



universität
wien

Diplomarbeit

Titel der Arbeit

Entwicklung der Akkuratheit von Lehrereinschätzungen
in der Grundschule.

Welchen Einfluss hat externalisierendes Verhalten?

Verfasserin

Christina Köhl

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag^a. rer. nat.)

Wien, am 22.3.2011

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

Danksagung

Im Rahmen dieser Arbeit möchte ich mich bei all jenen bedanken, die mich in meinem Studium unterstützt haben. Ich danke meiner Mutter für ihren unermüdlichen und sehr motivierenden Glauben an meine akademischen Fähigkeiten und ihre uneingeschränkte Unterstützung in allen Lebenslagen. Ich danke meinem Vater für seine langjährige Unterstützung, die mir ermöglichte, mich voll und ganz auf mein Studium zu konzentrieren und es abzuschließen. Ebenso großer Dank gilt meinem Lebenspartner Thomas, der mir vor allem den richtigen Blick auf das Studium ermöglichte und mir stets stützend zur Seite stand. Er sorgte für die nötige Motivation sowie die Rahmenbedingungen, die zur Verfassung der vorliegenden Arbeit und zu dem Abschluss des Studiums wesentlich waren.

Ebenso motivierend und stützend standen mir viele andere wichtige Familienmitglieder und meine lieben Freunde zur Seite, sei es mit Worten des Trostes und Verständnisses oder mit Worten der Motivation.

Herrn Ao. Univ.-Prof. Dr. Alfred Schabmann danke ich für seine methodische und praktische Unterstützung sowie sein Verständnis während der Betreuung meiner Diplomarbeit.

Erklärung

In der vorliegenden Arbeit wird aufgrund der besseren Lesbarkeit auf eine gendersensible Sprache verzichtet. Bei Verwendung des männlichen Geschlechts ist selbstverständlich stets auch das weibliche angesprochen.

Inhaltsverzeichnis

1	Einleitung	6
2	Beurteilung von Schülerleistungen.....	9
2.1	Lehrerurteile.....	9
2.2	Akkuratheit von Lehrereinschätzungen	12
2.3	Einflüsse auf die Akkuratheit von Lehrereinschätzungen.....	18
2.3.1	Schülermerkmale.....	19
3	Zielsetzung der Untersuchung	23
4	Methode	26
4.1	Stichprobe.....	26
4.2	Erhebungsinstrumente	27
4.2.1	Schülerleistungen im Lesen und Rechtschreiben.....	27
4.2.2	Lehrereinschätzungen von Schülerleistungen und Schülerverhalten	28
5	Statistische Analyse.....	30
5.1	Beschreibung des verwendeten statistischen Verfahrens.....	30
5.2	Bildung der Skalen.....	33
5.2.1	Bildung der Skalen zur Schülerleistung	33
5.2.2	Bildung der Skalen zur Lehrereinschätzung	35
5.2.3	Bildung der Skalen zur Verhaltenseinschätzung	37
5.3	Kriterien der Modellbeurteilung	39
6	Ergebnisse.....	40
6.1	Darstellung und Interpretation der Ergebnisse.....	40
6.1.1	Pfadmodell Lesesicherheit	41
6.1.2	Pfadmodell Lesegeschwindigkeit.....	47
6.1.3	Pfadmodell Rechtschreibsicherheit	51
6.2	Beantwortung der Fragestellungen	58
6.2.1	Lesesicherheit	58
6.2.2	Lesegeschwindigkeit	59
6.2.3	Rechtschreibsicherheit	60
7	Zusammenfassung und Diskussion	62
8	Abbildungsverzeichnis	66
9	Tabellenverzeichnis	67
10	Literaturverzeichnis	68
11	Anhang.....	72

Abstract

Die vorliegende Studie untersucht die Fragestellung, wie gut sich Lehrereinschätzungen bezüglich des Lesens und Rechtschreibens aus Schülerleistungen erklären lassen beziehungsweise wie akkurat Lehrereinschätzungen sind. Weiters wird erhoben, ob die durch den Lehrer wahrgenommenen externalisierenden Verhaltensweisen der Schüler einen Einfluss auf die Lehrereinschätzungen sowie auch auf die Schülerleistungen haben. Zu diesem Zweck wurden die Lese- und Rechtschreibfähigkeiten mittels standardisierter Tests und Lehrereinschätzungen bezüglich der genannten Schülerfähigkeiten und externalisierenden Verhaltensweisen mittels Fragebögen an 15 niederösterreichischen Grundschulen erhoben. Es nahmen 32 Klassen und insgesamt 293 Schüler an der Erhebung teil, welche am Anfang und Ende der ersten Klasse sowie am Ende der zweiten und vierten Klasse stattfanden. Zur statistischen Analyse der Fragestellungen wurde die Methode der Strukturgleichungsmodelle herangezogen. Es zeigte sich, dass die Akkuratheit der Lehrereinschätzungen bezüglich der Schülerleistungen nur gering bis mäßig war. Insgesamt betrachtet waren die Einschätzungen im Rechtschreiben etwas akkurater als im Lesen. Auch das externalisierende Verhalten konnte zumindest gering zur Erklärung der Lehrereinschätzungen beitragen. Sein Einfluss auf die Leistungen der Schüler war nur teilweise vorhanden und sehr gering.

Aus diesen Ergebnissen wird der Schluss gezogen, dass Lehrer Schwierigkeiten haben spezifische Fähigkeiten ihrer Schüler akkurat einzuschätzen und sich in ihren Urteilen auch durch sachfremde Kriterien wie das externalisierende Verhalten der Schüler leiten lassen.

1 Einleitung

Warum ist es wichtig Lehrereinschätzungen unter das Mikroskop von Forschern zu nehmen?

Die Antwort ist einfach: Lehrereinschätzungen bzw. Lehrerurteile haben eine große praktische Relevanz im Leben eines jeden Schülers!

Die Gestaltung und Planung des Unterrichts (Schrader, 2006), die Einleitung frühzeitiger Fördermaßnahmen (Deutsches PISA-Konsortium, 2001; Spinath, 2005), Schullaufbahnentscheidungen (Thomas, 2007) – all jene pädagogischen Entscheidungen hängen von der Akkuratheit des Lehrerurteils ab (Feinberg & Shapiro, 2003).

Hoge und Coladarci (1989) zeigten mit ihrer Metaanalyse über 16 Studien, dass Lehrer zu akkuraten Einschätzungen von Schülerleistungen fähig sind. Allerdings gibt es sehr widersprüchliche Ergebnisse in der Literatur.

Aus der Studie von Hecht und Greenfield (2002) ist bekannt, dass nicht nur der tatsächliche Leistungsstand der Schüler im Lesen das Lehrerurteil beeinflusst, sondern ebenso Schülermerkmale wie Verhalten einen direkten Einfluss auf die Einschätzungen eines Lehrers bezüglich eines Schüler haben und somit die Akkuratheit des Lehrerurteils beeinflussen.

Die vorliegende Diplomarbeit soll dazu dienen, die langfristige Entwicklung der Akkuratheit von Lehrereinschätzungen bezüglich der Lese- und Rechtschreibleistungen von Grundschulkindern über einen Zeitraum von vier Jahren betrachten zu können. Strukturgleichungsmodelle dienen zur Veranschaulichung dieser Entwicklung und der verschiedenen Einflussfaktoren hinsichtlich der Lehrereinschätzungen.

Im Kontext der Studie wird einerseits die Fragestellung untersucht, wie gut sich Lehrereinschätzungen aus Schülerleistungen erklären lassen beziehungsweise wie akkurat Lehrereinschätzungen sind, andererseits wird erhoben, ob das durch den Lehrer wahrgenommene Schülerverhalten ebenfalls einen Einfluss auf die Lehrereinschätzungen als auch auf die Schülerleistungen hat.

Dazu liegen drei Strukturgleichungsmodelle vor, die einen Blick auf die Akkuratheit der Lehrereinschätzungen bezüglich der Schülerleistungen in den Bereichen Lesesicherheit, Lesegeschwindigkeit und Rechtschreibsicherheit über vier Jahre Grundschulzeit hinweg ermöglichen.

Einleitung

Da sich die vorliegende Arbeit mit der Akkuratheit von Lehrereinschätzungen und ihren Einflussfaktoren beschäftigt, wird im Folgenden das Forschungsgebiet der Lehrerurteile näher beleuchtet und der aktuelle Wissenstand dargelegt. Darauf folgen die Darstellung des methodischen Teils der Studie sowie ihre Ergebnisse.

Im Anschluss an die Untersuchungsergebnisse wird eine kritische Betrachtung der Studie dargelegt und auch ein Ausblick darauf gegeben, welche Aspekte in weiterführenden Forschungsarbeiten im Bereich des Lehrerurteils von Interesse sein könnten.

2 Beurteilung von Schülerleistungen

Das folgende Kapitel beschäftigt sich mit den grundlegenden Erkenntnissen und Theorien bezüglich der Lehrerurteile und geht danach über in das Forschungsgebiet der Akkuratheit der Lehrerurteile.

2.1 Lehrerurteile

Zunächst erscheint die Definition einer Reihe von Begrifflichkeiten im Feld des pädagogischen Urteilens sinnvoll. Der Begriff Beurteilung bezieht sich speziell in pädagogischen Handlungsfeldern häufig auf Personen und stellt die offizielle Bezeichnung für die Notengebung dar. Die Bezeichnung ist im Gegensatz zum Begriff Bewertung mit einer gewissen Endgültigkeit konnotiert. Ein ebenfalls häufig synonym gebrauchter Ausdruck ist die Diagnose. Eine Diagnose ist dadurch definiert, dass sie auf einer genauen und begründeten Fragestellung sowie einer theoriegeleiteten und kontrollierten Datenerhebung basiert. Im Gegensatz dazu finden Prozesse, bei denen sich Lehrer ein einfaches Bild über einen Schüler machen, also eine globale vor-diagnostische Einschätzung vornehmen, aufgrund derer dann schneller auf Handlungen reagiert werden kann, ständig statt. Sie stellen aber eher unhinterfragte Anschauungen und naive Verhaltenstheorien dar (Kleber, 1992).

Die nun beschriebenen Begrifflichkeiten werden in der vorliegenden Arbeit ebenfalls synonym verwendet sowie auch die beiden Begriffe Lehrereinschätzung und Lehrurteil.

Historisch bedeutsam zeigt sich die Entwicklung der pädagogischen Diagnostik hinsichtlich ihrer Selektionseigenschaft in der Gesellschaft. Wo früher die Ständegesellschaft jegliche soziale Mobilität verhinderte, entscheidet heute mit Hilfe der Einführung der allgemeinen Schulpflicht der schulische Leistungsnachweis über sozialen Aufstieg, und so reicht nun der Leistungsdruck hinunter bis in die Grundschule. Daher wurde die Schulleistung zum zentralen diagnostischen Kriterium und deren Beurteilung zunehmend problembehafteter (Kleber, 1992).

Schülerbezogene Leistungsbeurteilungen stellen einen wesentlichen Aspekt der Lehrertätigkeit dar (Schrader & Helmke, 1990). Man unterscheidet zwei verschiedene Formen von Beurteilung, die formelle und die informelle. Während sich die formelle Urteilsleistung des Lehrers auf die Notengebung bezieht, finden informelle Urteilsprozesse schon während des Unterrichts statt. Auch Thomas (2007) sieht das „alltägliche Diagnostizieren“ (S.85) des Lehrers eingebettet in die Komplexität des Unterrichtsgeschehens, welches nach dem Unterricht auch noch korrigierbar ist, und auch oft laut Kleber (1992) mit Beratungstätigkeiten gekoppelt ist.

Die Kompetenz eines Lehrers, einen Schüler hinsichtlich seines Leistungsstandes genau einschätzen zu können, stellt einerseits die Grundlage für die Planung und Gestaltung des pädagogischen Handelns (Schrader, 2006), wie z. B. die Unterrichtssteuerung, die Wahl der Unterrichtsmaterialien und Unterrichtstechniken (Clark & Peterson, 1986; Sharpley & Edgar, 1986), und eine optimale schulische Förderung dar (Deutsches PISA-Konsortium, 2001), andererseits die Voraussetzung für Schullaufbahnentscheidungen (Thomas, 2007). Die Akkuratheit von Lehrereinschätzungen ist also die Basis für effektive pädagogische Entscheidungen (Feinberg & Shapiro, 2003).

Aufgrund dieser enormen Bedeutung von Lehrerurteilen im Leben von Kindern und Jugendlichen gilt es nun zu beleuchten, welche Prozesse rund um das Urteil eines Lehrers ablaufen können.

Laut Kleber (1992) beruhen alle Urteile auf Beobachtung und Interpretation. Frühere Beobachtungen, systematisch erworbenes Wissen als auch Meinungen haben Einfluss auf das Urteil. Nur am Rande sei erwähnt, dass unter anderem der explizite theoretische Hintergrund und auch implizite Persönlichkeitstheorien sowie Fehler in der Wahrnehmung das Urteil verzerren.

Der Autor betrachtet den Prozess der Urteilsbildung in vier Stufen, welche von der Wahrnehmung bis hin zur Beurteilung reichen. Zu Beginn wird ein Wahrnehmungsurteil gebildet, welches durch die Bewertung des ausgewählten Wahrgenommenen entsteht. Darauf folgt ein Vorausurteil, welches im Falle der Entstehung durch eine stereotype Bewertung einem Vorurteil entspricht und einen stereotypen Urteilsprozess mit sich bringt. Auf der nächsten Stufe entsteht abermals durch Bewertung aus dem Vorausurteil das Urteil.

Folgt dieses wie eben beschrieben wiederum einer stereotypen Bewertung, so entsteht daraus ein ungeprüftes Vorurteil. Findet allerdings eine evaluative Bewertung im gesamten Urteilsprozess statt, so werden die Wahrnehmungsurteile hinterfragt. Im Weiteren kommt es zur Bildung eines echten Vorausurteils, welches zu beurteilende Hypothesen liefert und ein argumentatives Urteil zur Folge hat. Eine diagnostisch begründete Beurteilung ergibt sich dann durch weitere evaluative Bewertungen von argumentativen Urteilen und bleibt stets überprüfbar und vorläufig. Im Alltag kommt es häufig zu Vergleichsurteilen, welche den Vorurteilen sehr ähnlich sind.

Ähnlichkeiten mit Klebers Modell (1992) weisen in der Forschung zwei weitere gegensätzliche Standpunkte bezüglich der Urteilskompetenz des Lehrers und der damit verbundenen Prozesse auf.

Brophy & Good, (1974) vertreten den Ansatz, dass Lehrer sich bei der Beurteilung von Schülerleistung und -verhalten von stereotypen Erwartungen, welche aufgrund nur weniger Informationen gebildet werden, leiten lassen. Wie Rosenthal und Jacobson (1968) in ihrem berühmten „Oak-School“- Experiment zeigen konnten, beeinflussen diese Erwartungen des Lehrers das unterrichtsbezogene Handeln und können im Sinne eines Erwartungseffektes als sich selbsterfüllende Prophezeiungen auf die Schülerleistungen Einfluss nehmen. Für gewöhnlich sind Erwartungen das Ergebnis der Beobachtung des Lehrers, sind sie jedoch nicht flexibel, so können sie auch eine erwartungsgemäße Leistung verursachen (Brophy & Good, 1974).

Der gegensätzliche Standpunkt in der Forschung sieht den Lehrer als Unterrichtsexperten, dessen kognitive Fähigkeiten ihn dazu befähigen, die Herausforderungen des Unterrichts zu bewältigen (Bromme, 1987). Diese Expertise soll im Laufe der beruflichen Ausbildung und Erfahrung entstehen. Es wird eine spezifische Wissensstruktur ausgebildet, die aus organisierten und integrierten Kategorien besteht, und ihrerseits spezifische Verarbeitungsprozesse mit sich bringt (Dawson, Zeitz & Wright, 1989, zitiert nach Krolak-Schwerdt & Rummer, 2005). Die Kategorien bestehen im Spezialwissen über das Schulfach, in fachdidaktischem Wissen und Wissen um typische Unterrichtsepisoden als auch im Wissen über die Schülerpersönlichkeit. Was erfahrene Lehrer bzw. Experten von Laien laut Krolak-Schwerdt und Rummer (2005) unterscheidet, ist die Fähigkeit, situationsspezifisch zwischen top-down (kategorienbasierten) und

bottom-up (merkmalsbasierten) Verarbeitungsstrategien zu variieren. Die beiden Verarbeitungsstrategien unterscheiden sich dadurch, dass die top-down Verarbeitung schnell verläuft und zu geringer differenzierten Urteilen führt, die bottom-up Verarbeitung aufwendiger ist und sich auf Besonderheiten des spezifischen Falles konzentriert. Welche Strategie gewählt wird, hängt zum einen von der Kategorisierbarkeit der Merkmalsinformation ab, zum anderen vom Ziel, unter welchem die Merkmalsverarbeitung stattfinden soll. In ihrer Studie finden Krolak-Schwerdt und Rummer (2005) heraus, dass erfahrene Lehrer nur dann zu vorurteilsgeleiteten Urteilen neigen, wenn aufgrund des Verarbeitungsziels eine schnelle und ressourcenschonende Beurteilung gewählt wird.

Nachdem nun ein Überblick über die Grundlagen, Determinanten und Gefahren von Lehrurteilen und der damit verbundenen kognitiven Prozesse gegeben wurde, werden im Folgenden Lehrerurteile hinsichtlich ihrer Akkuratheit und die dazu vorliegenden Ergebnisse aus zahlreichen Studien dargestellt.

2.2 Akkuratheit von Lehrereinschätzungen

Unter dem Begriff Akkuratheit im Zusammenhang mit Leistungsbeurteilungen versteht man ein psychometrisches Konzept, welches die Abweichung des Urteils vom wahren Wert einer Leistung meint. Dienlich scheint es dann, wenn anhand von Genauigkeit die Qualität von Leistungsbeurteilungen gemessen werden soll (Schuler & Marcus, 2004).

Befasst man sich mit dem Thema Akkuratheit von Lehrereinschätzungen, so stößt man unweigerlich auf den Begriff der diagnostischen Kompetenz. Schrader (2006) definiert diagnostische Kompetenz als „... die Fähigkeit eines Urteilers, Personen treffend zu beurteilen. Sie ist damit Grundlage für die Genauigkeit diagnostischer Urteile oder Diagnosen“ (S.95). Als adäquates Kriterium zur Messung der Diagnosegenauigkeit eignet sich die Schülerleistung, weil sie valide und reliabel mittels Tests erhoben werden kann (Schrader 2009). Deshalb kann auch davon ausgegangen werden, dass diagnostische Kompetenz mittels Diagnosegenauigkeit gut erfasst werden kann.

Meist werden allerdings die beiden Begriffe Diagnosegenauigkeit bzw. Akkuratheit und diagnostische Kompetenz – wie auch in der vorliegenden Arbeit – gleichgesetzt (Helmke, Hosenfeld & Schrader, 2004).

Es gibt in der Forschung selten aber doch Interesse an der Akkuratheit von Lehrerurteilen, die sich auf nicht-kognitive wie motivationale und affektive Schülermerkmale (z. B. Ängstlichkeit und Lernmotivation) beziehen (Spinath, 2005; Helmke, 1994). Aufgrund der Tatsache, dass sich die vorliegende Studie mit Lehrereinschätzungen hinsichtlich kognitiver Schülermerkmale beschäftigt, wird auf diesen Aspekt nicht spezifisch eingegangen.

Zur Bestimmung der diagnostischen Kompetenz bzw. Akkuratheit von Lehrerurteilen verwendete Zielinski (1974, zitiert nach Schrader & Helmke, 1990) in den 70er Jahren noch das Notenurteil, welches nur bedingt zur Einschätzung der Urteilsfähigkeit eines Lehrers herangezogen werden kann. Die Notengebung des Lehrers berücksichtigt meist einen größeren Zeitraum wie auch verschiedene Aspekte der fachlichen Leistung als dies jene zur Messung der Schülerfähigkeiten herangezogenen Testleistungen tun. Diese Methode zur Erhebung von Lehrerurteilen zählt nach Feinberg und Shapiro (2003) zu den indirekten Methoden, da die Einschätzung des Lehrers und die Testleistung des Schülers nicht mit demselben Maß gemessen werden. Neben den Schulnoten zählen auch Rating-Skalen zu den indirekten Methoden (Begeny, Eckert, Montarello & Storie, 2008), welche auch in der vorliegenden Arbeit zur Erhebung der Lehrereinschätzungen verwendet werden. In anderen Studien (z. B. Hoge & Butcher, 1984; Graney, 2008; Feinberg & Shapiro, 2003) verwendet man anstelle von Noten die direkte Einschätzung der zum Vergleich herangezogenen Testleistung durch den Lehrer (Schrader & Helmke, 1990; vgl. Feinberg & Shapiro, 2003). Man vergleicht hierbei die Leistung des Schülers bei bestimmten Items mit der Einschätzung des Lehrers bezüglich der Schülerleistung bei denselben Items (Feinberg & Shapiro, 2003). Nach Begeny et. al. (2008) beziehen sich direkte Methoden zur Erhebung der Lehrereinschätzungen auf Kompetenzen, die die Schüler im täglichen Unterricht erlernen, wie z. B. Leseverständnis oder Lesegeschwindigkeit.

Weiters unterscheidet man zwischen peerabhängigen und peerunabhängigen Lehrereinschätzungen (Begeny et. al., 2008). Sollen Lehrer die Leistungen ihrer Schüler im Vergleich zu jenen ihrer Klassenkameraden einschätzen, so sprechen wir von peerabhängigen Einschätzungen.

Soll bei der Einschätzung aber das Leistungsniveau der Klassenkameraden unberücksichtigt bleiben, so sprechen wir von peerunabhängigen Lehrereinschätzungen.

Eine Methode, welche den direkten und peerunabhängigen Ansatz vereint, ist der CBM (Curriculum-Based Measurement) (Deno, 1985; Deno & Mirkin, 1977, zitiert nach Begeny et al., 2008). Der CBM misst die gelernte Buchstabier- und Lesefähigkeit, die schriftliche Ausdrucksfähigkeit und mathematische Kompetenzen der Schüler.

Die diagnostische Kompetenz von Lehrern kann anhand von drei Urteilskomponenten, wie sie Schrader und Helmke (1987) unterscheiden, untersucht werden. Die erste Komponente meint die Tendenz des Lehrers, die Merkmalsausprägungen der Schüler verglichen mit ihren tatsächlichen Ausprägungen zu unter- oder zu überschätzen, und wird als *Niveauekomponente* bezeichnet. Zur Errechnung der Niveauekomponente werden Mittelwerte anhand von t-Tests gegenübergestellt. Wird die Übereinstimmung zwischen der Streuung der Merkmalsausprägungen und jener der tatsächlichen Merkmalsausprägungen herangezogen, so handelt es sich um die *Differenzierungskomponente*. Die dritte Komponente ist die *Rangkomponente* oder Vergleichskomponente und betrachtet die Fähigkeit der Lehrer, die relativen Merkmalsausprägungen der Schüler im Vergleich zueinander richtig zu positionieren. Berechnet wird diese Komponente anhand von Produkt-Moment-Korrelationen oder Rangkorrelationen. Trotz der Tatsache, dass die Rangkomponente das absolute Niveau der Lehrerurteile bzw. Merkmalsausprägungen außer Acht lässt, stellt sie den zentralen Akkuratheitsindikator dar (Schrader & Helmke, 1987; Spinath, 2005).

Zur Erhebung der Akkuratheit werden, wie eben gezeigt, Distanzmaße und korrelative Zusammenhänge zwischen wahren Werten und Urteilen sowie Prozentwerte von bspw. korrekt klassifiziertem Verhalten verwendet.

Die Herausforderung dabei ist allerdings die Ermittlung des Wertes der wahren Leistung, welcher zur Berechnung herangezogen werden soll (Schuler & Marcus, 2004).

Zur Akkuratheit von Lehrereinschätzungen bzw. Lehrerurteilen gibt es sehr unterschiedliche Ansichten und Studienergebnisse in der Fachliteratur.

Hoge und Coladarci (1989) sprechen in ihrer Studie von einer moderaten bis hohen Akkuratheit von Lehrereinschätzungen. Mit ihrem „Review of Literature“ (S.297) präsentieren die Autoren eine Metaanalyse über 16 Studien, welche bei einer Spanne von Korrelationen zwischen .28 und .92 eine durchschnittliche Korrelation von .66 zeigt. In einer anderen Untersuchung mit einer Stichprobe von 12 Lehrern wurde das Lehrpersonal gebeten, die intellektuellen Fähigkeiten, die Arbeitsmotivation in der Schule und die Testergebnisse ihrer Schüler einzuschätzen (Hoge & Butcher, 1984). Der Vergleich ihrer Einschätzungen mit den tatsächlichen Werten eines Lesefähigkeitstest und eines Intelligenztests zeigte eine hohe Akkuratheit der Lehrereinschätzungen. Ebenso sehen Egan und Archer (1985) Lehrer grundsätzlich als gute Diagnostiker bezüglich der Englisch- und Mathematikleistungen ihrer Schüler an.

Weiters ergab sich auch beim Vergleich von tatsächlicher mündlicher Lesefähigkeit und Lehrereinschätzungen bezüglich der Lesefähigkeit, welche mittels CBM (siehe S.14) wie auch einer Rating-Skala erhoben wurden, in beiden Fällen ein akkurates Ergebnis (Feinberg & Shapiro, 2003). Allerdings erhärtete sich der Verdacht, dass sich die Lehrereinschätzungen bezüglich spezifischer Leistungsniveaus der Schüler nicht so großer Akkuratheit erfreuen dürften. Dieser Verdacht bestätigte sich bei Begeny et.al. (2008), welche die Akkuratheit von Lehrereinschätzungen bezüglich der Lesefähigkeit mittels direkten und indirekten Methoden erhoben. Als direkte Methode diente ihnen der CBM. Auch diese Studie zeigte eine moderate bis hohe Akkuratheit von Lehrereinschätzungen, unabhängig von der verwendeten Methode. Eine weitere Analyse ergab jedoch, dass Lehrer nicht akkurat einschätzen. Es gelingt ihnen zwar gute Leser zu identifizieren, jedoch haben sie Schwierigkeiten, schwache Leser herauszufiltern und überschätzen diese eher. Beim Vergleich von Lehrereinschätzungen mit etwas weniger spezifischen Leistungen – also allgemein mit Schülerleistungen - in einem standardisierten Leistungstest ergab sich ebenfalls die Eigenschaft der Lehrer, Schüler mit guten Leistungen am akkuratesten, Schüler mit schlechten Leistungen am wenigsten akkurat einzuschätzen (Coladarci, 1986). Ähnliche Ergebnisse fanden sich auch bei Demaray und Elliott (1998), da bei Verwendung von prozentualen Übereinstimmungen und Korrelationen auch hier Lehrer ihre guten Schüler akkurater einschätzten als ihre schwachen Schüler.

Allgemein fanden sie heraus, dass auch bei der Einschätzung der Schülerleistung anhand von Rating-Skalen wertvolle Informationen und akkurate Ergebnisse erzielt werden können.

Eine eher geringe Kompetenz der Lehrkräfte, ihre Schüler genau einzuschätzen, zeigt beispielsweise auch die PISA-Studie (Deutsches PISA-Konsortium, 2001, S.119f.), in welcher nur 11% der schwachen Leser von Hauptschullehrern richtig eingeschätzt wurden. Der Großteil der zur Risikogruppe zählenden Schüler, welche voraussichtlich massive Schwierigkeiten beim Übergang ins Berufsleben haben dürften, wird von den Lehrern nicht identifiziert. Weiters stellten Schrader und Helmke (1990) fest, dass es Lehrern schwer fällt, eine über den Bezugsrahmen der Schulklasse hinausgehende Leistungseinschätzung einzelner Schüler vorzunehmen. Sie sind also nicht in der Lage, das absolute Niveau der Klasse einzuschätzen, wobei hinzugefügt werden muss, dass in dieser Studie keine direkte Einschätzung des Klassenniveaus erhoben wurde. Bei der Untersuchung der Schülermerkmale Intelligenz, Fähigkeitsselbstwahrnehmung, schulische Lernmotivation und Leistungsängstlichkeit anhand der drei bereits beschriebenen Akkuratheitsindikatoren, erreichten die Lehrereinschätzungen ebenfalls nur geringe Akkuratheit, weshalb man nicht von einem Konstrukt der diagnostischen Kompetenz von Lehrern sprechen sollte (Spinath, 2005). Das Ergebnis einer neueren Studie von Schmidt und Schabmann (2010) war eine nur begrenzte Akkuratheit der Lehrereinschätzungen, und zwar speziell dann, wenn es sich um die Einschätzung von Kindern mit isolierten Leseschwierigkeiten handelt. Vor allem zu Beginn sind die Einschätzungen des Leistungsstandes ungenau und Probleme werden fälschlicher Weise als vorübergehend betrachtet. Im Laufe der Grundschulzeit bessert sich dies. Außerdem zeigt sich, dass die Lehrereinschätzung stärker von den Rechtschreibleistungen abhängt als von den Leseleistungen. Eine weitere Studie von Schabmann und Schmidt (2009) stützt das Ergebnis, dass Lehrer besonders am Anfang der ersten Klasse nur geringe Akkuratheit in ihren Einschätzungen zeigen.

Der Grund für unterschiedliche Ergebnisse bezüglich der Akkuratheit von Lehrereinschätzungen in der Forschung liegt in der Verwendung der bereits beschriebenen unterschiedlichen Methoden zur Erhebung der Lehrerurteile (vgl. Hoge und Coladarci, 1989) als auch zur Messung der Akkuratheit (vgl. Spinath, 2005).

Spinath (2005) sieht die Ursache für die widersprüchlichen Aussagen teilweise in der Verwendung unterschiedlicher Akkuratheitsmaße, welche unterschiedliche Einschätzungsfehler mit sich bringen, und teilweise im Fehlen von Richtlinien zur Bewertung der unterschiedlichen Akkuratheitsmaße.

Vergleiche zwischen indirekten und direkten Methoden der Erhebung der Lehrerurteile werden in der Metaanalyse von Hoge und Coladarci (1989, S.305) thematisiert. Die Spannbreite der Korrelationen zwischen Lehrerurteil und Schülerleistung bei indirekten Erhebungsmethoden reicht von .28 bis .86, mit einer mittleren Korrelation von .62. Im Gegensatz dazu befinden sich die Zusammenhänge in Studien mit direkten Messmethoden in einem Spektrum von .48 bis .92, mit einer durchschnittlichen Korrelation von .69. Feinberg und Shapiro (2003, S.61) erhalten in ihrer Untersuchung ähnliche Ergebnisse. Bei Verwendung der direkten Methode des CBM ergibt sich eine Korrelation von .70, bei der Verwendung von Rating-Skalen (indirekte Methode) erhalten sie einen Zusammenhang von .60.

Unterschiede ergeben sich ebenso bei der Art der Berechnung der Akkuratheit. Bei korrelativen Studien zum Vergleich von Lehrereinschätzungen mit normbezogenen Messwerten bezüglich der Schülerleistung (Hoge & Butcher, 1984; Hoge & Coladarci, 1989) ergibt sich eine größere Akkuratheit als bei Studien, die mit prozentualer Übereinstimmung von Lehrerurteilen arbeiten (Eckert, Dunn, Coddington, Begeny & Kleinmann, 2006; Feinberg & Shapiro, 2003). Letztere führen hingegen zu einem Gewinn an detaillierteren Informationen bezüglich der diagnostischen Kompetenz von Lehrern (Coladarci, 1986; Demaray & Elliot, 1998; Eckert et al., 2006). Geringere Akkuratheit ergibt sich auch, wenn Lehrer spezifische Fähigkeiten wie die Lesegeschwindigkeit von Schülern einschätzen sollen (Eckert et al., 2006; Feinberg & Shapiro, 2003).

Der Nachteil der korrelativen Studien besteht allerdings darin, dass sie das absolute Niveau der Schülerleistungen vernachlässigen und somit der Lehrer als akkurater Beurteiler erscheint, obwohl es sich eigentlich um systematische Über- oder Unterschätzungen handelt (Spinath, 2005). Auch andere Studien sprechen von einer Überschätzung der Akkuratheit der Lehrerurteile bei Verwendung dieses Ansatzes (vgl. Begeny et al., 2008; Eckert et al., 2006; Feinberg & Shapiro, 2003).

Trotz der Einschränkungen von korrelativen Analysen werden diese in der vorliegenden Studie zur Beurteilung der Akkuratheit von Lehrereinschätzungen herangezogen.

Zusammenfassend stellte Spinath (2005) fest, dass je stärker die schulischen Leistungen mit dem einzuschätzenden Merkmal einhergehen, desto größer sind die Zusammenhänge zwischen den Lehrereinschätzungen und den Schülermerkmalen.

Neben methodischen Aspekten finden sich im nächsten Kapitel noch weitere Ursachen für die Varianz der Forschungsergebnisse und die umstrittene Akkuratheit von Lehrereinschätzungen. Hier wird nun einerseits der Frage nachgegangen, welche Schülermerkmale, die nicht die Schülerleistung betreffen, auf das Lehrerurteil einen Einfluss üben. Diese Frage stellt den Schwerpunkt der vorliegenden Arbeit dar, da die Einflüsse auf die Akkuratheit von Lehrereinschätzungen den Mittelpunkt des Interesses bilden. Andererseits wird auch der Einfluss von Lehrermerkmalen aufgezeigt.

2.3 Einflüsse auf die Akkuratheit von Lehrereinschätzungen

Es gibt unterschiedliche Einflüsse auf die Akkuratheit von Lehrereinschätzungen. Einerseits können diese Einflüsse von Schülermerkmalen ausgehen, andererseits können sie von Lehrermerkmalen ausgehen.

Bezüglich der Lehrermerkmale sei nur ein kurzer Überblick gewährt, da dies für die aktuelle Studie nicht von zentralem Interesse ist. Beispielsweise wurde in der Studie von Demaray und Elliott (1998) der Einfluss der Lehrererfahrung sowie des Ausbildungsgrades eines Lehrers untersucht. Es zeigte sich jedoch kein systematischer Einfluss der beiden Variablen auf die Genauigkeit eines Lehrerurteils.

Krolak-Schwerdt und Rummer (2005) überprüften den Einfluss der Verarbeitungsstrategien von Schülermerkmalen auf die Leistungsbeurteilung bei Lehrern im Vergleich zu Laien. Man stellte fest, dass Lehrer (Experten) je nach Verarbeitungsziel (Prognose künftiger Schülerleistungen oder lediglich eine Eindrucksbildung) unterschiedliche Verarbeitungsstrategien anwenden. Bei Laien fanden sich keine diesbezüglichen Unterschiede, was bedeutet, dass die Experten über eine spezifische Urteilskompetenz verfügen.

Im Folgenden werden einzelne Schülermerkmale, die einen Einfluss auf die Akkuratheit von Lehrereinschätzungen haben, eingehend dargestellt.

2.3.1 Schülermerkmale

Trotz der Tatsache, dass dem Lehrerurteil beachtliche Validität zukommt, gibt es einen erheblichen Anteil an Varianz, der unerklärt bleibt (Egan & Archer, 1985). Daraus kann geschlossen werden, dass bei der Beurteilung der Schülerleistungen auch andere Variablen eine Rolle spielen, die zu systematischen Fehlern in der Beurteilung der Schülerleistung führen (Schrader & Helmke, 1990). Es gibt eine Reihe von Schülermerkmalen, die einen Einfluss auf die Einschätzung des Lehrers bezüglich der Leistung seiner Schüler haben, wie z. B. Leistungsniveau, Intelligenz, Geschlecht und Verhalten (z. B. Demaray & Elliott, 1998; Hecht & Greenfield, 2002; Hoge & Butcher, 1984; Schabmann & Schmidt, 2009; Schrader & Helmke 1990; Tournaki & Podell, 2005).

Besonders der Einfluss des Verhaltens auf die Leistungseinschätzungen der Lehrer ist in der aktuellen Studie von Relevanz.

2.3.1.1 Intelligenz und Leistungsniveau

Das Schülermerkmal Intelligenz scheint einen nicht unerheblichen Einfluss auf die Lehrereinschätzungen bezüglich der Schülerleistungen zu haben. Hoge & Butcher, (1984) erhoben die Schülerfähigkeit mittels eines Intelligenztests und stellten fest, dass sie zur Verzerrung so mancher Lehrereinschätzung beiträgt, indem diese Lehrer Schüler mit geringerer Intelligenz unterschätzten und jene mit höheren Werten im Intelligenztest überschätzten. Auch bei Schrader und Helmke (1990) ergeben sich umso günstigere Leistungseinschätzungen, je höher die Intelligenz des Schülers ausgeprägt ist, und umso schlechtere Einschätzungen, je geringer die Intelligenz ist.

Bezüglich der Einschätzung des Leistungsniveaus in der mündlichen Lesefähigkeit teilten Begeny et. al (2008) die Grundschüler der zweiten und dritten Klasse anhand von Leseproben in drei verschiedene Niveaustufen bezogen auf die Klassenstufe ein: „frustrational“ (schlechte Lesefähigkeit), „instructional“ (durchschnittlich) und „mastery“ (sehr gut). Die Lehrer zeigten äußerst akkurate Einschätzungen bei Schülern des „mastery“-Levels, konnten aber nicht einmal die Hälfte jener Schüler identifizieren, welche sich auf dem „instructional“- oder „frustrational“-Level befanden.

Widersprüchliche Aussagen dazu finden sich bei Eckert et. al (2006), welche herausfanden, dass Lehrer jene Schüler mit schlechten Lesefähigkeiten („frustrational“-Level) besser einschätzen können als Schüler mit durchschnittlichen oder sehr guten Lesefähigkeiten.

2.3.1.2 Geschlecht

Mit dem Einfluss des Geschlechts auf die Lehrereinschätzung haben sich unter anderem Schrader und Helmke (1990) beschäftigt. Bei ihnen zeigt sich kein signifikanter Einfluss des Geschlechts auf die Lehrereinschätzung. Auch Demaray und Elliott (1998) sowie Hoge und Butcher (1984) unterstützen dieses Ergebnis hinsichtlich des Geschlechts der Schüler. Entgegen dieser Ergebnisse legt die Untersuchung von Bennett, Gottesmann, Rock und Cerullo (1993) einen indirekten Einfluss des Schülergeschlechts auf die Lehrereinschätzung hinsichtlich der Schülerleistung offen. Das durch den Lehrer wahrgenommene Verhalten der Schüler zeigte einen Effekt auf die Einschätzung hinsichtlich der Schülerkompetenzen. Da jenes durch den Lehrer wahrgenommene Verhalten eines Buben oft problematischer eingestuft wurde als jenes der Mädchen, wurden Buben auch öfter negativer hinsichtlich ihrer Leistungsfähigkeit beurteilt. Im Gegensatz dazu moderiert das Schülergeschlecht in der Studie von Schabmann und Schmidt (2009) den Einfluss des Verhaltens auf die Lehrereinschätzungen nur geringfügig.

Tournaki und Podell (2005) wiederum konstatieren positivere Vorhersagen des schulischen Erfolgs für Mädchen. Auch in der Erwachsenenbildung hat das Geschlecht eine Wirkung auf die Bewertung durch die Lehrer (Bernard, 1979). Das geschlechtsspezifische Rollenverhalten zeigt einen positiven Einfluss auf die Bewertung schriftlicher Leistungen an der Universität von Melbourne.

Männer werden in ihren Leistungen positiver bewertet, wenn sie ein maskulines Rollenverhalten zeigen, also ein typisches Studium für Männer belegen, wie bspw. Physik. Gleiches zeigte sich in dieser Studie auch für Frauen.

2.3.1.3 Verhalten

Wie auch die Variablen Geschlecht und Leistungsniveau bzw. Intelligenz hat laut Brophy und Good (1974, S.17-20) das durch den Lehrer wahrgenommene Verhalten in der Klasse einen direkten Einfluss auf seine Leistungseinschätzungen.

Der Einfluss des Verhaltens zeigt sich beispielsweise in einem Experiment von Tournaki und Podell (2005). 384 Lehrern wurden Beschreibungen des Geschlechts, der Lesefähigkeit, des Sozialverhaltens und der Aufmerksamkeit eines Schülers zur Vorhersage des schulischen Erfolgs vorgelegt. Die Studie zeigte, dass Unaufmerksamkeit eher bei freundlichen Schülern toleriert wird als in Kombination mit aggressivem Verhalten. Dies trifft vor allem dann zu, wenn es sich um durchschnittlich gute Schüler handelt. Bei schwachen Schülern gibt es keine derartigen Unterschiede in den Vorhersagen des schulischen Erfolgs. Die schlechtesten Vorhersagen bekamen aggressive und unaufmerksame Schüler mit schwachen Lesefähigkeiten.

Die Längsschnittstudie von Hecht und Greenfield (2002) kam zu dem Ergebnis, dass die Lehrereinschätzungen aus Verhalten und Fähigkeiten resultieren. Die Stichprobe setzte sich aus 170 unterprivilegierten Schülern zusammen, welche aus einer ersten, zweiten oder dritten Klasse Grundschule stammten. Bezüglich der Vorhersagegenauigkeit der künftigen Lesefähigkeiten war in der ersten Klasse deutlich zu sehen, dass die Lesefähigkeit den größten Teil der Varianz der künftigen Lesefähigkeit erklärte und somit den größten Einfluss ausübte.

Eine der für die vorliegende Untersuchung äußerst interessante Studie stammt von Bennett et. al (1993) und wird nun etwas genauer beschrieben. Die herangezogene Stichprobe setzt sich aus 794 Kindern vom Kindergarten bis hin zur zweiten Klasse Grundschule zusammen. Mittels der Methode von Strukturgleichungsmodellen wird unter anderem der Einfluss des wahrgenommenen Verhaltens auf die schulische Leistungsbeurteilung erhoben. Die Autoren gehen davon aus, dass die tatsächlichen Schülerleistungen einerseits eine direkte Wirkung auf die Lehrereinschätzungen haben, andererseits indirekt über die Mediatorvariable Verhalten auf die Lehrereinschätzungen wirken.

Die Pfadmodelle zeigen, dass das wahrgenommene Verhalten einen direkten Einfluss auf die Lehrereinschätzungen hat, jedoch entgegen der Hypothese keine Mediatorrolle zwischen Testleistung und Lehrereinschätzung einnimmt. An diesem Ergebnis orientiert sich auch die vorliegende Untersuchung, da sie unter anderem den direkten Einfluss bestimmter Verhaltensweisen auf die Beurteilungen der Lehrer untersucht.

Auch Schabmann und Schmidt (2009) zeigen, dass das Verhalten am Anfang der ersten Klasse Grundschule die Lehrereinschätzungen mindestens genauso stark bedingt wie die Testleistungen im Bereich der Lese- und Rechtschreibkompetenz, welche nur gering mit den Einschätzungen zusammenhängen. Als bedeutsames Verhalten betrachten sie primär externalisierende Verhaltensweisen (oppositionelles Verhalten, Aggression) und Hyperaktivität. Zwei der drei übrigen Verhaltenskategorien (Viktimisierung und Depression) zeigen keine signifikanten Effekte auf die Lehrereinschätzungen, und die Kategorie sozialer Rückzug hat einen nur geringen Einfluss auf die Lehrereinschätzungen, vor allem beim Rechtschreiben. Den größten Einfluss auf die Einschätzungen zeigt allerdings die Variable Hyperaktivität. Ebenso signifikant zeigt sich ihre Wirkung auf die Leseleistungen. Die beiden anderen Verhaltensausrägungen haben keine direkten Effekte auf die Leistungen der Schüler.

In Anlehnung an die Untersuchung von Bennett et. al. (1993) und Schabmann und Schmidt (2009) werden in der vorliegenden Studie die Akkuratheit der Lehrereinschätzungen untersucht sowie auch der Einfluss des durch den Lehrer wahrgenommenen Verhaltens. Wie Bennett et. al. (1993) bereits zeigen konnte, stellt das Verhalten der Schüler keine Mediatorvariable zwischen Leistung und Lehrereinschätzung dar, sondern hat eine direkte Wirkung auf die Lehrerurteile. Weiters wird auch jene Erkenntnis herangezogen, die Schabmann und Schmidt (2009) machten, als sie einen direkten Effekt des wahrgenommenen hyperaktiven Verhaltens auf die Leseleistungen feststellten. Außerdem werden, ähnlich wie bei diesen Autoren, nur die Verhaltenskategorien oppositionelles Verhalten, Aggression und Hyperaktivität als externalisierende Verhaltensweisen herangezogen.

3 Zielsetzung der Untersuchung

In der folgenden Untersuchung wird grundsätzlich der Frage nachgegangen, wie gut Lehrereinschätzungen durch Schülerleistungen über einen Zeitraum von 4 Jahren Grundschule erklärbar sind. Kernpunkt der Studie ist die Frage, wie akkurat Lehrereinschätzungen sind, und ob die Wahrnehmung des Schülerverhaltens durch den Lehrer einen Einfluss auf seine Leistungseinschätzungen sowie auch auf die Leistungen der Schüler hat. Dabei werden die Bereiche Lesesicherheit, Lesegeschwindigkeit und Rechtschreibe-sicherheit getrennt voneinander betrachtet.

Seit geraumer Zeit beschäftigt sich die Forschung mit dem Thema Akkuratheit von Lehrereinschätzungen und daher gibt es, wie bereits erwähnt, zahlreiche Studien zu diesem Forschungsgebiet. Manche sehen Lehrereinschätzungen als äußerst akkurat an (Hoge & Coladarci, 1989; Hoge & Butcher, 1984; Egan & Archer, 1985), andere wiederum sprechen von einer eher geringen Akkuratheit, die sich besonders auf die Identifikation von schwachen Schülern bezieht (Deutsches PISA-Konsortium, 2001; Schabmann & Schmidt, 2009; Schmidt & Schabmann, 2010; Schrader & Helmke, 1990; Spinath, 2005). Die Unterschiedlichkeit der Studienergebnisse hat in erster Linie methodische Gründe (Hoge & Coladarci, 1989; Spinath, 2005).

Inhaltlich betrachtet liegen die Ursachen für die geringe Akkuratheit aber vor allem darin, dass sich Lehrereinschätzungen nicht nur aus den tatsächlichen Schülerleistungen bilden, sondern auch andere Schülermerkmale, wie z. B. Leistungsniveau, Intelligenz, Geschlecht und Verhalten, einen Einfluss auf das Urteil eines Lehrers haben (z. B. Demaray & Elliott, 1998; Hecht & Greenfield, 2002; Hoge & Butcher, 1984; Schabmann & Schmidt, 2009; Schrader & Helmke 1990; Tournaki & Podell, 2005).

Besonders interessant für die vorliegende Untersuchung sind die Ergebnisse der Studien von Bennett et. al. (1993) als auch von Schabmann und Schmidt (2009). Bennett et. al. (1993) überprüften die Hypothese, ob das Verhalten eine Mediatorfunktion zwischen Schülerleistungen und Lehrereinschätzungen übernimmt. Das Ergebnis der Studie zeigt, dass die Schülerleistung direkt auf die Lehrereinschätzung wirkt und nicht über die Variable Verhalten vermittelt wird.

Weiters konnten Schabmann und Schmidt (2009) den direkten Einfluss von hyperaktivem Verhalten auf die Leistungen der Schüler im Lesen belegen sowie auch den direkten Einfluss des Verhaltens auf die Lehrereinschätzungen. Der direkte Effekt der tatsächlichen Leistungen auf die Lehrereinschätzungen – die Akkuratheit - fiel gering aus.

Das Besondere der aktuellen Untersuchung ist der Zeitraum der längsschnittlichen Betrachtung von 4 Jahren Grundschule, da keine Längsschnittuntersuchungen von vergleichbarer Dauer bezüglich der Akkuratheit vorliegen. Aufgrund des Zeitraumes von vier Jahren können umfassende Einsichten in die Akkuratheit von Lehrerurteilen gewonnen werden, da sie zu vier verschiedenen Zeitpunkten betrachtet wird.

Zentrale Fragestellung für Lesesicherheit

- Wie akkurat sind die Lehrereinschätzungen bezüglich der Lesesicherheit über die 4 Jahre Grundschule hinweg?
- Welchen Einfluss üben die durch den Lehrer wahrgenommenen externalisierenden Verhaltensweisen auf die Lehrereinschätzungen bezüglich der Lesesicherheit in der Grundschule aus?
- Welchen Einfluss haben externalisierende Verhaltensweisen auf die Schülerleistungen im Bereich Lesesicherheit?

Zentrale Fragestellung für Lesegeschwindigkeit

- Wie akkurat sind die Lehrereinschätzungen bezüglich der Lesegeschwindigkeit über die 4 Jahre Grundschule hinweg?
- Welchen Einfluss üben die durch den Lehrer wahrgenommenen externalisierenden Verhaltensweisen auf die Lehrereinschätzungen bezüglich der Lesegeschwindigkeit in der Grundschule aus?
- Welchen Einfluss haben externalisierende Verhaltensweisen auf die Schülerleistungen im Bereich Lesegeschwindigkeit?

Zentrale Fragestellung für Rechtschreibsicherheit

- Wie akkurat sind die Lehrereinschätzungen bezüglich der Rechtschreibsicherheit über die 4 Jahre Grundschule hinweg?
- Welchen Einfluss üben die durch den Lehrer wahrgenommenen externalisierenden Verhaltensweisen auf die Lehrereinschätzungen bezüglich der Rechtschreibsicherheit in der Grundschule aus?
- Welchen Einfluss haben externalisierende Verhaltensweisen auf die Schülerleistungen im Bereich Rechtschreibsicherheit?

4 Methode

Der folgende Abschnitt beschreibt einerseits die Zusammensetzung der Stichprobe und andererseits werden die Instrumente, welche zur Erhebung der Daten verwendet wurden, dargestellt.

4.1 Stichprobe

Die vorliegende Untersuchung beruht auf einer Längsschnittstudie der Fakultät für Psychologie der Universität Wien. Die Erhebung fand in 15 niederösterreichischen Grundschulen statt, es nahmen 32 Klassen und insgesamt 293 Schüler daran teil. Die Untersuchung gliedert sich in zwei Kohorten, eine Kohorte besteht aus Schülern deren Einschulungsjahr 1998/99 war, die andere Kohorte wurde ein Jahr später, 1999/00, eingeschult. Die Längsschnittstudie erstreckt sich in Abhängigkeit von den beiden Einschulungszeitpunkten über jeweils 4 Jahre und läuft somit über die gesamte Grundschulzeit hinweg. Die Daten wurden jeweils im Dezember der ersten Klasse Grundschule und im Juni der ersten, zweiten Klasse und vierten Klasse Grundschule erhoben. Es wurden also zu Beginn der ersten Klasse (A1) sowie am Ende der ersten (E1), zweiten (E2) und vierten Klasse (E4) Grundschule Datenerhebungen durchgeführt.

Hinsichtlich der Geschlechterverteilung kann festgestellt werden, dass mehr Buben ($n=163$) als Mädchen ($n=130$) teilnahmen, was einer prozentuellen Verteilung von 56,6% zu 44,4% entspricht.

Die Stichprobe der aktuellen Studie besteht aus 179 Kindern. Dies ergibt sich einerseits aufgrund von fehlenden Werten, andererseits wurden Schüler aus Klassen ausgeschlossen, die nach einem ganzheitlichen Konzept unterrichtet werden, da es zu unterschiedlichen Entwicklungsverläufen des Lesens bei synthetischem und ganzheitlichem Unterricht am Anfang der Grundschule kommt (Schabmann, 2001).

4.2 Erhebungsinstrumente

Für die Erhebung der tatsächlichen Schülerleistungen sowie auch der Lehrereinschätzungen dieser wurden unterschiedliche Instrumente verwendet, die im Folgenden beschrieben werden.

4.2.1 Schülerleistungen im Lesen und Rechtschreiben

Die Erhebungsinstrumente bezüglich der Schülerleistungen werden getrennt nach den Bereichen Lesen und Rechtschreiben beschrieben.

4.2.1.1 Lesefähigkeiten

Zur Erhebung der Lesefähigkeiten am Anfang der ersten Klasse Grundschule (A1) wurde der Wiener Früherkennungstest von Klicpera, Humer, Gasteiger-Klicpera und Schabmann (2008) herangezogen. Dieser Test besteht aus bekannten Buchstaben und Wörtern, neuen (unbekannten) und Pseudowörtern, wobei die Anzahl der richtig gelesenen Wörter kodiert wurde. Die Wörter waren auf das verwendete Lehrbuch abgestimmt.

Am Ende der ersten (E1), zweiten (E2) und vierten (E4) Klasse wurde der „Individuelle Lesetest“ (Klicpera, Gasteiger-Klicpera & Schabmann, 1993) zur Erfassung der Lesefähigkeiten eingesetzt. Den Schülern wurden 6 Wortlisten zu je 15 Wörtern vorgegeben, welche aus jeweils kurzen und langen, häufigen (enger Grundwortschatz), seltenen (erweiterter Grundwortschatz) und Pseudowörtern (sinnlose Silben mit hoher orthographischer Regularität) bestehen (Klicpera et. al., 1993, S.60). Die verwendeten Wörter des „Individuellen Lesetests“ orientierten sich ebenfalls an den im Unterricht bereits durchgenommenen Wörtern. Die Aufgabe der Kinder war es, die Wörter fehlerfrei und so schnell wie möglich vorzulesen. Erhoben wurde einerseits wie viele Wörter richtig gelesen wurden, andererseits wie viele richtige Wörter pro Minute vom Schüler laut vorgelesen wurden.

4.2.1.2 Rechtschreibfähigkeiten

Zur Erhebung der Rechtschreibfähigkeiten wurde ebenfalls ein „Individueller Rechtschreibtest“ von Klicpera et. al. (1993) zusammengestellt.

Am Anfang der ersten Klasse Grundschule wurden den Schülern sowohl bekannte Buchstaben und Wörter als auch neue sowie auch Pseudowörter, die sie zuvor als Leseaufgabe erhalten hatten, diktiert. Die richtig geschriebenen Wörter wurden dann zur Analyse herangezogen.

Am Ende der ersten Klasse wurden ebenfalls die richtig geschriebenen bekannten, neuen und Pseudowörter ermittelt.

Für die letzten beiden Zeitpunkte, E2 und E4, zog man den Salzburger Lese- und Rechtschreibtest (Landerl, Wimmer & Moser, 1997) heran, wobei wiederum die richtig geschriebenen Wörter von Bedeutung waren.

4.2.2 Lehrereinschätzungen von Schülerleistungen und Schülerverhalten

Folgend werden die Erhebungsinstrumente bezüglich der Lehrereinschätzungen getrennt nach Schülerleistungen (Lese- und Rechtschreibfähigkeiten) und Schülerverhalten beschrieben.

4.2.2.1 Lese- und Rechtschreibfähigkeiten

Bezüglich der Lese- und Rechtschreibfähigkeiten wurden den Lehrern Fragebögen mit einer 5-stufigen Rating Skala vorgegeben, wobei die Skala von 0 (stimmt nicht) bis 4 (stimmt genau) reichte.

Es handelt sich hier um indirekte und peerunabhängige Lehrereinschätzungen, welche eine allgemeine Einschätzung in den Bereichen Lesen und Rechtschreiben verlangt und die Lehrer auffordert, die Schüler unabhängig von ihren Klassenkameraden zu beurteilen.

Am Anfang der ersten Klasse Grundschule wurden die Lehrereinschätzungen im Bereich Lesesicherheit durch die Beurteilung von 6 Items mittels der 5-stufigen Rating Skala erhoben.

Die Lehrereinschätzungen im Dezember des ersten Schuljahres zum Bereich Rechtschreibsicherheit wurden durch die Einschätzung von 6 Items (vgl. Tabelle 4, S. 36) auf der 5-stufigen Rating Skala ermittelt.

Zu den anderen Erhebungszeitpunkten - Ende erste, zweite und vierte Klasse - wurden jeweils Lesesicherheit, Lesegeschwindigkeit und Rechtschreibsicherheit nur allgemein mit einem Rating zwischen den Werten 0 und 4 eingeschätzt.

Zur Lesegeschwindigkeit liegen zum ersten Erhebungszeitpunkt keine Daten vor.

4.2.2.2 Schülerverhalten

Wie auch die Schülerleistungen wurde das Schülerverhalten von den Lehrern anhand eines Fragebogens mit einer 5-stufigen Rating Skala zu allen Erhebungszeitpunkten eingeschätzt, wobei ein Wert von 0 für „stimmt nicht“ und ein Wert von 4 für „stimmt genau“ stand. Die Lehrer beurteilten 26 Items, wovon 17 zu den Variablen oppositionelles Verhalten, Aggression, Hyperaktivität, sozialer Rückzug, Depression und Viktimisierung zusammengefasst wurden (vgl. Schabmann & Schmidt, 2009). In Anlehnung an Schabmann und Schmidt (2009), die zeigten, dass vier dieser Variablen einen Einfluss auf das Lehrerurteil haben, werden aufgrund einer Faktorenanalyse (siehe Tabelle 6) die Variablen oppositionelles Verhalten, Aggression und Hyperaktivität zu einem Faktor, dem externalisierenden Verhalten, zusammengefasst.

5 Statistische Analyse

Das Kapitel der Statistischen Analyse befasst sich mit der Beschreibung des verwendeten statistischen Verfahrens, dem Zustandekommen der einzelnen Skalen und den Kriterien zur Modellbeurteilung.

5.1 Beschreibung des verwendeten statistischen Verfahrens

Strukturgleichungsmodelle dienen nach Weiber und Mühlhaus (2010) zur statistischen Analyse von a-priori-Fragestellungen. Sie stellen ein komplexes Ursache-Wirkungsgefüge zwischen manifesten und/oder latenten Variablen dar, welches aus mehreren wissenschaftlichen Hypothesen besteht. Die Komplexität ergibt sich dadurch, dass mehrere Kausalhypothesen gleichzeitig betrachtet werden können. Abhängige (endogene) und unabhängige (exogene) Variablen stehen in einer zeitlichen Abfolge zueinander und bilden ein lineares Gleichungssystem ab. Es handelt sich hierbei also um die Überprüfung kausaler Zusammenhänge, wobei Richtung und Stärke der Wirkungszusammenhänge geschätzt werden. Auch Messfehler finden in dieser Modellprüfung Berücksichtigung, die zum einen Messungenauigkeiten und zum anderen nicht berücksichtigte Einflussgrößen auf die abhängige Variable beinhalten. Die Pfade zwischen den latenten Dimensionen sind Effektgrößen der vermuteten Kausalbeziehungen. Die Schätzung der Modellparameter soll zeigen, wie gut die zu den Variablen erhobenen Daten im Strukturgleichungsmodell reproduziert werden.

Die Berechnungen der statistischen Analyse wurden mit dem Programm AMOS 17.0 (Analysis of Moment Structures) von James L. Arbuckle durchgeführt.

Die in der Ergebnisdarstellung (siehe S. 40) abgebildeten Strukturgleichungsmodelle basieren auf jenem theoretischen Modell, welches nun im Folgenden kurz skizziert werden soll (siehe Abbildung 1).

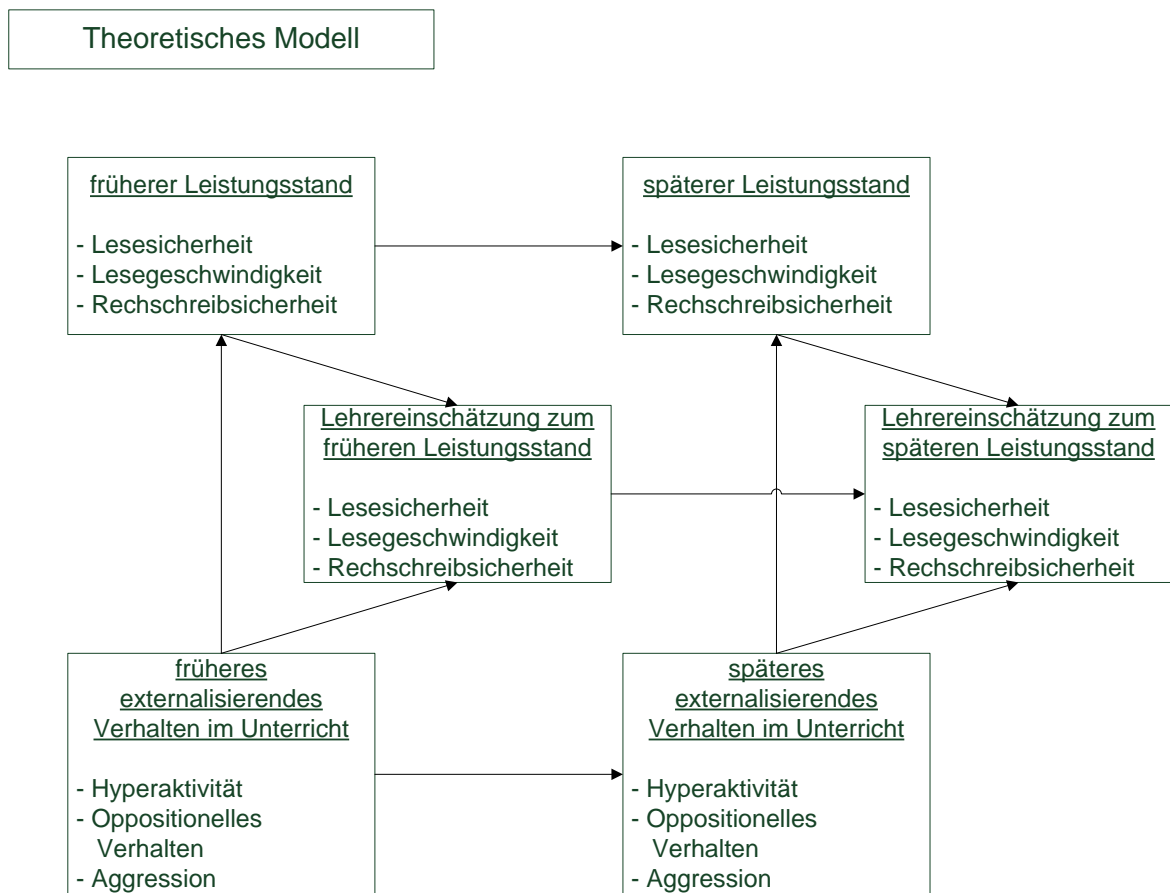


Abbildung 1: Theoretisches Modell

Theoretisches Modell über den Zusammenhang zwischen Lehrereinschätzungen und dem tatsächlichen Leistungsstand in den Lese- und Rechtschreibleistungen sowie den durch Lehrer im Unterricht wahrgenommenen externalisierenden Verhaltensauffälligkeiten

Wie anhand des theoretischen Teils (siehe 2.3) der Arbeit ersichtlich wurde, entstehen Lehrerurteile nicht nur aus der zur Kenntnisnahme des wahren Leistungsstandes ihrer Schüler, sondern spiegeln auch andere Einflussfaktoren wieder. Das vorliegende theoretische Modell (siehe Abbildung 1) soll zeigen, dass die Lehrereinschätzungen in den Bereichen Lesesicherheit, Lesegeschwindigkeit und Rechtschreibsicherheit einerseits von den tatsächlichen Schülerleistungen (siehe Pfeil von früherem bzw. späterem Leistungsstand zu Lehrereinschätzung zu früherem bzw. späterem Leistungsstand), andererseits auch vom durch den Lehrer im Unterricht wahrgenommenen externalisierenden Verhalten der Kinder abhängen (siehe Pfeil von früherem bzw. späterem externalisierendem Verhalten im Unterricht zu Lehrereinschätzung zu früherem bzw. späterem Leistungsstand).

Außerdem hat das externalisierende Verhalten auch einen Einfluss auf die Leistungen der Schüler (siehe Pfeil von früherem bzw. späterem externalisierendem Verhalten im Unterricht zu früherem bzw. späterem Leistungsstand).

Darüber hinaus hängen spätere Leistungen von früheren Leistungen ab (siehe Pfeil von früherem Leistungsstand zu späterem Leistungsstand) sowie früheres Verhalten auch späteres Verhalten erklärt (siehe Pfeil von früherem externalisierendem Verhalten im Unterricht zu späterem externalisierendem Verhalten im Unterricht) und ebenso bspw. am Anfang der ersten Klasse getroffene Lehrerurteile auch Lehrerurteile am Ende der ersten Klasse beeinflussen (Pfeil von Lehrereinschätzung zu früherem Leistungsstand zu Lehrereinschätzung zu späterem Leistungsstand).

In Abbildung 1 wird also von „früher“ und „später“ bezüglich Lehrereinschätzung, Leistungsstand und Verhalten gesprochen. Damit ist längsschnittliche Betrachtung über vier Zeitpunkte während der Grundschule gemeint (Anfang und Ende der ersten Klasse sowie Ende der zweiten und vierten Klasse). Spätere Lehrereinschätzungen sind wie bereits erwähnt von früheren Lehrereinschätzungen abhängig. Selbiges trifft auf die Schülerleistungen sowie auf das externalisierende Verhalten zu.

Es soll untersucht werden, wie sich dieses Bedingungsgefüge über die vier Zeitpunkte hinweg entwickelt, wobei jede Dimension der Leistung – Lesesicherheit, Lesegeschwindigkeit und Rechtschreibeisicherheit - für sich alleine betrachtet wird. Die Akkuratheit der Lehrereinschätzungen zeigt sich dadurch, wie gut die Lehrereinschätzungen durch die tatsächlichen Leistungen erklärt werden können. Der Einfluss des Verhaltens zeigt sich durch seine Effektstärke auf die Leistungen und die Lehrereinschätzungen.

5.2 Bildung der Skalen

Die einzelnen Skalen der Strukturgleichungsmodelle wurden durch die Berechnung von Summenscores aus den einzelnen manifesten Variablen gebildet. Nachfolgend wird nun jede Skala vorgestellt sowie die Items, aus denen sie gebildet wurde. Eine Skala wurde durch jene Variablen gebildet, welche hoch in einem Faktor laden. Nur Variablen, die anhand einer Faktorenanalyse - mittels der Extraktionsmethode der Hauptkomponentenanalyse - die Eindimensionalität der Skala bestätigen, werden in die Skala aufgenommen.

5.2.1 Bildung der Skalen zur Schülerleistung

Bildung der *Skala Leistung-Lesesicherheit* am Beginn der ersten Klasse

(*Leistung A1*):

Wie in Tabelle 1 ersichtlich, wird zuerst eine Faktorenanalyse (Hauptkomponentenanalyse) mit den manifesten Variablen Prozentrang richtig gelesene bekannte, neue und Pseudowörter, welche aus den Items der Lesetests gebildet wurden, gemacht, um deren Eindimensionalität zu überprüfen. Es zeigt sich, dass diese Variablen hoch in einem Faktor laden und somit zur Bildung der Skala herangezogen werden können. Im Weiteren wird der Summenscore über diese drei manifesten Variablen gebildet, welcher dann als *Skala Leistung Lesesicherheit A1* bezeichnet wird.

Bildung der *Skalen Leistung-Lesesicherheit* am Ende der ersten Klasse (*Leistung E1*), am Ende der zweiten Klasse (*Leistung E2*) und Ende der vierten Klasse (*Leistung E4*):

Zur Bestätigung der Eindimensionalität der manifesten Variablen Prozentrangwerte richtig gelesener häufiger, seltener und Pseudowörter (kurz und lang) jeweils zu den Zeitpunkten am Ende der ersten, zweiten und vierten Klasse werden Faktorenanalysen (Hauptkomponentenanalysen) verwendet (vgl. Tabelle 1). Dadurch zeigt sich, dass diese manifesten Variablen zur Bildung der *Skalen Leistung E1, Leistung E2 und Leistung E4* herangezogen werden können. Dies wird wiederum mittels Summenscores gemacht.

Tabelle 1: Resultat der Faktorenanalyse über den Faktor Leistung Lesesicherheit am Anfang und Ende der ersten sowie am Ende der zweiten und vierten Klasse

Skala Leistung-LESESICHERHEIT	Manifeste Variablen	Faktor
		1
Leistung A1	PR richtig gelesene bekannte Wörter	.77
	PR richtig gelesene neue Wörter	.90
	PR richtig gelesene Pseudo-Wörter	.89
Leistung E1	PR richtig gelesene häufige Wörter (kurz/lang)	.93
	PR richtig gelesene seltene Wörter (k/l)	.95
	PR richtig gelesene Pseudo-Wörter (k/l)	.91
Leistung E2	PR richtig gelesene häufige Wörter (k/l)	.79
	PR richtig gelesene seltene Wörter (k/l)	.88
	PR richtig gelesene Pseudo-Wörter (k/l)	.86
Leistung E4	PR richtig gelesene häufige Wörter (k/l)	.81
	PR richtig gelesene seltene Wörter (k/l)	.85
	PR richtig gelesene Pseudo-Wörter (k/l)	.86

Bildung der *Skala Leistung-Lesegeschwindigkeit* am Ende der ersten Klasse (*Leistung E1*):

Es wird vorerst mittels einer Faktorenanalyse (Hauptkomponentenanalyse) die Eindimensionalität der manifesten Variablen häufige, seltene und Pseudowörter pro Minute überprüft, was sich auch in Tabelle 2 bestätigt. Die *Skala Leistung Lesegeschwindigkeit E1* wird aus dem Summenscore über diese drei manifesten Variablen gebildet.

Ebenso werden auch die *Skalen Leistung E2* und *Leistung E4* gebildet (siehe Tabelle 2).

Tabelle 2: Resultat der Faktorenanalyse über den Faktor Leistung Lesegeschwindigkeit am Anfang und Ende der ersten sowie am Ende der zweiten und vierten Klasse

Skala Leistung-LESEGESCHWINDIGKEIT	Manifeste Variable	Faktor
		1
Leistung E1	Häufige Wörter	.98
	Seltene Wörter	.99
	Pseudo-Wörter	.98
Leistung E2	Häufige Wörter	.94
	Seltene Wörter	.97
	Pseudo-Wörter	.92
Leistung E4	Häufige Wörter	.93
	Seltene Wörter	.96
	Pseudo-Wörter	.90

Bildung der *Skala Leistung-Rechtschreibsicherheit* am Beginn der ersten Klasse (*Leistung A1*):

Wie auch bei den anderen beiden Leistungsdimensionen zum Lesen erfolgt zuerst eine Überprüfung der Eindimensionalität der manifesten Variablen Prozentrangwert richtig geschriebener bekannter, neuer und Pseudowörter mittels einer Faktorenanalyse (Hauptkomponentenanalyse) (siehe Tabelle 3). Wie sich zeigt, laden die drei Variablen in einem Faktor hoch und es kann somit von Eindimensionalität ausgegangen werden. Weiters wird aus diesen manifesten Variablen die *Skala Leistung-Rechtschreibsicherheit A1* für den Anfang der ersten Klasse gebildet.

Tabelle 3: Resultat der Faktorenanalyse über den Faktor Leistung Rechtschreibsicherheit am Anfang der ersten Klasse

Skala Leistung- RECHTSCHREIBSICHERHEIT	Manifeste Variable	Faktor
		1
Leistung A1	PR richtig geschriebener bekannter Wörter	.73
	PR richtig geschriebener neuer Wörter	.82
	PR richtig geschriebener Pseudo-Wörter	.82

Am Ende der ersten, zweiten und vierten Klasse besteht die jeweilige *Skala (Leistung E1, E2 und E4)* nur aus der manifesten Variable Prozentsatz richtig geschriebener Wörter.

5.2.2 Bildung der Skalen zur Lehrereinschätzung

Zur Bildung der *Skala Einschätzung-Lesesicherheit A1* am Beginn der ersten Klasse wird eine Faktorenanalyse (Hauptkomponentenanalyse) über folgende 6 Items berechnet: Das Kind...

- kann die Position eines Lautes im Wort erkennen
- kann kurze Wörter in Laute aufgliedern
- kann zwei Laute zusammenschleifen
- kann alle durchgenommenen Buchstaben lesen
- kann alle bisher im Unterricht durchgenommenen Wörter sicher lesen
- kann auch bisher im Unterricht noch nicht durchgenommene Wörter, die aus bekannten Buchstaben zusammengesetzt sind, erlesen.

Wie Tabelle 4 zeigt, wird die Eindimensionalität der Items bestätigt und somit aus ihrem Summenscore die *Skala Einschätzung-Lesesicherheit A1* gebildet.

Tabelle 4: Resultat der Faktorenanalyse über den Faktor Einschätzung der Lesesicherheit am Anfang der ersten Klasse

Skala Einschätzung- Lesesicherheit	Manifeste Variable	Faktor
		1
Einschätzung A1	Kann Position eines Lautes im Wort erkennen	.84
	Kann kurze Wörter in Laute aufgliedern	.88
	Kann 2 Laute zusammenschleifen	.82
	Kann alle durchgenommen Buchstaben lesen	.78
	Kann alle durchgenommen Wörter sicher lesen	.85
	Kann auch noch nicht durchgenommene Wörter lesen	.80

Die *Skalen Einschätzung-Lesesicherheit E1, E2 und E4* (Ende erster, zweiter und vierter Klasse) setzen sich jeweils nur aus einer manifesten Variable zusammen, da die Lehrer hier die Lesesicherheit für jedes Kind allgemein auf einer 5-stufigen Skala einschätzten.

Bildung der Skalen Einschätzung-Lesegewindigkeit:

Ebenso wie die *Skalen Einschätzung-Lesesicherheit E1, E2 und E4* besteht die *Skala Einschätzung-Lesegewindigkeit E1, E2 und E4* am Ende der ersten, zweiten und vierten Klasse aus nur einer manifesten Variablen. Zum Beginn der ersten Klasse war die Erhebung der Lesegewindigkeit noch nicht möglich, daher liegen dazu auch keine Daten vor.

Bildung der *Skala Einschätzung-Rechtschreibsicherheit A1* am Anfang der ersten Klasse:

Über die folgenden 3 Items wird eine Faktorenanalyse (Hauptkomponentenanalyse) gerechnet:

Das Kind...

- kann Wörter, die wir im Unterricht bereits geschrieben haben, bei nochmaliger Vorgabe richtig schreiben.
- kann Wörter, die im Unterricht nur gelesen wurden, richtig schreiben.
- kann auch kurze Wörter, die wir noch nicht gelesen haben, aber deren Buchstaben wir schon gelernt haben, richtig schreiben.

Die Eindimensionalität der Itemstruktur wird in Tabelle 5 bestätigt. Daher kann aus dem Summenscore über diese Items die *Skala Einschätzung-Rechtschreibsicherheit A1 gebildet werden*.

Tabelle 5: Resultat der Faktorenanalyse über den Faktor Einschätzung der Rechtschreibsicherheit am Anfang der ersten Klasse

Skala Einschätzung- RECHTSCHREIBSICHERHEIT	Manifeste Variable	Faktor
		1
Einschätzung A1	Kann geübte Wörter richtig schreiben	.91
	Kann gelesene Wörter richtig schreiben	.95
	Kann unbekannte kurze Wörter richtig schreiben	.94

Die *Skalen Einschätzung der Rechtschreibsicherheit E1, E2 und E4* bestehen aus je einer manifesten Variablen.

5.2.3 Bildung der Skalen zur Verhaltenseinschätzung

In der vorliegenden Untersuchung wird der Einfluss von oppositionellem, hyperaktivem und aggressivem Verhalten, welches unter dem Begriff externalisierende Verhaltensweisen zusammengefasst wird, untersucht.

Um die Eindimensionalität der Variablen Hyperaktivität, Opposition und Aggression zu überprüfen, wird eine Faktorenanalyse (Hauptkomponentenanalyse) gerechnet (siehe Tabelle 6). Zu allen vier Zeitpunkten (A1, E1, E2 und E4) kann die Eindimensionalität der manifesten Variablen bestätigt werden.

Zur Bildung der *Skalen externalisierendes Verhalten A1, E1, E2 und E4* wird dann jeweils der Summenscore aus den drei manifesten Variablen herangezogen.

Tabelle 6: Resultat der Faktorenanalyse über den Faktor externalisierendes Verhalten

Skala externalisierendes Verhalten	Manifeste Variable	Faktor
		1
Verhalten A1	Opposition Aggression Hyperaktivität	.89 .79 .82
Verhalten E1	Opposition Aggression Hyperaktivität	.92 .88 .88
Verhalten E2	Opposition Aggression Hyperaktivität	.95 .84 .88
Verhalten E4	Opposition Aggression Hyperaktivität	.93 .87 .88

5.3 Kriterien der Modellbeurteilung

Die Güte der Anpassung der theoretischen Modellstruktur wird durch folgende Parameter und ihre Cut-off-Werte bestimmt (siehe **Fehler! Verweisquelle konnte nicht gefunden werden.**) (Weiber & Mühlhaus, 2010, S.176):

Tabelle 7: Gütekriterien
Gütekriterien zur Beurteilung des Gesamtfits eines Modells
nach Weiber und Mühlhaus (2010, S.176)

Art des Kriteriums	Kriterium	Cut-off-Wert	Quelle
interferenzstatistisches Gütekriterium	RMSEA	$\leq .05 - .08$	Browne und Cudeck (1993)
deskriptiver, absoluter Fit-Index	Chi-Quadrat / df	≤ 3	Homburg und Giering (1996)
inkrementelles Fitmaß	CFI	$\geq .09$	Homburg und Baumgartner (1995)

Der Root-Mean-Square-Error of Approximation (RMSEA) als interferenzstatistisches Maß legt die Güte des Modells anhand folgender Wertebereiche fest (Browne & Cudeck, 1993):

- $RMSEA \leq .05$: gute Modellanpassung
- $RMSEA \leq .08$: akzeptable Modellanpassung
- $RMSEA \geq .10$: inakzeptable Modellanpassung

Das Verhältnis des Chi-Quadrat-Wertes zur Anzahl der Freiheitsgrade (Chi-Quadrat/df oder CMIN/df) sollte nach Homburg und Giering (1996) nicht größer als 3 sein, um von einer guten Modellanpassung sprechen zu können.

Der Schwellenwert für den Comparative Fit Index (CFI) liegt bei .09 (Homburg & Baumgartner, 1995). Ist der CFI eines Modells größer als der Schwellenwert, so kann man von einem guten Modell-Fit sprechen.

Anhand der beschriebenen Modellparameter werden die nun folgenden Strukturgleichungsmodelle beurteilt.

6 Ergebnisse

Dieses Kapitel beschäftigt sich mit der Darstellung und Interpretation der Ergebnisse sowie mit der Beantwortung der Fragestellungen.

6.1 Darstellung und Interpretation der Ergebnisse

In den nun folgenden Strukturgleichungsmodellen wurden anstatt latenter Dimensionen, welche üblicherweise durch Ellipsen veranschaulicht werden, Skalen aus einzelnen manifesten Variablen bzw. aus Summenscores von mehreren manifesten Variablen gebildet, die durch Rechtecke dargestellt sind. Die Messfehler der Skalen werden durch Kreise veranschaulicht, sind aber größtenteils hier aus Gründen der Übersichtlichkeit weggelassen worden. Lediglich jene Messfehler, welche Korrelationen untereinander aufweisen (gebogene Doppelpfeile), werden in den Strukturmodellen veranschaulicht. Die Pfeile zwischen den Skalen stehen für die Effekte einer Skala auf eine andere Skala.

Zu jeder Dimension der Schülerleistung – Lesesicherheit, Lesegeschwindigkeit, Rechtschreibe-sicherheit – wird ein Strukturgleichungsmodell gezeigt. Wie bereits bei der Beschreibung des verwendeten Verfahrens erläutert wurde (siehe 5.1), werden in den folgenden drei Strukturgleichungsmodellen die Effekte zwischen den Leistungen, Einschätzungen und externalisierenden Verhaltensweisen dargestellt. Von Interesse ist die Akkuratheit der Lehrereinschätzungen, welche vor allem durch die Effektstärke der tatsächlichen Schülerleistung auf die Lehrereinschätzung zu allen vier Zeitpunkten gezeigt werden kann. Aufgrund der Erkenntnis von Schabmann und Schmidt (2009) wie auch von Bennett et. al. (1993), dass das Verhalten der Schüler einen Einfluss auf die Lehrereinschätzungen hat, soll durch die Effekte des externalisierenden Verhaltens der Schüler auf die Einschätzungen der Lehrer zu allen Zeitpunkten dieser Einfluss überprüft werden. Dieser spräche für eine Verringerung der Akkuratheit. Ebenso werden Effekte des externalisierenden Verhaltens auf die Schülerleistungen dargestellt, die sich schon bei Schabmann und Schmidt (2009) zeigten.

Zur Erklärung der Ergebnisdarstellung muss vorweggenommen werden, dass sämtliche negativen Effekte in den folgenden Strukturgleichungsmodellen Ausdruck dafür sind, dass die betreffenden Skalen unterschiedlich gepolt sind.

6.1.1 Pfadmodell Lesesicherheit

6.1.1.1 Parameterschätzung für das Pfadmodell Lesesicherheit

Die Beziehungen zwischen den Skalen Leistung, Einschätzung und externalisierendes Verhalten am Anfang der ersten sowie am Ende der ersten, zweiten und vierten Klasse werden im vorliegenden Strukturgleichungsmodell veranschaulicht (siehe Abbildung 2). Sie werden als Pfadkoeffizienten mittels Maximum-Likelihood-Methode geschätzt und standardisiert wiedergegeben. Diese Effektstärken zwischen den Skalen werden inklusive ihrer Signifikanzniveaus in Tabelle 8 dargestellt.

Tabelle 8: Effekte zwischen den Skalen des Pfadmodells Lesesicherheit
 standardisierte Lösung, *** signifikant ($p \leq 0,001$), **signifikant ($p \leq .01$), *signifikant ($p \leq .05$), ns. (nicht signifikant)

Pfad von	Pfad zu	Effekt e	Signifikanz (p)
Leistung A1	Leistung E1	.47	***
Leistung E1	Leistung E2	.46	***
Leistung E2	Leistung E4	.52	***
Einschätzung A1	Einschätzung E1	-.36	***
Einschätzung E1	Einschätzung E2	.42	***
Einschätzung E2	Einschätzung E4	.69	***
externalisierendes Verhalten A1	externalisierendes Verhalten E1	.78	***
externalisierendes Verhalten E1	externalisierendes Verhalten E2	.59	***
externalisierendes Verhalten E2	externalisierendes Verhalten E4	.56	***
Leistung A1	Einschätzung A1	.36	***
Leistung E1	Einschätzung E1	-.29	***
Leistung E2	Einschätzung E2	-.16	.009**
Leistung E4	Einschätzung E4	-.13	.008**
externalisierendes Verhalten A1	Leistung A1	-.21	.004**
externalisierendes Verhalten E1	Leistung E1	.07	.334 ns
externalisierendes Verhalten E2	Leistung E2	-.09	.194 ns.
externalisierendes Verhalten E4	Leistung E4	-.17	.005**
externalisierendes Verhalten A1	Einschätzung A1	-.33	***
externalisierendes Verhalten E1	Einschätzung E1	.13	.055 ns.
externalisierendes Verhalten E2	Einschätzung E2	.31	***
externalisierendes Verhalten E4	Einschätzung E4	.08	.117 ns.

In Tabelle 9 werden die Korrelationen der Messfehler untereinander und ihre Signifikanzniveaus dargestellt.

Tabelle 9: Messfehlerkorrelationen des Pfadmodells Lesesicherheit

*** signifikant ($p \leq 0,001$), **signifikant ($p \leq .01$), *signifikant ($p \leq .05$), ns. (nicht signifikant)

Messfehlerkorrelationen zwischen den Skalen		r	Signifikanz (p)
externalisierendes Verhalten A1	externalisierendes Verhalten E2	.37	.002**
externalisierendes Verhalten A1	externalisierendes Verhalten E4	.32	***
externalisierendes Verhalten E1	externalisierendes Verhalten E4	.11	.156 ns.

Anschließend folgt nun die Darstellung des Strukturgleichungsmodells zum Bereich Lesesicherheit über vier Jahre Grundschule hinweg, welches im Folgenden näher beschrieben wird.

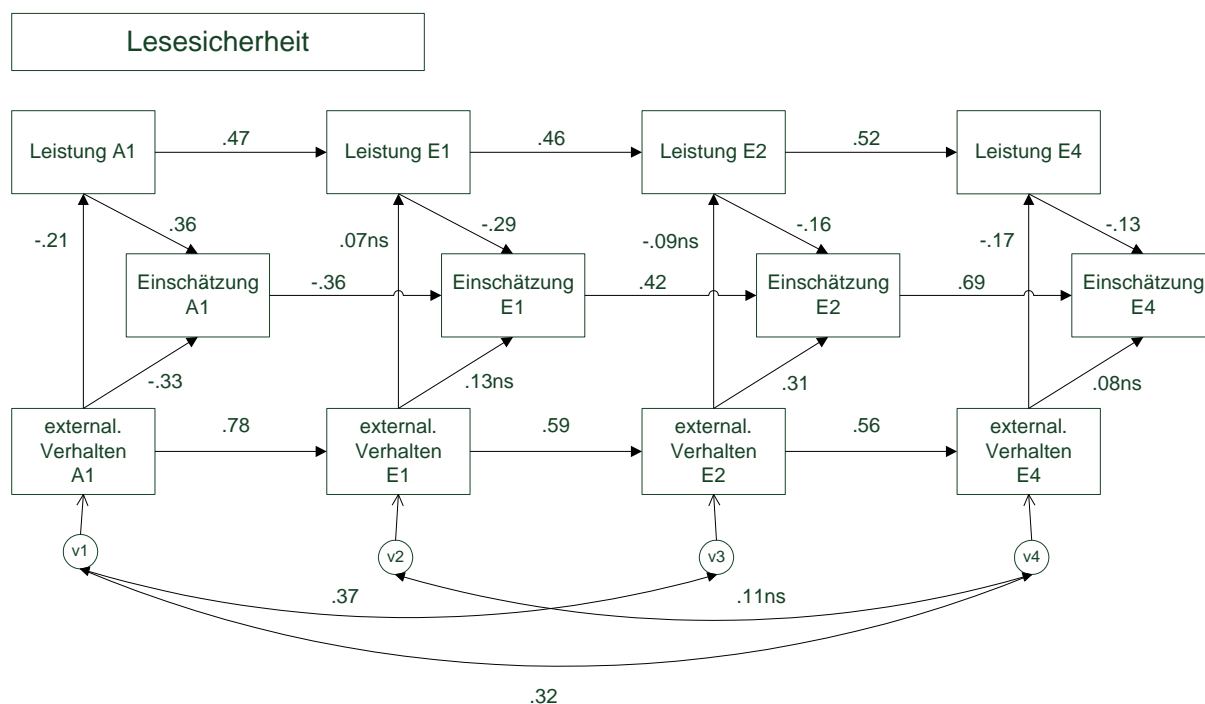


Abbildung 2: Strukturgleichungsmodell Lesesicherheit

Strukturgleichungsmodell zur Akkuratheit der Lehrereinschätzungen im Bereich der Lesesicherheit und zum Einfluss der wahrgenommenen externalisierenden Verhaltensweisen auf die Lehrereinschätzungen und die Schülerleistungen über vier Zeitpunkte während der Grundschulzeit. standardisierte Lösung; ns = nicht signifikante Koeffizienten ($p \geq 0.05$); Rechtecke stellen Skalen dar, Pfeile stellen Effekte zwischen den Skalen dar, Kreise stellen Messfehler dar; gebogene Doppelpfeile stellen unerklärte Korrelationen dar; Anmerkungen: A1 = Anfang erste Klasse, E1 = Ende erste Klasse, E2 = Ende zweite Klasse, E4 = Ende vierte Klasse

6.1.1.2 Ergebnisse des Pfadmodells Lesesicherheit

Das Pfadmodell Lesesicherheit (siehe Abbildung 2) beschreibt die Entwicklung der Akkuratheit der Lehrereinschätzungen bezüglich der Lesesicherheit anhand der Skalen *Leistung*, *externalisierendes Verhalten* und *Einschätzung* über die vier Zeitpunkte Anfang der ersten Klasse, Ende der ersten Klasse, Ende der zweiten Klasse und Ende der vierten Klasse Grundschule hinweg. In diesem Strukturgleichungsmodell wird ein direkter Einfluss der Leistung A1 (Anfang der ersten Klasse) auf die Lehrereinschätzung A1 vermutet. Ebenso wird angenommen, dass die Verhaltenseinschätzung durch die Lehrer (A1) einen Einfluss auf die Lehrereinschätzung A1 hat. Das bedeutet, dass sich die Lehrereinschätzung nicht nur aus der tatsächlichen Schülerleistung erklären lässt, sondern auch die Wahrnehmung des Verhaltens der Schüler auf die Lehrereinschätzung wirkt und somit die Akkuratheit beeinträchtigt wird. Außerdem wird davon ausgegangen, dass die Verhaltenseinschätzung A1 auf die Leistung A1 bezüglich Lesesicherheit Auswirkungen hat.

Von denselben Beziehungen ist zu den anderen drei Zeitpunkten am Ende der ersten, zweiten und vierten Klasse auszugehen.

Weiters wird die Entwicklung der Stabilität der Lehrereinschätzungen, der Verhaltenseinschätzungen und der Schülerleistungen über die vier Zeitpunkte A1, E1, E2 und E4 hinweg beschrieben.

Die Stabilität der Leistungen im Lesen vom ersten bis zum letzten Testzeitpunkt erweist sich mit signifikanten Werten zwischen .46 und .52 als mäßig gut, wie Tabelle 8 zeigt. Daraus kann geschlossen werden, dass ein Kind mit guten Leistungen zu Beginn der ersten Klasse auch gute Leistungen zu späteren Zeitpunkten erbringen wird.

Ein ähnliches Bild ergibt sich rund um die Stabilität der Lehrereinschätzungen, welche sich signifikant zwischen -.36 und .69 bewegen. Kinder, die zu Beginn der Schulzeit von den Lehrern als leistungsstark eingeschätzt werden, werden auch zu den späteren Zeitpunkten positiv eingeschätzt. Am stabilsten und daher den größten Effekt (.69) zeigen die Einschätzungen am Ende der zweiten Klasse auf jene am Ende der vierten Klasse.

Betrachtet man die Entwicklung der Verhaltenseinschätzungen über den Zeitraum von vier Jahren, so stellt man eine hohe Stabilität (.78) zwischen den ersten beiden Zeitpunkten fest, die dann kontinuierlich leicht abnimmt bis hin zu einem mittelmäßigen Ergebnis (.56). Diese vorliegenden Effekte bedeuten, dass eine positive Verhaltenseinschätzung am Beginn der ersten Klasse speziell auch am Ende der ersten Klasse aufrechterhalten wird, sowie auch bis zur vierten Klasse eher positiv bleibt.

Die Korrelationen zwischen den Messfehlern der Verhaltensskalen (siehe Tabelle 9) untereinander über mehrere Zeitpunkte hinweg zeigen sich folgendermaßen: Die Korrelation (Signifikanzniveau von 1%) zwischen den Messfehlern am Anfang der ersten Klasse und am Ende der zweiten Klasse beträgt .37, jene zwischen Anfang erster Klasse und Ende vierter Klasse beträgt .32. Eine nicht signifikante Korrelation zwischen dem Ende der ersten und dem Ende der vierten Klasse zeigt sich mit .11. Die Messfehlerkorrelationen stehen für durch das Modell nicht erklärte Varianzanteile des externalisierenden Verhaltens.

Der direkte Einfluss der tatsächlichen Leseleistungen der Schüler auf die Lehrereinschätzungen zeigt sich am höchsten zu Beginn der ersten Klasse mit einem moderaten sowie signifikanten Effekt von .36. Bereits ab dem Ende der ersten Klasse kann man nur mehr von einem geringen, signifikanten Effekt sprechen (-.29), am Ende der zweiten (-.16) und vierten (-.13) Klasse nur mehr von einem sehr geringen, auf einem Niveau von 1% signifikanten Effekt. Die Werte bedeuten, dass Lehrer ihre Schüler zu Beginn der ersten Klasse mäßig akkurat einschätzen können, also die Einschätzungen mäßig durch die Leistungen erklärt werden. Im Laufe der Grundschulzeit nimmt diese Akkuratheit noch mehr ab.

Betrachtet man die Effekte vom Verhalten auf die Lehrereinschätzungen, so ergibt sich zu Beginn der ersten Klasse ein mäßiger, aber signifikanter Effekt (-.33) sowie auch am Ende der zweiten Klasse (.31). Am Ende der ersten (.13) und vierten (.08) Klasse Grundschule hat das Verhalten keinen signifikanten Effekt auf die Lehrereinschätzungen. Am Anfang der ersten und Ende der zweiten Klasse scheinen also aggressive, oppositionelle oder hyperaktive Kinder schlechtere Leistungseinschätzungen von den Lehrern zu erhalten.

Der Einfluss des externalisierenden Verhaltens auf die Leseleistungen der Schüler ist gering und nur am Anfang der ersten Klasse mit $-.21$ und am Ende der vierten Klasse mit $-.17$ auf einem Alphaniveau von 1% signifikant. Trotzdem diese Effekte nur gering sind, bedeutet dies, dass ein unauffälliges Verhalten bezüglich externalisierender Verhaltensweisen wie Aggression, Hyperaktivität und Opposition eher mit guten Leistungen einhergeht. Am Ende der ersten Klasse ($.07$) und am Ende der zweiten Klasse ($-.09$) ist kein signifikanter Effekt vom Verhalten auf die Lesesicherheit festzustellen. Insgesamt scheint das Verhalten wenig bis gar keinen Einfluss auf die Schülerleistungen im Bereich der Lesesicherheit zu haben.

6.1.1.3 Modellbeurteilung für das Pfadmodell Lesesicherheit

Anhand der Modellgütewerte CMIN/DF mit einem Wert von 2.494, RMSEA (.092) und CFI (.941) kann man insgesamt von einem akzeptablen Modell-Fit ausgehen (siehe Tabelle 10). Das theoretische Modell zeigt somit eine akzeptable Anpassung an die empirischen Daten.

Tabelle 10: Werte der Modellgüte für das Pfadmodell Lesesicherheit

CMIN/DF	2.494
CFI	.941
RMSEA	.092

Zusammenfassend kann festgestellt werden, dass das Pfadmodell Lesesicherheit über eine akzeptable Modellanpassung verfügt. Das Verhalten zeigt lediglich zu zwei Zeitpunkten (A1 und E1) einen geringen Einfluss auf die Schülerleistungen sowie zu den beiden Zeitpunkten A1 und E2 einen moderaten Effekt auf die Lehrereinschätzungen bezüglich der Lesesicherheit. Die Leistung erklärt lediglich am Anfang der ersten Klasse die Lehrereinschätzungen mittelmäßig gut, im weiteren Verlauf der Grundschule kommt es zur Abnahme bis hin zu einem sehr geringen Erklärungswert. Somit kann von einer eher geringen Akkuratheit der Lehrereinschätzungen im Bereich der Lesesicherheit ausgegangen werden und zeitweise ein Einfluss des Verhaltens auf die Leistungseinschätzungen der Lehrer konstatiert werden.

6.1.2 Pfadmodell Lesegeschwindigkeit

6.1.2.1 Parameterschätzung für das Pfadmodell Lesegeschwindigkeit

Die Beziehungen zwischen den Skalen Leistung Lesegeschwindigkeit, Einschätzung und externalisierendes Verhalten am Ende der ersten, zweiten und vierten Klasse werden im vorliegenden Strukturgleichungsmodell veranschaulicht (siehe Abbildung 3). Mittels Maximum-Likelihood-Methode werden die Pfadkoeffizienten geschätzt und standardisiert wiedergegeben. In Tabelle 11 werden die Effektstärken und ihre Signifikanzen dargestellt.

Tabelle 11: Effekte zwischen den Skalen des Pfadmodells Lesegeschwindigkeit
standardisierte Lösung, *** signifikant ($p \leq 0,001$), **signifikant ($p \leq .01$),
*signifikant ($p \leq .05$), ns. (nicht signifikant)

Pfad von	Pfad zu	Effekte	Signifikanz (p)
Leistung E1	Leistung E2	.78	***
Leistung E2	Leistung E4	.81	***
Einschätzung E1	Einschätzung E2	.36	***
Einschätzung E2	Einschätzung E4	.55	***
externalisierendes Verhalten E1	externalisierendes Verhalten E2	.78	***
externalisierendes Verhalten E2	externalisierendes Verhalten E4	.45	***
Leistung E1	Einschätzung E1	-.34	***
Leistung E2	Einschätzung E2	-.33	***
Leistung E4	Einschätzung E4	-.20	***
externalisierendes Verhalten E1	Leistung E1	-.12	.116 ns.
externalisierendes Verhalten E2	Leistung E2	-.12	.006**
externalisierendes Verhalten E4	Leistung E4	-.08	.051 ns.
externalisierendes Verhalten E1	Einschätzung E1	.18	.008**
externalisierendes Verhalten E2	Einschätzung E2	.25	***
externalisierendes Verhalten E4	Einschätzung E4	.15	.006**

In Tabelle 12 werden die Korrelationen der Messfehler untereinander und ihre Signifikanzniveaus dargestellt.

Tabelle 12: Messfehlerkorrelationen des Pfadmodells Lesegeschwindigkeit

*** signifikant ($p \leq 0,001$), **signifikant ($p \leq .01$), *signifikant ($p \leq .05$), ns. (nicht signifikant)

Messfehlerkorrelationen zwischen den Skalen		r	Signifikanz (p)
externalisierendes Verhalten E1	externalisierendes Verhalten E4	.42	***

Weiters folgt nun die Darstellung des Strukturgleichungsmodells zum Bereich Lesegeschwindigkeit vom Ende der ersten bis zum Ende der vierten Klasse Grundschule, welches auch näher erläutert wird.

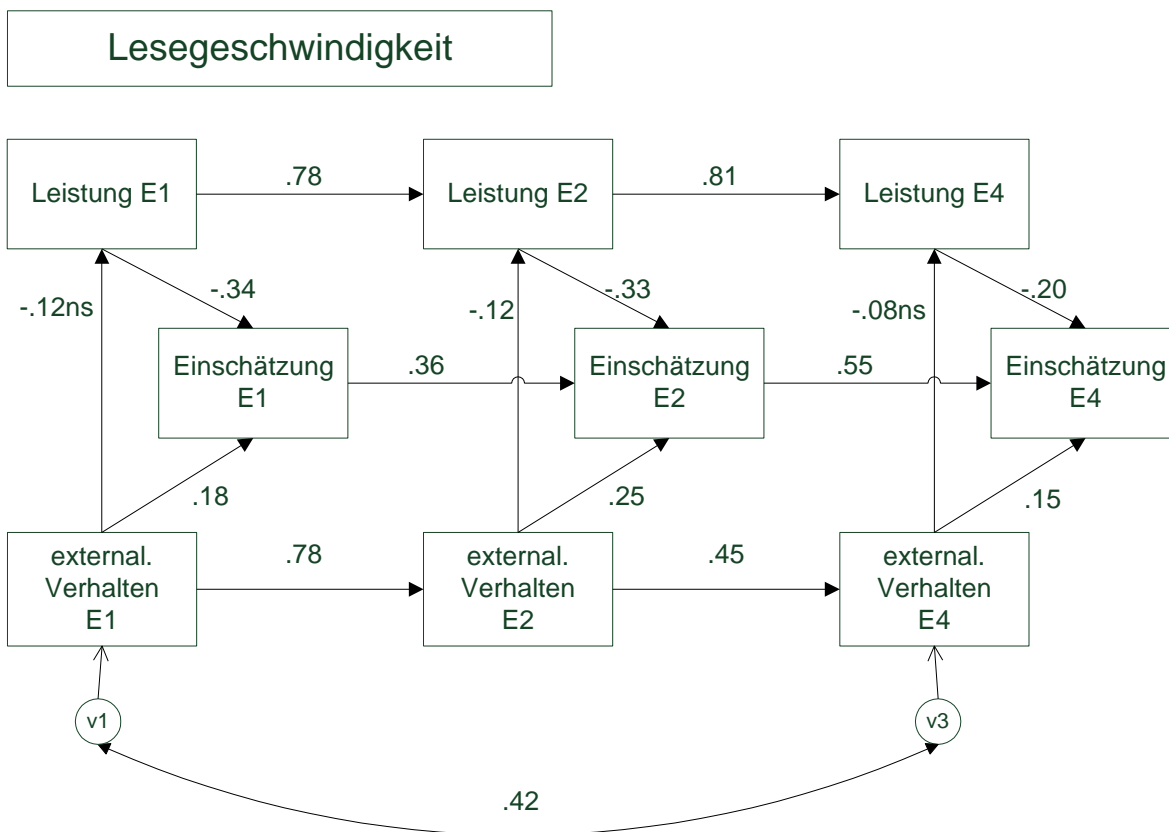


Abbildung 3: Strukturgleichungsmodell Lesegeschwindigkeit

Strukturgleichungsmodell zur Akkuratheit der Lehrereinschätzungen im Bereich der Lesegeschwindigkeit und zum Einfluss der wahrgenommenen externalisierenden Verhaltensweisen auf die Lehrereinschätzungen und die Schülerleistungen über vier Zeitpunkte während der Grundschulzeit. standardisierte Lösung; ns = nicht signifikante Koeffizienten ($p \geq 0.05$); Rechtecke stellen Skalen dar, Pfeile stellen Effekte der Skalen untereinander dar, Kreise stellen Messfehler dar, gebogene Doppelpfeile stellen unerklärte Korrelationen dar; Anmerkungen: A1 = Anfang erste Klasse, E1 = Ende erste Klasse, E2 = Ende zweite Klasse, E4 = Ende vierte Klasse

6.1.2.2 Ergebnisse des Pfadmodells Lesegeschwindigkeit

Im Pfadmodell Lesegeschwindigkeit (siehe Abbildung 3) wird die Entwicklung der Akkuratheit der Lehrereinschätzungen bezüglich der Lesegeschwindigkeit anhand der Skalen *Leistung*, *externalisierendes Verhalten* und *Einschätzung* über die drei Zeitpunkte Ende der ersten Klasse, Ende der zweiten Klasse und Ende der vierten Klasse Grundschule beschrieben. Es gibt deshalb nur drei Erhebungszeitpunkte, weil die Lesegeschwindigkeit am Beginn der ersten Klasse noch kein gut messbares Leistungskriterium wäre. Ein direkter Einfluss der Leistung E1 (Ende der ersten Klasse) auf die Lehrereinschätzung E1 wird im Strukturgleichungsmodell angenommen. Weiters geht man davon aus, dass die Verhaltenseinschätzung durch den Lehrer E1 einen Einfluss auf die Leistungseinschätzung E1 hat, was wiederum bedeutet, dass nicht nur die tatsächlichen Leistungen sondern auch externalisierende Verhaltensweisen der Schüler zur Erklärung der Lehrereinschätzungen bezüglich der Lesegeschwindigkeitsleistungen beitragen. Dies würde eine geringere Akkuratheit der Lehrereinschätzungen bedeuten. Außerdem wird angenommen, dass die Verhaltenseinschätzung E1 auf die Leistung E1 bezüglich Lesesicherheit Auswirkungen hat.

Von denselben Beziehungen ist zu den anderen beiden Zeitpunkten am Ende der zweiten und vierten Klasse auszugehen.

Weiters wird die Entwicklung der Stabilität der Lehrereinschätzungen, der Verhaltenseinschätzungen und der Schülerleistungen über die drei Zeitpunkte E1, E2 und E4 hinweg beschrieben. Die Leistungen bezüglich der Lesesicherheit erweisen sich mit signifikanten Werten von .78 und .81 als sehr stabil, wie Tabelle 11 zeigt. Daraus kann geschlossen werden, dass ein Kind mit guten Leistungen am Ende der ersten Klasse auch gute Leistungen am Ende der zweiten Klasse und im Weiteren gute am Ende der vierten Klasse erbringen wird.

Nicht ganz so stabil erweisen sich die Lehrereinschätzungen mit signifikanten Werten von .36 vom Ende der ersten zum Ende der zweiten Klasse hin und mit .55 vom Ende der zweiten zum Ende der vierten Klasse. Dies bedeutet, dass vor allem Lehrereinschätzungen am Ende der zweiten Klasse bis zum Ende der vierten Klasse beibehalten werden.

Betrachtet man die Entwicklung der Verhaltenseinschätzungen über den Zeitraum von drei Jahren, so stellt man eine hohe Stabilität (.78) zwischen dem Ende der ersten und zweiten Klasse fest, die dann kontinuierlich leicht abnimmt bis hin zu einem mittelmäßigen Ergebnis (.45) zwischen dem Ende der zweiten und vierten Klasse. Diese vorliegenden Effekte bedeuten, dass eine positive Verhaltenseinschätzung am Ende der ersten Klasse speziell auch am Ende der zweiten Klasse erhalten bleibt sowie auch von der zweiten bis zur vierten Klasse eher positiv bleibt.

Die Korrelation von .42 zwischen den Messfehlern der Verhaltensskalen (siehe Tabelle 12) am Ende der ersten und vierten Klasse bedeutet, dass unerklärte Varianzanteile des externalisierenden Verhaltens vorhanden sind.

Der direkte Einfluss der tatsächlichen Schülerleistungen punkto Lesegeschwindigkeit auf die Lehrereinschätzungen bewegt sich mit signifikanten Werten zwischen -.34 und -.20. Am Ende der ersten (-.34) und zweiten Klasse (-.33) kann man noch von einem mittleren, am Ende der vierten Klasse (-.20) nur mehr von einem geringen Effekt sprechen. Lehrer schätzen demnach ihre Schüler am Ende der ersten und zweiten Klasse mäßig akkurat ein, am Ende der Grundschule weisen diese Einschätzungen bezüglich der Lesegeschwindigkeit jedoch nur mehr geringe Akkuratheit auf.

Der Effekt des Verhaltens auf die Lehrereinschätzungen ist zu allen drei Zeitpunkten gering, aber signifikant. Am Ende der ersten und vierten Klasse zeigen sich geringe auf einem Signifikanzniveau von 1% signifikante Effekte von .18 und .15. Am Ende der zweiten Klasse ist der Effekt von .25 sogar auf einem Alphaniveau von 0,1% signifikant. Aggressive, oppositionelle oder hyperaktive Kinder dürften demnach öfters schlechtere Leistungseinschätzungen der Lehrer erhalten als Kinder mit unauffälligem Verhalten.

Die Beeinflussung der Lesegeschwindigkeitsleistungen durch externalisierende Verhaltensweisen ist nur am Ende der zweiten Klasse gegeben und selbst da sehr gering mit einem signifikanten Effekt von -.12 (Signifikanzniveau von 1%).

Am Ende der ersten Klasse (-.12) und am Ende der vierten Klasse (-.08) gibt es keine signifikanten Werte. Ein Kind mit unauffälligem externalisierendem Verhalten am Ende der zweiten Klasse wird eher gute Leistungen zeigen. Das externalisierende Verhalten hat also wenig bis gar keinen Einfluss auf die tatsächliche Lesegeschwindigkeit.

6.1.2.3 Modellbeurteilung für das Pfadmodell Lesegeschwindigkeit

Der absolute Fit-Index CMIN/DF mit einem Wert von 2.375 sowie das inkrementelle Fitmaß CFI mit .972 stehen für eine gute Modellanpassung. Der Wert RMSEA mit .088 befindet sich im akzeptablen Bereich (siehe Tabelle 13). Insgesamt kann man von einem guten Modell sprechen.

Tabelle 13: Werte der Modellgüte für das Pfadmodell Lesegeschwindigkeit

CMIN/DF	2.375
CFI	.972
RMSEA	.088

Zusammenfassend betrachtet zeigt das Pfadmodell Lesegeschwindigkeit eine gute Modellanpassung. Das Verhalten zeigt lediglich am Ende der zweiten Klasse einen sehr geringen, signifikanten Einfluss auf die Schülerleistungen. Auch auf die Lehrereinschätzungen übt die Wahrnehmung des Verhaltens durch die Lehrer nur einen geringen, aber signifikanten Einfluss aus. Der größte Wirkfaktor ist die tatsächliche Leistung, aus der sich die Lehrereinschätzung bildet. Jedoch zeigen sich hier zu allen Zeitpunkten auch nur mittelmäßige bis geringe Effekte, was für eine mäßige Akkuratheit spricht.

Es kann also von einer mäßigen Akkuratheit der Lehrerurteile gesprochen werden, wenn es sich um Einschätzungen der Lesegeschwindigkeit handelt. Auch das externalisierende Verhalten der Schüler hat zumindest zeitweise einen geringen Einfluss auf die Leistungen und die Leistungseinschätzungen.

6.1.3 Pfadmodell Rechtschreibsicherheit

6.1.3.1 Parameterschätzung für das Pfadmodell Rechtschreibsicherheit

Im vorliegenden Strukturgleichungsmodell werden die Beziehungen zwischen den Skalen *Leistung Rechtschreibsicherheit*, *Einschätzung* und *externalisierendes Verhalten* am Anfang der ersten sowie am Ende der ersten, zweiten und vierten Klasse veranschaulicht. Die Schätzung der Pfadkoeffizienten erfolgt mittels Maximum-Likelihood-Methode. Sie sind im Strukturmodell standardisiert wiedergegeben (siehe Abbildung 4).

In Tabelle 14 werden die Effektstärken und Signifikanzniveaus dargestellt.

Tabelle 14: Effekte zwischen den Skalen des Pfadmodells Rechtschreibsicherheit
standardisierte Lösung, *** signifikant ($p \leq 0,001$), **signifikant ($p \leq .01$),
*signifikant ($p \leq .05$), ns. (nicht signifikant)

Pfad von	Pfad zu	Effekte	Signifikanz (p)
Leistung A1	Leistung E1	.13	.081 ns.
Leistung E1	Leistung E2	.47	***
Leistung E2	Leistung E4	.67	***
Einschätzung A1	Einschätzung E1	-.32	***
Einschätzung E1	Einschätzung E2	.39	***
Einschätzung E2	Einschätzung E4	.44	***
externalisierendes Verhalten A1	externalisierendes Verhalten E1	.78	***
externalisierendes Verhalten E1	externalisierendes Verhalten E2	.59	***
externalisierendes Verhalten E2	externalisierendes Verhalten E4	.56	***
Leistung A1	Einschätzung A1	.23	***
Leistung E1	Einschätzung E1	-.32	***
Leistung E2	Einschätzung E2	-.33	***
Leistung E4	Einschätzung E4	-.40	***
externalisierendes Verhalten A1	Leistung A1	-.18	.013*
externalisierendes Verhalten E1	Leistung E1	-.11	.147 ns.
externalisierendes Verhalten E2	Leistung E2	-.23	***
externalisierendes Verhalten E4	Leistung E4	-.19	***
externalisierendes Verhalten A1	Einschätzung A1	-.36	***
externalisierendes Verhalten E1	Einschätzung E1	.20	.002**
externalisierendes Verhalten E2	Einschätzung E2	.14	.028*
externalisierendes Verhalten E4	Einschätzung E4	.10	.073 ns.

Die Korrelationen der Messfehler und ihre Signifikanzen werden in Tabelle 15 dargestellt.

Tabelle 15: Messfehlerkorrelationen des Pfadmodells Rechtschreibsicherheit
 *** signifikant ($p \leq 0,001$), **signifikant ($p \leq .01$), *signifikant ($p \leq .05$), ns. (nicht signifikant)

Messfehlerkorrelationen zwischen den Skalen		r	Signifikanz (p)
externalisierendes Verhalten A1	externalisierendes Verhalten E2	.37	.002**
externalisierendes Verhalten A1	externalisierendes Verhalten E4	.32	***
externalisierendes Verhalten E1	externalisierendes Verhalten E4	.11	.156 ns.

Das bedeutet, dass Lehrereinschätzungen sich nicht nur aus den tatsächlichen Schülerleistungen zusammensetzen, sondern auch andere Parameter wie die Wahrnehmung des Schülerverhaltens eine Rolle spielen, was eine geringere Akkuratheit bedeuten würde. Weiters wird angenommen, dass die Verhaltenseinschätzung A1 auch einen Teil der Varianz der Rechtschreibleistungen A1 erklärt.

Von eben diesen Beziehungen ist zu den anderen drei Zeitpunkten - am Ende der ersten, zweiten und vierten Klasse - auszugehen.

Auch die Entwicklung der Stabilität der Lehrereinschätzungen, der Verhaltenseinschätzungen und der Schülerleistungen über die vier Zeitpunkte A1, E1, E2 und E4 wird im Strukturmodell beschrieben.

Die Stabilität der Leistungen im Rechtschreiben ist im ersten Schuljahr zwischen Anfang und Ende erster Klasse mit einem nicht signifikanten Wert von .13 nicht gegeben. Hier kann aus den Leistungen zu Beginn der ersten Klasse nicht auf die Leistungen am Ende der ersten Klasse geschlossen werden. Mittlere Stabilitäten finden sich dann mit signifikanten Werten von .47 und .67 zwischen den anderen drei Zeitpunkten. Kinder mit guten Leistungen am Ende der ersten Klasse werden auch eher gute Leistungen am Ende der zweiten Klasse zeigen. Gleiches gilt für gute Leistungen am Ende der zweiten Klasse, welche relativ stabil bis zum Ende der vierten Klasse sind.

Was die Stabilität der Lehrereinschätzungen betrifft, so sprechen signifikante Werte zwischen -.32 und .44 für eine mäßige bis geringe Ausprägung. Das bedeutet, dass Kinder, die zu einem Zeitpunkt als sichere Schreiber eingeschätzt wurden, werden auch zu anderen Zeitpunkten so gesehen.

Betrachtet man die Entwicklung der Verhaltenseinschätzungen über den Zeitraum von vier Jahren, so stellt man eine hohe Stabilität (.78) zwischen Anfang und Ende der ersten Klasse fest, die dann abnimmt bis hin zu einer mittleren Ausprägung (.56). Vor allem ein Kind, das am Anfang der ersten Klasse als verhaltensauffällig gilt, wird auch am Ende der Klasse derart eingeschätzt. Ähnliches trifft auch auf die Entwicklung der Einschätzungen am Ende der ersten bis zum Ende der zweiten und hin zum Ende der vierten Klasse zu.

Die Messfehlerkorrelationen der Verhaltensskalen (siehe Tabelle 15) zeigen zwischen Anfang der ersten Klasse und Ende der zweiten Klasse einen Wert von .37 (Signifikanzniveau von 1%), jene zwischen Anfang erster Klasse und Ende vierter Klasse beträgt .32 und ist signifikant. Eine nicht signifikante Korrelation zwischen dem Ende der ersten und dem Ende der vierten Klasse gibt es mit einem Wert von .11. Messfehlerkorrelationen stellen nicht erklärte Varianzanteile des externalisierenden Verhaltens im vorliegenden Modell dar.

Betrachtet man den direkten Einfluss der tatsächlichen Rechtschreibleistungen auf die Lehrerurteile, so kommt es zu einem leichten Anstieg der Effekte von .23 bis -.40 innerhalb von ca. 4 Jahren. Die Effekte sind durchwegs signifikant und leider nur gering bis mäßig stark. Das bedeutet, dass Lehrer ihre Schüler am Anfang der Grundschule sehr wenig akkurat einschätzen können, erst am Ende der Grundschule haben ihre Beurteilungen einen stärkeren Bezug zur tatsächlichen Leistungsfähigkeit ihrer Schüler.

Die Effekte des Verhaltens auf die Lehrereinschätzungen ergeben zu Beginn der ersten Klasse eine mittlere Stärke (-.36). Doch schon am Ende der ersten Klasse kann man nur noch von einem geringen, signifikanten Einfluss sprechen, der dann im Laufe der Grundschule bis hin zu einem nicht signifikanten, sehr geringen Effekt abnimmt. Am Ende der ersten Klasse zeigt sich demnach ein geringer Effekt auf die Einschätzungen von .20 (Signifikanzniveau von 1%), am Ende der zweiten Klasse ebenfalls (.14 auf einem Niveau von 5%).

Vor allem am Anfang der ersten Klasse scheinen also aggressive, oppositionelle oder hyperaktive Kinder schlechtere Leistungseinschätzungen bezüglich der Rechtschreibe-sicherheit von den Lehrern zu erhalten. Dieser Zusammenhang nimmt über die vier Jahre hinweg ab.

Der Einfluss der wahrgenommenen externalisierenden Verhaltensweisen auf die Schülerleistungen im Rechtschreiben ist am Anfang der ersten Klasse mit -.18 auf einem Alphaniveau von 5% signifikant, aber gering. Am Ende der zweiten (-.23) und vierten Klasse (-.19) gibt es ebenfalls geringe, signifikante Effekte vom Verhalten auf die Leistung. Nur am Ende der ersten Klasse zeigt das externalisierende Verhalten keinen Einfluss auf die Schülerleistungen. Die signifikanten Werte bedeuten, dass aggressives, oppositionelles und hyperaktives Verhalten zu den drei Zeitpunkten eher die Leistungsfähigkeit beeinträchtigen.

Insgesamt scheint das Verhalten aber nur einen geringen Einfluss auf die Schülerleistungen im Bereich der Rechtschreibe sicherheit zu haben.

6.1.3.3 Modellbeurteilung für das Pfadmodell Rechtschreibe sicherheit

Sowohl der Fit-Index CMIN/DF mit einem Wert von 2.749 als auch der CFI mit .935 und der RMSEA mit .099 stehen für eine akzeptable Modellanpassung. (siehe Tabelle 16).

Aufgrund der genannten Fit-Indices bzw. Gütekriterien kann insgesamt von einem eher guten Modell-Fit ausgegangen werden.

Tabelle 16: Werte der Modellgüte für das Pfadmodell Rechtschreibe sicherheit

CMIN/DF	2.749
CFI	.935
RMSEA	.099

Zusammenfassend gesprochen ergibt das Pfadmodell Rechtschreibe sicherheit eine gute Modellanpassung. Das Verhalten zeigt zu allen Zeitpunkten zumindest einen geringen Einfluss auf die Rechtschreibleistungen der Schüler, außer am Ende der zweiten Klasse. Auswirkungen des durch den Lehrer wahrgenommenen externalisierenden Schülerverhaltens auf seine Leistungseinschätzungen zeigen sich bei den Rechtschreibleistungen mäßig am Anfang der ersten Klasse. Danach scheint das Verhalten nur mehr einen sehr geringen bis gar keinen signifikanten Effekt mehr zu haben.

Die Kompetenz der Lehrer ihre Schüler entsprechend ihres Leistungsstandes im Rechtschreiben einzuschätzen, nimmt im Laufe der Grundschulzeit zu. Schüler am Ende der vierten Klasse werden von ihren Lehrer im Rechtschreiben mäßig gut eingeschätzt.

Der vorhandene Einfluss des Verhaltens und der geringe bis mäßige Einfluss der Leistungen auf die Leistungsbeurteilungen sprechen für eine geringe bis mäßige Akkuratheit der Lehrereinschätzungen im Bereich der Rechtschreibe sicherheit.

Nachdem die einzelnen Modelle ausführlich beschrieben wurden, kann nun zur Beantwortung der Forschungsfragen übergegangen werden.

6.2 Beantwortung der Fragestellungen

Die Fragestellungen wurden mit Hilfe von Strukturgleichungsmodellen untersucht. Diese dienen dazu mehrere Kausalhypothesen gleichzeitig zu überprüfen (Weiber und Mühlhaus, 2010).

Die verschiedenen Modelle veranschaulichen, welche Einflussfaktoren auf die Lehrereinschätzungen wirken und wie akkurat Lehrereinschätzungen bzw. –urteile sind.

6.2.1 Lesesicherheit

Wie akkurat sind die Lehrereinschätzungen bezüglich der Lesesicherheit über die 4 Jahre Grundschule hinweg?

Zur Beantwortung der Fragestellung wurden die direkten Effekte der Leistungen auf die Lehrereinschätzungen im Pfadmodell Lesesicherheit herangezogen. Die Akkuratheit der Lehrereinschätzungen bezüglich der Lesesicherheit ist zu Beginn der Grundschulzeit mittelmäßig (.36) und zeigt sich im weiteren Verlauf der Grundschule als gering (E1: -.29, E2: -.16 und E4: -.13). Es kommt zu einer Abnahme der Akkuratheit im Laufe der Grundschule, insgesamt kann man von einer eher geringen Akkuratheit der Lehrereinschätzungen sprechen.

Welchen Einfluss üben die durch den Lehrer wahrgenommenen externalisierenden Verhaltensweisen auf die Lehrereinschätzungen bezüglich der Lesesicherheit in der Grundschule aus?

Diese Fragestellung kann dahingehend beantwortet werden, dass das externalisierende Verhalten einen Beitrag zur Erklärung der Lehrereinschätzungen bezüglich der Schülerleistungen leistet. Allerdings gilt dies nur am Anfang der ersten (-.33) und am Ende der zweiten Klasse (.31). Zu den anderen beiden Zeitpunkten konnte dies nicht festgestellt werden.

Welchen Einfluss haben externalisierende Verhaltensweisen auf die Schülerleistungen im Bereich Lesesicherheit?

Die Fragestellung kann durch die Betrachtung der Effekte des durch den Lehrer eingeschätzten externalisierenden Verhaltens auf die tatsächliche Schülerleistung beantwortet werden.

Lediglich am Anfang der ersten (-.21) und am Ende der vierten Klasse (-.14) zeigen sich signifikante Effekte vom Verhalten auf die Leistung, und selbst diese sind nur gering.

Insgesamt kann man also von einem geringen bis keinem Einfluss des externalisierenden Verhaltens auf die Leseleistungen ausgehen.

6.2.2 Lesegeschwindigkeit

Wie akkurat sind die Lehrereinschätzungen bezüglich der Lesegeschwindigkeit über die 4 Jahre Grundschule hinweg?

Die Frage nach der Akkuratheit der Lehrereinschätzungen bezüglich der Lesegeschwindigkeit kann anhand der direkten Effekte der Leistungen auf die Einschätzungen dahingehend beantwortet werden, dass sich am Ende der ersten (-.34) und zweiten Klasse (-.33) eine mäßige Akkuratheit, am Ende der vierten Klasse (-.20) nur mehr eine geringe Akkuratheit zeigt.

Insgesamt kann also von einer mäßigen Akkuratheit der Lehrereinschätzungen bezüglich der Lesegeschwindigkeit ausgegangen werden, jedoch zeigt sich ein stetiges Abnehmen der Akkuratheit im Laufe der Grundschule.

Welchen Einfluss üben die durch den Lehrer wahrgenommenen externalisierenden Verhaltensweisen auf die Lehrereinschätzungen bezüglich der Lesegeschwindigkeit in der Grundschule aus?

Im Pfadmodell Lesegeschwindigkeit zeigen sich über alle drei Erhebungszeitpunkte hinweg geringe, signifikante Effekte (zwischen .18 und .25) von den wahrgenommenen externalisierenden Verhaltensweisen auf die Lehrereinschätzungen. Das externalisierende Verhalten trägt also gering zur Erklärung der Lehrereinschätzung bezüglich der Lesegeschwindigkeit bei.

Welchen Einfluss haben externalisierende Verhaltensweisen auf die Schülerleistungen im Bereich Lesegeschwindigkeit?

Um dies beantworten zu können, müssen die Effekte des externalisierenden Verhaltens auf die Leistungen im Bereich der Lesegeschwindigkeit betrachtet werden. Diese sind eigentlich nicht vorhanden, lediglich am Ende der zweiten Klasse (-.12) kann man von einem sehr geringen Einfluss des Verhaltens auf die Leistungen sprechen. Insgesamt ist von einem nicht vorhandenen bis minimalen Einfluss auszugehen.

6.2.3 Rechtschreibsicherheit

Wie akkurat sind die Lehrereinschätzungen bezüglich der Rechtschreibsicherheit über die 4 Jahre Grundschule hinweg?

Um die Akkuratheit der Lehrereinschätzungen beurteilen zu können, werden die direkten Effekte des Pfadmodells Rechtschreibsicherheit von der Leistung zur Einschätzung herangezogen. Es zeigt sich zu Beginn der ersten Klasse eine geringe Akkuratheit (.23), am Ende der ersten Klasse mit -.32, der zweiten mit -.33 und der vierten mit -.40 ein deutlicher Anstieg im Laufe der Grundschule hin zu einer moderaten Akkuratheit der Lehrereinschätzungen bezüglich der Rechtschreibsicherheit.

Welchen Einfluss üben die durch den Lehrer wahrgenommenen externalisierenden Verhaltensweisen auf die Lehrereinschätzungen bezüglich der Rechtschreibsicherheit in der Grundschule aus?

Zur Beantwortung der Fragestellung werden die Effekte des externalisierenden Verhaltens auf die Leistungseinschätzungen der Lehrer betrachtet. Zu Beginn der ersten Klasse Grundschule zeigt sich noch ein mittlerer Effekt von -.36, am Ende der ersten nur ein geringer von .20 und am Ende der zweiten ebenfalls nur ein geringer Effekt von .14. Am Ende der vierten Klasse kommt es nicht mehr zu einem signifikanten Wert. Der Einfluss des externalisierenden Verhaltens auf die Lehrereinschätzungen nimmt demnach im Laufe der Grundschuljahre ab.

Zu Beginn ist er noch mäßig stark, am Ende der Grundschule ist er gar nicht mehr vorhanden.

Welchen Einfluss haben externalisierende Verhaltensweisen auf die Schülerleistungen im Bereich Rechtschreibsicherheit?

Um diese Fragestellung beantworten zu können, werden wiederum die direkten Effekte des externalisierenden Verhaltens auf die Leistungen der Schüler betrachtet.

Am Beginn der ersten Klasse (-.18), am Ende der zweiten (-.23) und am Ende der vierten Klasse (-.19) zeigen sich geringe, signifikante Effekte des Verhaltens auf die Leistungen. Nur am Ende der ersten Klasse gibt es keinen signifikanten Effekt. Man kann daher insgesamt von einem geringen Einfluss des externalisierenden Verhaltens auf die Schülerleistung ausgehen.

Im nächsten Kapitel folgen die Zusammenfassung der Ergebnisse der vorliegenden Untersuchung sowie die kritische Betrachtung des wissenschaftlichen Vorgehens und Implikationen für weitere Forschung auf dem Gebiet der Akkuratheit von Lehrereinschätzungen.

7 Zusammenfassung und Diskussion

Trotzdem dem Lehrerurteil erhebliche Validität zukommt, gibt es auch Anteile der Varianz der Lehrerurteile, die nicht erklärt werden können (Egan & Archer, 1985). Ziel der vorliegenden Arbeit war es daher die Akkuratheit der Lehrereinschätzungen zu untersuchen und festzustellen, ob das durch den Lehrer wahrgenommene externalisierende Verhalten einen Beitrag zur Erklärung dieser Varianz leisten kann. Dazu wurde der Einfluss des externalisierenden Verhaltens auf die Lehrereinschätzungen und außerdem auch auf die tatsächlichen Schülerleistungen überprüft, welcher zu einer Beeinträchtigung der Akkuratheit führen würde.

Die Ergebnisse der Studie deuten darauf hin, dass die Lehrer die Leistungen ihrer Schüler mit geringer bis mäßiger Akkuratheit einschätzen. In den Bereichen Lesegeschwindigkeit und Lesesicherheit scheint die Akkuratheit speziell zu Beginn der ersten Klasse Grundschule mittelmäßig zu sein und im Laufe der Grundschule abzunehmen, bis hin zu einer geringen Akkuratheit. Beim Rechtschreiben zeigt sich der umgekehrte Effekt, die Akkuratheit der Lehrereinschätzungen steigt von geringen bis hin zu mäßigen Werten gegen Ende der Grundschule an. Wie bei Schmidt und Schabmann (2010) zeigte sich, dass die Einschätzungen im Rechtschreiben insgesamt betrachtet etwas akkurater waren als im Lesen und die Akkuratheit am Ende der Grundschule am höchsten war. Diese mittelmäßigen Ergebnisse deuten daraufhin, dass Lehrer spezifische Fähigkeiten wie beispielsweise Lesegeschwindigkeit schwer einschätzen können (vgl. Eckert et al., 2006; Feinberg & Shapiro, 2003).

Weiters zeigt sich in den drei Kompetenzbereichen, dass das externalisierende Verhalten zu fast allen Erhebungszeitpunkten mindestens gering zur Erklärung der Lehrereinschätzungen beiträgt, wie es unter anderem auch schon bei Brophy und Good (1974), Hecht und Greenfield (2002) sowie Tournaki und Podell (2005) gezeigt wurde. Lehrer scheinen ihre Schüler negativer zu beurteilen, wenn sie aggressives, oppositionelles oder hyperaktives Verhalten bei ihren Schülern wahrnehmen. Vor allem aber beim Rechtschreiben nimmt der Einfluss des externalisierenden Verhaltens über vier Jahre Grundschule hinweg ab und ist am Ende der vierten Klasse nicht mehr signifikant. Außerdem geht dies gleichzeitig mit einem Anstieg der Akkuratheit einher.

Vermutlich fällt es Lehrern leichter, diese Schülerfähigkeiten richtig einzuschätzen und sich vor allem in der vierten Klasse nicht von anderen Faktoren beeinflussen zu lassen.

Insgesamt kann man von einer eher geringen bis maximal mäßigen Akkuratheit der Lehrereinschätzungen sprechen, wie sich auch schon in anderen Studie zeigte (z.B. Deutsches PISA-Konsortium, 2001; Schabmann & Schmidt, 2009; Schmidt & Schabmann, 2010; Schrader & Helmke, 1990; Spinath, 2005).

Ebenso wie bei Bennett et. al. (1993) zeigt das durch den Lehrer wahrgenommene externalisierende Verhalten einen direkten Einfluss auf die Beurteilungen der Lehrer zu haben. Es trägt wie in der Studie von Schabmann und Schmidt (2009) zu den meisten Zeitpunkten zumindest gering zur Erklärung der Lehrereinschätzungen bezüglich Schreiben und Lesen bei.

Der Einfluss des wahrgenommenen externalisierenden Verhaltens auf die Leistungen der Schüler ist aber nur teilweise gegeben und sehr gering. Aber es zeigte sich auch schon bei Schabmann und Schmidt (2009) eine signifikante Wirkung von wahrgenommenem hyperaktivem Verhalten auf die Leseleistungen.

Insgesamt scheinen durch den Lehrer wahrgenommene externalisierende Verhaltensweisen die Beziehung zwischen Leistung und Lehrereinschätzung zu moderieren, allerdings nur in geringem Maß und nicht über die gesamte Grundschule hinweg.

Im Zusammenhang mit der vorliegenden Studie gilt es nun auch, eine kritische Betrachtung anzustreben. Ein methodischer Kritikpunkt der Studie liegt im korrelativen Ansatz der Strukturgleichungsmodelle. Der Nachteil von korrelativen Verfahren besteht darin, dass systematische Über- oder Unterschätzungen der Schülerleistungen nicht enttarnt werden können, da Korrelationen das absolute Niveau der Schülerleistungen unberücksichtigt lassen (Spinath, 2005). Auch Eckert et. al (2006) und Feinberg und Shapiro (2003) sind der Ansicht, dass Lehrereinschätzungen nicht ausreichend durch Korrelationskoeffizienten abgebildet werden. Außerdem zog man zur Erhebung der Lehrereinschätzungen in der aktuellen Studie indirekte Methoden, Rating-Skalen, heran (vgl. Begeny et. al., 2008).

Egan & Archer (1985, S.309) stellten zwar kaum Unterschiede in der Akkuratheit je nach verwendeter Methode (direkt oder indirekt) fest, sprachen sich aber trotzdem dafür aus, dass die der Beurteilung und dem Kriterium zugrundeliegenden Messskalen übereinstimmen sollten. Hoge & Coladarci (1989) fanden durchschnittlich höhere Korrelationen bei der direkten Einschätzung der Testleistungen durch die Lehrer und auch Feinberg und Shapiro (2003) sprechen sich für die CBM-Methode aus, da sie im Gegensatz zu Rating-Skalen und standardisierten Leistungstests einen direkten Bezug zu dem im Unterricht Gelernten haben.

Eine weitere Einschränkung stellt die Tatsache dar, dass sich die Studie nur mit sehr spezifischen Fähigkeiten befasst und sicher mehrere Teilkompetenzen, die bspw. für das Lesen von Bedeutung sind, wie Leseverständnis und Enkodieren, außer Acht lässt (vgl. Shinn, Good, Knuston, Tilly & Collins, 1992, zitiert nach Feinberg & Shapiro, 2003). Außerdem sind die Ergebnisse zeitlich als auch räumlich begrenzt zu interpretieren wie bei Hecht und Greenfield (2002), da die Erhebung der Daten schon ungefähr zehn Jahre zurückliegt und sich auf ein Bundesland (Niederösterreich) beschränkt.

Aus den Ergebnissen der Studie können einige Schlüsse für den zukünftigen Umgang mit Lehrerurteilen gezogen werden. Wünschenswert wäre es inakkurate Beurteilungen zu vermeiden und wie Kleber (1992) schon sagte evaluative Bewertungen anzustreben, was bedeutet, dass stets nur ein vorläufiges Urteil abgegeben werden soll, dass immer wieder neu überprüft wird und nicht aus Vorurteilen heraus gebildet werden sollte. Auch Brophy & Good, (1974) waren der Meinung, dass Lehrer sich nicht von stereotypen Erwartungen leiten lassen dürfen, die aus nur wenigen Informationen gebildet werden.

Weiters erscheint es sinnvoll, die Urteilskompetenz der Lehrer durch Bildungs- und Weiterbildungsangebote zu fördern (vgl. Krolak-Schwerdt & Rummer, 2005). Besonders die Stärkung der diagnostischen Kompetenz im Bereich Lesen wäre nach den Ergebnissen der vorliegenden Studie wünschenswert, doch leider zeigt die Lehrerausbildung dahingehend Lücken auf (vgl. Klicpera et. al., 2008). Eine Stütze für Lehrer in der Praxis könnten standardisierte Testwerte sein (vgl. Bennett et. al., 1993) sowie andere objektive Daten, die urteilsverzerrende Variablen ins Bewusstsein holen (vgl. Demaray & Elliott, 1998).

Für die künftige Forschung auf dem Gebiet der Lehrereinschätzungen wäre auch die genauere Untersuchung des Einflusses der lehrerbezogenen Merkmale wie Geschlecht und Persönlichkeitstheorien auf die Lehrereinschätzung denkbar (vgl. Kleber, 1992). Ebenso wäre es sinnvoll, eine größere und flächendeckende Stichprobe für derartige Studien zu generieren sowie auch die eingehendere Untersuchung der Lehrerurteile hinsichtlich anderer schulischer Leistungsfähigkeiten als Lesen und Rechtschreiben. Eine intensivere Beschäftigung mit der Einschätzung nicht-kognitiver Schülermerkmale kann möglicherweise auch Aufschluss über die Leistungseinschätzungen der Lehrer geben.

Im Sinne der großen Bedeutung und Selektionseigenschaft (Kleber, 1992), die Lehrerurteile für das Leben von Kindern haben, ist eine weitere Forschung sowie Umsetzung der daraus gewonnenen Erkenntnisse wünschenswert.

8 Abbildungsverzeichnis

Abbildung 1: Theoretisches Modell.....	31
Abbildung 2: Strukturgleichungsmodell Lesesicherheit.....	43
Abbildung 3: Strukturgleichungsmodell Lesegeschwindigkeit.....	48
Abbildung 4: Strukturgleichungsmodell Rechtschreibsicherheit.....	54

9 Tabellenverzeichnis

Tabelle 1: Resultat der Faktorenanalyse über den Faktor Leistung Lesesicherheit am Anfang und Ende der ersten sowie am Ende der zweiten und vierten Klasse	34
Tabelle 2: Resultat der Faktorenanalyse über den Faktor Leistung Lesegeschwindigkeit am Anfang und Ende der ersten sowie am Ende der zweiten und vierten Klasse	34
Tabelle 3: Resultat der Faktorenanalyse über den Faktor Leistung Rechtschreibsicherheit am Anfang der ersten Klasse	35
Tabelle 4: Resultat der Faktorenanalyse über den Faktor Einschätzung der Lesesicherheit am Anfang der ersten Klasse	36
Tabelle 5: Resultat der Faktorenanalyse über den Faktor Einschätzung der Rechtschreibsicherheit am Anfang der ersten Klasse	37
Tabelle 6: Resultat der Faktorenanalyse über den Faktor externalisierendes Verhalten	38
Tabelle 7: Gütekriterien	39
Tabelle 8: Effekte zwischen den Skalen des Pfadmodells Lesesicherheit	42
Tabelle 9: Messfehlerkorrelationen des Pfadmodells Lesesicherheit	43
Tabelle 10: Werte der Modellgüte für das Pfadmodell Lesesicherheit	46
Tabelle 11: Effekte zwischen den Skalen des Pfadmodells Lesegeschwindigkeit	47
Tabelle 12: Messfehlerkorrelationen des Pfadmodells Lesegeschwindigkeit	48
Tabelle 13: Werte der Modellgüte für das Pfadmodell Lesegeschwindigkeit	51
Tabelle 14: Effekte zwischen den Skalen des Pfadmodells Rechtschreibsicherheit	52
Tabelle 15: Messfehlerkorrelationen des Pfadmodells Rechtschreibsicherheit.....	53
Tabelle 16: Werte der Modellgüte für das Pfadmodell Rechtschreibsicherheit	57

10 Literaturverzeichnis

- 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., Gottesmann, 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.
- Bernard, M.E. (1979). Does sex role behavior influence the way teachers evaluate students? *Journal of Educational Psychology*, 71, 553-562.
- Bortz, J & Döring, N. (2006). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler*. 4. Auflage. Heidelberg: Springer Medizin Verlag.
- Bromme, R. (1987). Der Lehrer als Experte – Entwurf eines Forschungsansatzes. In H. Neber (Hrsg.), *Angewandte Problemlösungspsychologie* (S.127-151). Münster: Aschendorff.
- Brophy, J.E. & Good, T.L. (1974). *Teacher-Student-Relationships*. New York: Holt, Rinehart & Winston.
- Browne, M. & Cudeck, R. (1993). Alternative ways of assessing equation model fit. In K. A. Bollen und J.S. Long (eds.), *Testing Structural Equation Models* (S. 136-162). Newbury Park: Sage.
- Clark, C.M. & Peterson, P.L. (1986). Teacher's thought processes. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (pp.255-296). New York: Macmillan.
- Coladarci, T. (1986). Accuracy of teacher judgments of student responses to standardized test items. *Journal of Educational Psychology*, 78, 141-146.
- 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.
- Deutsches PISA-Konsortium (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.

- Eckert, T.L., Dunn, E.K., Coddling, R.S., Begeny, J.C. & Kleinmann, A.E. (2006). Assessment of mathematics and reading performance: An examination of the correspondence between direct assessment of student performance and teacher report. *Psychology in the Schools*, 43, 247-265.
- Egan, O. & Archer, P. (1985). The accuracy of teachers' ratings of ability: A regression model. *American Educational Research Journal*, 22, 25-34.
- Feinberg, A.B. & Shapiro, E.S. (2003). Accuracy of teacher judgments in predicting oral reading fluency. *School Psychology Quarterly*, 18, 52-65.
- 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.
- Helmke, A. (1994). Student needs: Diagnosing. In T. Husen & T.N. Postlethwaite (Eds.), *The international encyclopedia of education* (pp.5768-5772). Oxford: Pergamon.
- Helmke, A., Hosenfeld, I. & Schrader, F.W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. Arnold & C. Griese (Hrsg.), *Schulleitung und Schulentwicklung – Voraussetzungen, Bedingungen und Erfahrungen* (S.119-144). Hohengehren: Schneider Verlag.
- Hoge, R.D. & Butcher, R. (1984). Analysis of teacher judgments of pupil 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.
- Homburg Ch. & Giering A. (1996). Konzeptualisierung und Operationalisierung komplexer Konstrukte - Ein Leitfaden für die Marketingforschung. In L. Hildebrandt und Ch. Homburg (1998, Hrsg.), *Die Kausalanalyse* (S.111-146). Stuttgart: Schäffer Poeschel Verlag.

- Homburg Ch. & Baumgartner H. (1995), Beurteilung von Kausalmodellen: Bestandsaufnahme und Anwendungsempfehlungen. In L. Hildebrandt und Ch. Homburg (1998, Hrsg.), *Die Kausalanalyse* (S.111-146). Stuttgart: Schäffer Poeschel Verlag.
- Kleber, E.W. (1992). *Diagnostik in pädagogischen Handlungsfeldern. Einführung in Bewertung, Beurteilung, Diagnose und Evaluation*. Weinheim: Juventa Verlag.
- Klicpera, C., Gasteiger-Klicpera, B. & Schabmann, A. (1993). *Lesen und Schreiben. Entwicklung und Schwierigkeiten*. Bern: Verlag Hans Huber.
- 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.
- Krolak-Schwerdt, S. & Rummer, R. (2005). Der Einfluss von Expertise auf den Prozess der schulischen Leistungsbeurteilung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37 (4), 205-213.
- Landerl, K., Wimmer, H. & Moser, E. (1997). *SLRT – Salzburger Lese- und Rechtschreibtest*. Bern: Huber.
- Rosenthal, R. & Jacobson, L. (1968). *Pygmalion in the classroom: teacher expectation and pupils' intellectual development*. New York: Holt, Rinehart & Winston.
- Schabmann, A. (2001). Frühes Lesen und Rechtschreiben bei ganzheitlichem und synthetischem Erstleseunterricht. Habilitationsschrift an der Grund- und Integrativwissenschaftlichen Fakultät der Universität Wien, Institut für Klinische Psychologie.
- Schabmann, A., & Schmidt, B. M. (2009). Sind Lehrer gute Lese-Rechtschreibdiagnostiker? Der Einfluss von problematischem Schülerverhalten auf die Einschätzungen der Lesekompetenz durch Lehrkräfte. *Heilpädagogische Forschung*, XXXV, 3, 133-145.
- Schmidt, B.M. & Schabmann, A. (2010). „Es ist nur vorübergehend!“ Lehrereinschätzungen über mögliche Lese-Rechtschreibprobleme – eine klassifikatorische Analyse. *Heilpädagogische Forschung*, XXXVI, 3, 106-115.

- Schrader F.-W. (2009). Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. *Zeitschrift für Pädagogische Psychologie*, 23 (3-4), 237-245.
- Schrader, F.-W. (2006): Diagnostische Kompetenz von Eltern und Lehrern. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3. Aufl., S. 95–100). Weinheim: Beltz.
- Schrader, F.-W. und 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 (4), 312-324.
- Schrader, F.-W. und Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1, 27-52.
- Schuler, H. & Marcus, B. (2004). Leistungsbeurteilung. In H. Schuler (Hrsg.), *Enzyklopädie der Psychologie* (S. 947-1006), Göttingen: Hogrefe Verlag.
- Sharpley, C.F. & Edgar, E. (1986). Teachers` ratings vs. standardized tests: An empirical investigation of agreement between two indices of achievement. *Psychology in Schools*, 23, 106-111.
- Spinath, B. (2005). Akkuratheit der Einschätzungen von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19, 85-95.
- Thomas, L. (2007). Lern- und Leistungsdiagnostik. In T. Fleischer, N. Grewe, B. Jötten, K. Seifried & B. Sieland (Hrsg.), *Handbuch Schulpsychologie* (S.82-98). Stuttgart: Kohlhammer.
- Tournaki, N. & Podell, D.M. (2005).The impact of student characteristics and teacher efficacy on teacher`s predictions of student success. *Teaching and Teacher Education*, 21, 299-314.
- Weiber, R. & Mühlhaus, D. (2010). *Strukturgleichungsmodellierung. Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS*. Berlin, Heidelberg: Springer-Verlag.

11 Anhang

Curriculum Vitae

Persönliche Daten

Christina Ingrid Köhl

Wohnhaft in 1230 Wien

Geburtsdatum/-ort: 22.6.1981 / Wien

Staatsangehörigkeit: Österreich

Ausbildung

seit WS 2002	Studium der Psychologie an der Universität Wien
1999 – 2002	Pädagogische Akademie Ettenreichgasse, 1100 Wien
1991 – 1999	Neusprachliches Gymnasium Kollegium Kalksburg, 1230 Wien
1987 – 1991	Volksschule Haebergasse 1a, 1120 Wien

Berufliche Tätigkeiten

2009-2010	Praktikum an der Schulpsychologischen Beratungsstelle Wr. Neustadt
seit 2008	Volksschullehrerin
2007	Praktikum im Kaiser-Franz-Josef-Spital (Psychiatrie)
2005-2007	Trainerin der Lateintanzformation des FCP
2003	Sommerpraktikum im International Montessori Kindergarten
seit 2002	selbstständiger Nachhilfeunterricht und Kinderbetreuung

Sonstiges

2004	Seminar zur Anwendung des Rorschachtests
1998 – 2002	Jungschargruppenleiterin der Pfarre Rodaun
2004+2005	Staatsmeister im Formationstanzen 2004 & 2005