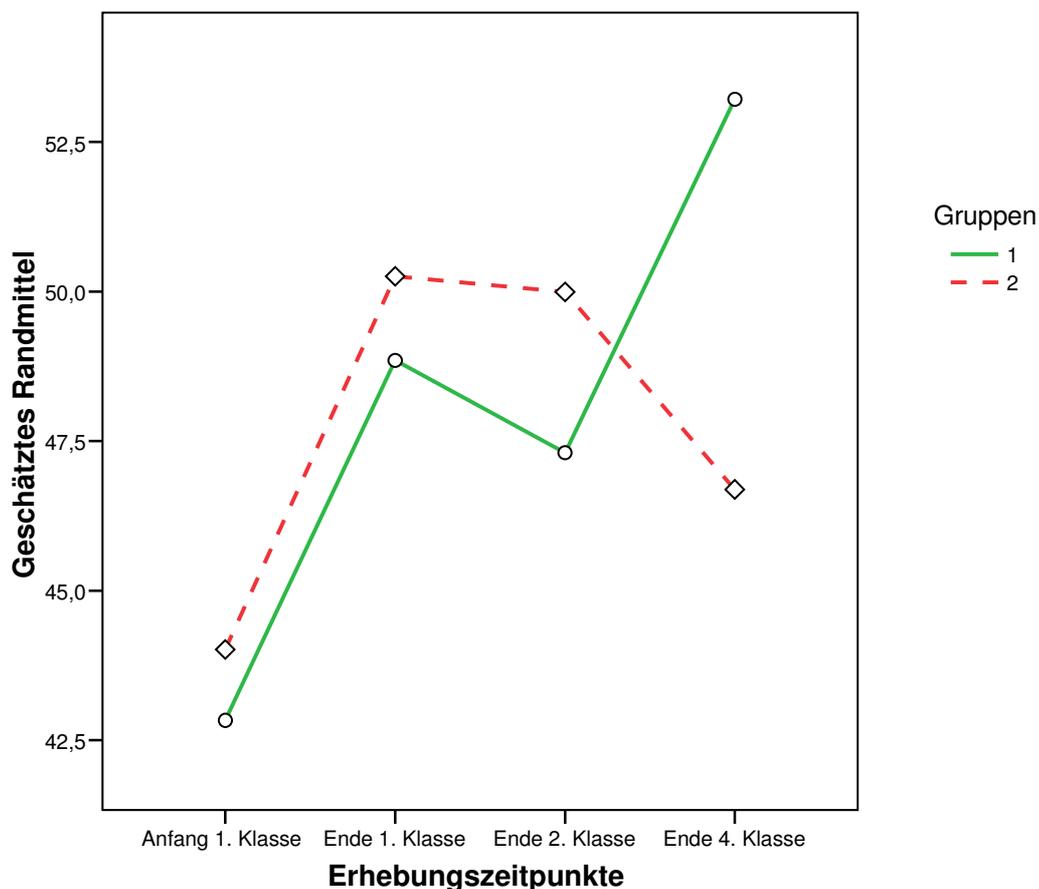
#### 9.1 Lesen

Die Fragestellungen in Bezug auf Lesen werden mit Hilfe einer Varianzanalyse mit 2 Messwiederholungsfaktoren (Erhebungszeitpunkt und Schülergruppe) untersucht. Dieses Verfahren vergleicht die Mittelwerte der überschätzten Kinder im Lesen mit den Leseleistungen der unterschätzten Schüler, über alle 4 Erhebungszeitpunkte hinweg.

*Unterscheiden sich die Leistungsverläufe der über- und unterschätzten Kinder über einen Zeitraum von 4 Jahren hinweg voneinander?*

Die Leistungsentwicklungen der über- und unterschätzten Schüler, betrachtet über 4 Schuljahre hinweg, unterscheiden sich nicht signifikant voneinander [ $F_{(3, 30)} = 1.481$ ;  $p = .240$ ;  $\alpha = .05$ ]. Das bedeutet, dass sich Kinder die in ihrer Leseleistung am Anfang der Volksschule überschätzt wurden, in ihrem Leistungsverlauf nicht von ihren unterschätzten Kollegen unterscheiden. Obwohl bei den Leistungsverläufen der über- und unterschätzten Kinder keine statistisch signifikanten Unterschiede zu sehen waren, zeigte sich trotzdem eine kleine Effektstärke ( $\eta^2 = .129$ ). In Diagramm 1 sind die soeben beschriebenen Ergebnisse auch noch graphisch dargestellt.

**Diagramm 1: Leistungsverläufe der über- und unterschätzten Kinder im Lesen**



Gruppe 1: überschätzte Kinder; Gruppe 2: unterschätzte Kinder

Auch anhand des Liniendiagramms, in dem die Leistungsverläufe der beiden Gruppen getrennt voneinander dargestellt werden, ist zu sehen, dass sich über- (Gruppe 1) und unterschätzte Kinder (Gruppe 2) im Lesen sehr ähnlich entwickeln. Vor allem bei Betrachtung der ersten 3 Messzeitpunkte zeigt sich, dass die Linien annähernd parallel zueinander verlaufen und nicht besonders weit auseinander liegen. Nach dem 3. Erhebungszeitpunkt (Ende 2. Klasse) zeigt sich im Liniendiagramm eine Tendenz, dass die überschätzten Schüler sich in ihrer Leseleistung plötzlich steigern, wohingegen die Leistung der Unterschätzten sich verschlechtert. Wie bereits schon erwähnt, wurde die Wechselwirkung zwischen Erhebungszeitpunkt und Gruppe aber nicht signifikant. Es zeigte sich jedoch eine kleine Effektstärke.

Da sich bezüglich der Leseleistung keine signifikante Wechselwirkung zwischen Erhebungszeitpunkten und den beiden Schülergruppen zeigt, werden auch die Fragestellungen betrachtet, die sich auf die Haupteffekte Zeit und Schülergruppe beziehen.

*Unterscheiden sich die Leseleistungen aller Schüler über einen Zeitraum von 4 Jahren hinweg voneinander?*

Bei gemeinsamer Betrachtung beider Schülergruppen, d.h. wenn die Leseleistungen der über- und unterschätzten Kinder zu jedem Messzeitpunkt zusammengefasst werden, zeigen sich keine signifikanten Leistungsunterschiede im Lesen zwischen den 4 Erhebungszeitpunkten [ $F_{(3, 30)} = 2.094$ ;  $p = .122$ ;  $\alpha = .05$ ]. Das bedeutet, dass sich die durchschnittlichen Leseleistungen, aller Kinder gemeinsam betrachtet, über alle 4 Erhebungszeitpunkte hinweg, nicht voneinander unterscheiden. Trotzdem zeigt sich auch hier zumindest ein kleiner Effekt von  $\eta^2 = .173$ .

*Unterscheiden sich über- und unterschätzte Kinder hinsichtlich ihrer Leseleistungen voneinander, unabhängig von den Erhebungszeitpunkten?*

Es zeigt sich, dass sich die Leseleistungen der über- und unterschätzten Kinder auch dann nicht voneinander unterscheiden, wenn die 4 Erhebungszeitpunkte nicht berücksichtigt werden [ $F_{(1, 10)} = .021$ ;  $p = .887$ ;  $\alpha = .05$ ], d.h. auch wenn der Mittelwert aller Leseleistungen der Überschätzten mit jenem der Unterschätzten verglichen wird, zeigen sich keine signifikanten Unterschiede zwischen den beiden Gruppen. Bei diesem Vergleich der beiden Schülergruppen zeigt sich eine Effektstärke von  $\eta^2 = .002$ , d.h. es zeigt sich kein Effekt.

Aufgrund dieser Ergebnisse kann gesagt werden, dass bei der vorliegenden Untersuchung beim Lesen kein Pygmalioneffekt zu sehen ist, also die fehlerhaften Lehrerurteile keine Auswirkungen auf die Leseleistungen der Schüler hatten.

## 9.2 Rechtschreiben

Auch die Fragestellungen in Bezug auf Rechtschreiben werden mittels einer Varianzanalyse mit 2 Messwiederholungsfaktoren (Erhebungszeitpunkt und

Schülergruppe) untersucht. Dieses Verfahren vergleicht die Mittelwerte im Rechtschreiben der überschätzten Kinder mit jenen der unterschätzten Schüler über alle 4 Erhebungszeitpunkte hinweg.

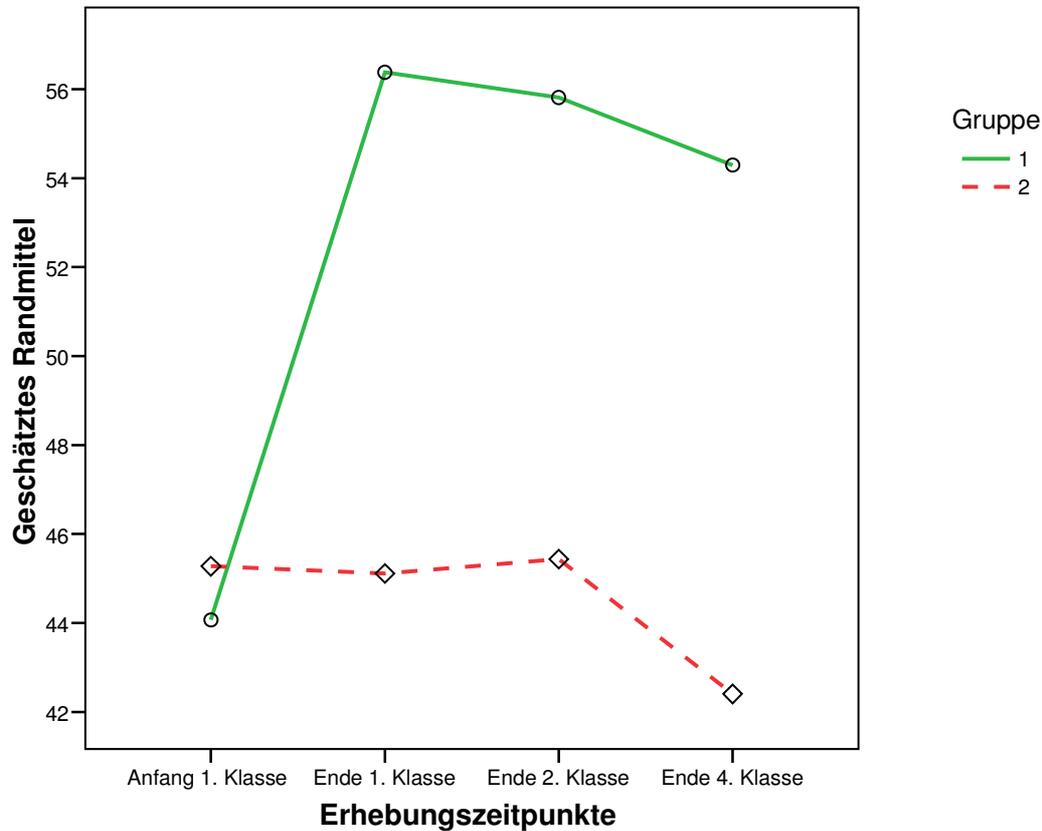
Da sich beim Rechtschreiben eine signifikante Wechselwirkung zwischen den Erhebungszeitpunkten und den Schülergruppen zeigt, wird hier nur auf die zentrale Fragestellung eingegangen, ob sich die Leistungsverläufe der über- und unterschätzten Kinder über alle 4 Erhebungszeitpunkte hinweg voneinander unterscheiden, da die Haupteffekte, die ebenfalls signifikant sind, durch eine signifikante Wechselwirkung verzerrt sein könnten. Das heißt, dass auf die beiden Fragestellungen, die sich auf den Haupteffekt Zeit und Schülergruppe beziehen, nicht näher eingegangen wird.

*Unterscheiden sich die Leistungsverläufe der über- und unterschätzten Kinder über einen Zeitraum von 4 Jahren hinweg voneinander?*

Wie eben schon erwähnt gibt es beim Rechtschreiben eine signifikante Wechselwirkung zwischen Zeit und Gruppe [ $F_{(1,601, 16)} = 4.435$ ;  $p = .036$ ;  $\alpha = .05$ ]. Das bedeutet, dass sich die Leistungsverläufe der über- und unterschätzten Kinder signifikant voneinander unterscheiden. Es zeigt sich außerdem eine moderate Effektstärke [ $\eta^2 = .307$ ].

Da dieses Ergebnis noch nichts darüber aussagt, wie sich die Leistungsverläufe der über- und unterschätzten Kinder voneinander unterschiedlich entwickeln, ist es sinnvoll, wie auch beim Lesen, zusätzlich noch das Liniendiagramm genauer zu betrachten.

**Diagramm 2: Leistungsverläufe der über- und unterschätzten Kinder im Rechtschreiben**



Gruppe 1: überschätzte Kinder; Gruppe 2: unterschätzte Kinder

Anhand des Liniendiagramms ist zu sehen, dass jene Kinder, die in ihrer Rechtschreibleistung am Anfang der 1. Klasse von ihren Lehrern überschätzt wurden (Gruppe 1), einen starken Leistungszuwachs im Rechtschreiben haben. Die unterschätzten Kollegen (Gruppe 2) zeigen hingegen keine relevanten Veränderungen in ihrer Rechtschreibleistung über den Zeitraum von 4 Jahren hinweg. Mit Hilfe des Liniendiagramms kann also auch die Frage beantwortet werden, ob das Über- oder Unterschätzen einen größeren Effekt auf die weitere Leistungsentwicklung von Schülern hat.

Die graphische Darstellung der Leistungsverläufe der beiden Gruppen spricht dafür, dass nur das Überschätzen der Rechtschreibleistungen in der 1. Klasse einen Einfluss auf die zukünftigen Leistungen der Kinder hat. Die Leistungen der

unterschätzten Kinder bleiben, dem Diagramm zufolge, über die Zeit relativ stabil. Bezüglich der überschätzten Schüler ist außerdem zu sehen, dass deren Rechtschreibleistungen, nach dem ersten Leistungszuwachs, ebenfalls ziemlich stabil bleiben. Das heißt, dass die Leistungen der Überschätzten zum 2., 3. und 4. Erhebungszeitpunkt, laut graphischer Darstellung, auch keine bedeutsamen Veränderungen mehr aufweisen.

Aufgrund der vorliegenden Ergebnisse kann gesagt werden, dass sich beim Rechtschreiben ein Erwartungseffekt gezeigt hat. Das Überschätzen der Lehrer, am Anfang der 1. Klasse, hat zu einer deutlichen Leistungssteigerung im Rechtschreiben geführt. Das heißt auch, dass sich nur ein positiver Erwartungseffekt, Galatea-Effekt, gezeigt hat.

### 10 Zusammenfassung und Diskussion

Ziel der vorliegenden Studie war es in erster Linie der Frage nachzugehen, ob sich Schülerleistungen, durch das Über- und Unterschätzen von Seiten der Lehrer, den Lehrererwartungen entsprechend entwickeln. Außerdem wurde untersucht, ob das Über- oder Unterschätzen von schulischen Leistungen einen stärkeren Effekt auf die zukünftige Leistungsentwicklung der Schüler hat.

Die Ergebnisse zeigen, dass fehlerhafte Lehrereinschätzungen beim Lesen keine Auswirkungen auf die weitere Leistungsentwicklung der Schüler in diesem Bereich haben. Das heißt, beim Lesen hat sich kein Erwartungseffekt gezeigt. Wie bereits im Ergebnisteil beschrieben, ist bei den überschätzten Kindern nach dem 3. Messzeitpunkt jedoch eine steile Leistungssteigerung zu sehen; die Leseleistungen der unterschätzten Schüler nehmen hingegen ab. Da der Leistungsverlauf der über- und unterschätzten Kinder jedoch davor sehr ähnlich ist und die statistische Analyse nur auf einer sehr kleinen Stichprobe basiert, wird diese doch eher starke Leistungsveränderung, im Vergleich zum vorigen Verlauf, nicht signifikant. Trotzdem zeigt sich zumindest eine kleine Effektstärke bezüglich der Wechselwirkung zwischen Erhebungszeitpunkten und Schülergruppen. Es ist jedoch fraglich, wodurch diese plötzliche Leistungssteigerung bzw. -senkung zustande kommt. Es ist eher unwahrscheinlich, dass die Lehrererwartungen, vom Anfang der 1. Klasse, sich erst am Ende der 2. Klasse Volksschule plötzlich so stark auswirken.

Beim Rechtschreiben zeigt sich hingegen, dass sich überschätzte Schüler signifikant besser in ihrer Rechtschreibleistung, innerhalb von 4 Jahren, entwickeln als unterschätzte Kinder. Das bedeutet, dass sich beim Rechtschreiben ein positiver Erwartungseffekt gezeigt hat (Galatea-Effekt). Außerdem zeigte sich eine mittlere Effektstärke hinsichtlich der Wechselwirkung zwischen Messzeitpunkten und den beiden Schülergruppen. Das Unterschätzen scheint, bei Betrachtung der graphischen Darstellung, keinen Effekt auf die weitere Rechtschreibleistung zu haben. Wie auch z.B. bei Madon et al. (1997), zeigt sich bezüglich des Rechtschreibens, dass das Überschätzen wirksamer ist als das Unterschätzen. Zu

## Zusammenfassung und Diskussion

beachten ist jedoch, dass das Überschätzen nur vom ersten zum zweiten Erhebungszeitpunkt zu einem starken Leistungsanstieg führt, danach bleiben die Leistungen der überschätzten Kinder ebenfalls stabil.

Bei der Untersuchung war außerdem zu sehen, dass, wie auch bei Bates und Nettlebeck (2001), leistungsschwache Kinder von ihren Lehrern eher überschätzt werden und durchschnittlich gute Schüler entweder richtig beurteilt oder eben unterschätzt werden. Dieses Ergebnis kann einerseits dafür sprechen, dass Lehrer Schwierigkeiten haben schwache Schüler zu identifizieren (vgl. Begeny et al., 2008;), andererseits könnte es ein Hinweis dafür sein, dass Lehrer versuchen die unterschiedlichen Leistungsniveaus der Kinder einander anzunähern, indem sie gute Schüler unter- und schlechte Schüler überschätzen.

Insgesamt ist zu sagen, dass Lehrer ihre Schüler größtenteils richtig beurteilen. Sowohl beim Lesen als auch beim Rechtschreiben schätzten Lehrer nur weniger als ein Drittel ihrer Schüler in ihren schulischen Leistungen falsch ein. Zu beachten ist jedoch, dass v.a. sehr schwache Schüler oft überschätzt werden und daher auch nicht erkannt werden. Dies kann in weiterer Folge dazu führen, dass keine entsprechenden Fördermaßnahmen ergriffen werden um die Defizite der leistungsschwachen Schüler bearbeiten zu können. Für den Schulalltag bedeutet das, dass Lehrer dabei Unterstützung bräuchten, schwache Schüler zu erkennen und in weiterer Folge diese auch entsprechend zu fördern.

Weiters wird sichtbar, dass beim Auftreten eines Erwartungseffekts, bei kontrolliertem Leistungsniveau und Verhalten, nur das Überschätzen relevante Auswirkungen auf die späteren Schülerleistungen hat. Dieses Ergebnis ist beruhigend, da, wie bereits weiter oben erwähnt, es natürlich günstiger ist, wenn durch das Überschätzen von schulischen Leistungen sich diese verbessern, durch das Unterschätzen aber die Leistungen der Schüler zumindest stabil bleiben und sich nicht verschlechtern.

Im Zusammenhang mit der vorliegenden Untersuchung ist zu berücksichtigen, dass sich die Ergebnisse im Lesen und Rechtschreiben auf jeweils sehr kleine Stichproben beziehen und daher eine Generalisierbarkeit der gefundenen

## Zusammenfassung und Diskussion

Ergebnisse nur eingeschränkt möglich ist. Es wäre interessant, in zukünftigen Studien diesen Fragestellungen mit einer größeren parallelisierten Stichprobe nachzugehen. Weiters wäre es von Vorteil, in folgenden Arbeiten auch das Lehrerverhalten bei der Untersuchung des Erwartungseffekts zu berücksichtigen, da sich in einigen Studien gezeigt hat, dass sich manche Lehrer differenzierter als andere ihren Schülern gegenüber verhalten (z.B. Brattesani et al., 1984). Außerdem wäre es sinnvoll die Lehrereinschätzungen auch bei den drei weiteren Erhebungszeitpunkten zu berücksichtigen, da dadurch die Stabilität der Lehrererwartungen überprüft werden und die Erwartungsverläufe mit den Leistungsverläufen der Schüler verglichen werden könnten. Die Analyse dieser beiden Aspekte ist bereits, mit dem hier verwendeten Datensatz, in Arbeit.

## 11 Abbildungsverzeichnis

Tabelle 1: Faktorenanalyse Lesen .....	41
Tabelle 2: Faktorenanalyse Rechtschreiben .....	41
Tabelle 3: Faktorenanalyse Verhalten.....	42
Tabelle 4: Deskriptivstatistik.....	47
Tabelle 5: Überprüfung der Parallelisierung mittels t-Tests für Lesen.....	48
Tabelle 6: Überprüfung der Parallelisierung mittels t-Tests für Rechtschreiben....	49
Diagramm 1: Leistungsverläufe der über- und unterschätzten Kinder im Lesen...51	
Diagramm 2: Leistungsverläufe der über- und unterschätzten Kinder im Rechtschreiben .....	54

## 12 Literatur

- Alvidrez, J., & Weinstein, R. S. (1999). Early Teacher Perceptions and Later Student Academic Achievement. *Journal of Educational Psychology*, 91, 731-746.
- Babad, E., Inbar, J., & Rosenthal, R. (1982). Pygmalion, Galatea and the Golem: Investigations of biased and unbiased teachers. *Journal of Educational Psychology*, 74, 459-474.
- Bates, C., & Nettelbeck, T. (2001). Primary school teachers' judgments of reading achievement. *Educational Psychology*, 21, 177-187.
- Begeny, J. C., Eckert, T. L., Montarello, S. A., & Storie, M. S. (2008). Teachers' perceptions of students' reading abilities: An examination of the relationship between teachers' judgments and students' performance across a continuum of rating methods. *School Psychology Quarterly*, 23, 43-55.
- Bennett, R. E., Gottesman, R. L., Rock, D. A., & Cerullo, F. (1993). Influence of behavior perceptions and Gender on teachers' judgments of students' academic skill. *Journal of Educational Psychology*, 85, 347-356.
- Berka, H. H., & Westhoff, K. (1981). Theorie und Methoden. Lehrererwartungen und Schülerverhalten. *Zeitschrift für Sozialpsychologie*, 12, 1-23.
- Brattesani, K., Weinstein, R., & Marshall, H. (1984). Student perceptions of differential teacher treatment as moderator of teacher expectation effects. *Journal of Educational Psychology*, 76, 236-247.
- Brophy, J. (1983). Research on the Self-Fulfilling Prophecy and Teacher Expectations. *Journal of Educational Psychology*, 75, 631-661.

- Brophy, J., & Good, T. (1970). Teacher`s communication of differential expectations for children`s classroom performance: Some behavioral data. *Journal of Educational Psychology*, 61, 365-374.
- Casparis, C. (1980). Diskussion: Eine theoretische Analyse von <<Pygmalioneffekten>> und <<sich selbst erfüllenden Voraussagen>>. *Zeitschrift für Sozialpsychologie*, 11, 124-128.
- Demaray, M. K., & Elliot, S. N. (1998). Teachers` judgments of students` academic functioning: A comparison of actual and predicted performances. *School Psychology Quarterly*, 13, 8-24.
- Eder, D. (1981). Ability grouping as a self-fulfilling prophecy: A micro-analysis of teacher-student interaction. *Sociology of Education*, 54, 151-161.
- Egan, O., & Archer, P. (1985). The accuracy of teachers` ratings of ability: A regression model. *American Educational Research Journal*, 22, 25-34.
- Elashoff, J. D., & Snow, R. E. (1972). *Pygmalion auf dem Prüfstand* (Übers.). München: Kösel. (Original erschienen 1971: Pygmalion reconsidered)
- Gilbert, D. T. (1995). Attribution and Interpersonal Perception. In A. Tesser (Ed.): *Advanced Social Psychology* (99-148). New York: Mrs. Simon Guggenheim Fund.
- Feinberg, A. B., & Shapiro, E. S. (2003). Accuracy of Teacher Judgments in Predicting Oral Reading Fluency. *School Psychology Quarterly*, 18, 52-65.
- Good, T. L., & Nichols, S. L. (2001). Expectancy Effects in the Classroom: A Special Focus on Improving the Reading Performance of Minority Students in First-Grade Classrooms. *Educational Psychologist*, 36, 113-126.

- Graney, S. (2008). General Education Teacher Judgments of their Low-Performing Students`Short-Term Reading Progress. *Psychology in the Schools*, 45, 537-549.
- Hecht, S. A., & Greenfield, D. B. (2002). Explaining the predictive accuracy of teacher judgments of their student`s reading achievement: The role of gender, classroom behavior, and emergent literacy skills in a longitudinal sample of children exposed to poverty. *Reading and Writing: An Interdisciplinary Journal*, 15, 789-809.
- Hinnant, J. B., O`Brien, M., & Ghazarian, S. R. (2009). The Longitudinal Relations of Teacher Expectations to Achievement in the Early School Years. *Journal of Educational Psychology*, 101, 662-670.
- Hoge, R. D., & Butcher, R. (1984). Analysis of teacher judgments of pupils achievement levels. *Journal of Educational Psychology*, 76, 777-781.
- Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research*, 59, 297-313.
- Jussim, L., & Eccles, J. (1992). Teacher Expectations II: Construction and reflection of student achievement. *Journal of Personality and Social Psychology*, 63, 947-961.
- Jussim, L., & Eccles, J. (1995). Naturally Occurring Interpersonal Expectancies. In Eisenberg, N. (Ed.): *Social Development* (74-108), Thousand Oaks: Sage Publications.
- Jussim, L. Eccles, J., & Madon, S. J. (1996). Social perception, social stereotypes, and teacher expectations: Accuracy and the quest for the powerful self-fulfilling prophecy. *Advances in experimental social psychology*, 29, 281-388.

- Jussim, L., & Harber, K. D. (2005). Teacher Expectations and Self-Fulfilling Prophecies: Knowns and Unknowns, Resolved and Unresolved Controversies. *Personality and Social Psychology Review*, 9, 131-155.
- Klicpera, C., & Gasteiger-Klicpera, B. (1994). Die langfristige Entwicklung der mündlichen Lesefähigkeit bei guten und schwachen Lesern. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 26, 278-290.
- Klicpera, C., Huemer, R., Gasteiger-Klicpera, B., & Schabmann, A. (2008). *WFT – Wiener Früherkennungstest. Leselehrgangsspezifisches Verfahren zur Früherkennung von Kindern mit Lese- und Rechtschreibschwierigkeiten im Schuleingangsbereich*. Wien: Verlag E. Dörner.
- Lekholm, A. K., & Cliffordson, C. (2008). Discrepancies between school grades and test scores at individual and school level: effects of gender and family background. *Educational Research and Evaluation*, 14, 181-199.
- Lexikon der Psychologie (2001). 4. Band, 170. Heidelberg: Spektrum akademischer Verlag GmbH
- Ludwig, P. H. (1991). *Sich selbst erfüllende Prophezeiungen im Alltagsleben*. Stuttgart: Verlag für Angewandte Psychologie.
- Ludwig, P. H. (1994). Pygmalion in der Odyssee? Zur Vermengung von Experiment und Alltagsrealität. *Zeitschrift für Pädagogische Psychologie*, 8, 99-102.
- Ludwig, P. H. (2006). Erwartungseffekt. In Rost, D. H. (Hrsg.): *Handwörterbuch Pädagogische Psychologie*. 3. Auflage (132-138). Weinheim: Beltz PVU

- Madon, S. J., Jussim, J., & Eccles, J. (1997). In search of the powerful self-fulfilling prophecy. *Journal of Personality and Social Psychology*, 72, 791-809.
- Raudenbush, S. W. (1984). Magnitude of teacher expectancy effects on pupil IQ as a function of the credibility of expectancy inductions: A synthesis of findings from 18 experiments. *Journal of Educational Psychology*, 76, 85-97.
- Rosenthal, R., & Jacobson, L. (1974). *Pygmalion im Unterricht. Lehrererwartung und Intelligenzentwicklung der Schüler* (Übers.). Weinheim: Beltz. (Original erschienen 1968: Pygmalion in the classroom)
- Rosenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects: the first 345 studies. *The Behavioral and Brain Sciences*, 3, 377-414.
- Rosenthal, R. (2000). Expectancy Effects. In Alan, E. K. (Hrsg.): *Encyclopedia of Psychology*, Vol. 3 (294-296), Oxford University Press. New York.
- Rubie-Davies, C., Hattie, J., & Hamilton, R. (2006). Expecting the best for students: Teacher expectations and academic outcomes. *British Journal of Educational Psychology*, 76, 429-444.
- Schabmann, A., & Schmidt, B. M. (2009). Sind Lehrer gute Leserechtschreibdiagnostiker? Der Einfluss von problematischem Schülerverhalten auf die Einschätzungen der Lesekompetenz durch Lehrkräfte. *Heilpädagogische Forschung*, XXXV, 3, 133-145.
- Schrader, F.-W. (1989). *Diagnostische Kompetenz von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Frankfurt: Lang.
- Schrader, F. W. (2001). Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 68-71). Weinheim: Beltz.

- Schrader, F.-W., & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1, 27-52.
- Schrader, F.-W., & Helmke, A. (1990). Lassen sich Lehrer bei der Leistungsbeurteilung von sachfremden Gesichtspunkten leiten? Eine Untersuchung zu Determinanten diagnostischer Lehrerurteile. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 22, 312-324.
- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19, 85-95.
- Spitz, H. H. (1999). Beleaguered Pygmalion: A History of the Controversy Over Claims That Teacher Expectancy Raises Intelligence. *Intelligence*, 27, 199-234.
- Sutherland, A., & Goldschmid, M. L. (1974). Negative teacher expectation and IQ change in children with superior intellectual potential. *Child Development*, 45, 852-856.
- Thorndike, R. L. (1972). Rezension von "Pygmalion im Unterricht". In J. Elashoff & R. Snow (Hrsg.), *Pygmalion auf dem Prüfstand* (85-88). München: Kösel.
- Tournaki, N., & Podell, D. M. (2005). The impact of student characteristics and teacher efficacy on teachers' predictions of student success. *Teaching and Teacher Education*, 21, 299-314.
- Weinstein, R. (1976). Reading group membership in first grade: Teacher behaviors and pupil experience over time. *Journal of Educational Psychology*, 68, 103-116.

Weinstein, R., & McKnown, C. (1998). Expectancy effects in "context:" Listening to the voices of students and teachers. In J. Brophy (Ed.), *Advances in research on teaching: Expectations in the classroom*, (7, 215-242). Greenwich, CT:JAI.

Weinstein, R., & Middlestadt, S. (1979). Student perceptions of teacher interactions with male high and low achievers. *Journal of Educational Psychology*, 71, 421-431.

Werner, B. (1999). Das Pygmalion-Motiv in der Aufklärung. In W.-D. Scholz & H. Schwab, (Hrsg.): *Bildung und Gesellschaft im Wandel*. (155-170). Oldenburg: Bibliotheks- und Informationssystem ([http://docserver.bis.uni-oldenburg.de/\\_\\_publikationen/bisverlag/2000/schbil99/schbil99.html](http://docserver.bis.uni-oldenburg.de/__publikationen/bisverlag/2000/schbil99/schbil99.html)); 28.12.09

# Lebenslauf

## Persönliche Daten

Madeleine Scholze, geboren am 7. August 1981 in Wien

Wohnhaft in A-2102 Bisamberg, Bundesstraße 50/5

Staatsbürgerschaft: Österreich

## Schulbildung/Ausbildung

1987-1991	Volksschule in 1210 Wien
1991-1995	Musikhauptschule in Korneuburg/NÖ
1995-2000	BHS Kindergartenpädagogik in 1100 Wien
9. Juni 2000	Maturaabschluss der BHS Kindergartenpädagogik
SS 2002	Psychologiestudium an der Universität Wien
WS 2002	FH für Sozialarbeit in Feldkirchen/Kärnten
seit SS 2003	Psychologiestudium an der Universität Wien
seit September 2009	Psychotherapeutisches Propädeutik (APG)

## Berufsweg:

Juli & August 2002	Kinderanimation / -betreuung am Campingplatz/Kärnten
Juni 2004-Juni 2005	Teilleistungs- und Lernbetreuung am Institut für Bildungsbegleitung (IBB), 1070 Wien
September 2005- August 2006	Horterzieherin im Schülertreff Bisamberg
Juli & August 2007	Ferialpraxis bei IBM, 1020 Wien
Juli-September 2008	6-Wochen Praktikum im SMZ-Ost an der Abteilung für Kinder- und Jugendheilkunde (240 Std.)
September & Oktober 2008	Praktikum an der Fakultät für Psychologie der Universität Wien im Arbeitsbereich Bildungspsychologie und Evaluation

## Sonstiges

August 2000- September 2001	Auslandsjahr als Au-Pair in den USA
--------------------------------	-------------------------------------