



universität
wien

Diplomarbeit

Titel der Arbeit

Die Qualität von LehrerInneneinschätzungen bei der
Identifikation von hoher kognitiver Leistungsfähigkeit bei
Kindern

Verfasserin

Katharina Klockgether

Angestrebter akademischer Grad

Magistra der Naturwissenschaften (Mag. rer. nat.)

Wien, im April 2011

Studienkennzahl: 298

Studienrichtung: Psychologie

Betreuerin: Ao. Univ.-Prof. Dipl.-Psych. Dr. Barbara Schober

Danksagung

Ich danke den Kindern und LehrerInnen, die mit viel Geduld und
großem Einsatz an der Studie teilgenommen haben.

Inhaltsverzeichnis

1 Einleitung.....	4
2 Theorie	6
2.1 Diagnostik von Leistungsfähigkeit.....	6
2.2 Diagnostische Kompetenz	8
2.3 Mögliche Einflussfaktoren auf die diagnostische Kompetenz.....	15
2.4 Identifikation von hoher Leistungsfähigkeit	18
2.5 Forschungsanliegen.....	25
3 Empirie	30
3.1 Methode	30
3.1.1 Geplante Stichprobe	30
3.1.2 Tatsächliche Stichprobe	30
3.1.3 Untersuchungsplan.....	32
3.1.4 Erhebungsinstrumente	33
3.2 Auswertungsverfahren	37
3.3 Ergebnisse	42
3.3.1 Deskriptive Ergebnisse der Stichprobe.....	42
3.3.2 Ergebnisse zur diagnostischen Kompetenz.....	43
3.3.4 Ergebnisse zur Identifikation der klassenintern leistungsfähigsten Kinder.....	49
3.3.5 Ergebnisse zur Identifikation der besonders leistungsfähigen Kinder.....	55
3.3.6 Ergebnisse zum Einfluss verschiedener LehrerInnenmerkmale auf die Einschätzung der LehrerInnen	60
4 Zusammenfassung und Diskussion	63
4.1 Kritik und Ausblick	68
4.2 Fördermöglichkeiten diagnostischer Kompetenz	69
5 Resümee	73
Literaturverzeichnis	75
Abbildungsverzeichnis	80
Tabellenverzeichnis	80
Anhang	84
Lebenslauf	104

1 Einleitung

Viele hohe kognitive Potenziale werden nicht erkannt. Ohne angepasste Förderung ist es wahrscheinlich, dass sich das Potenzial eines Kindes nicht ausreichend entfalten kann (Heller, 2001; Holling & Kanning, 1999). Neben der ExpertInnenmeinung wird in einer Standard-Umfrage (Seidl, 2010) deutlich, dass die meisten ÖsterreicherInnen der Meinung sind, dass in unseren Schulen viele Talente und Fähigkeiten übersehen werden und ungenutzt bleiben. Die Bedeutung, Kinder mit hoher Leistungsfähigkeit zu identifizieren, ergibt sich daraus, dass nur so eine spezifische Förderung erfolgen kann. Ziele hierbei sind, vorhandenes Talent zu entfalten, eventuellen Entwicklungsproblemen entgegenzuwirken sowie die Bildungslaufbahn betreffende Entscheidungen zu erleichtern (Wild, 1991).

Ausführliche Diagnostik mit Hilfe von Intelligenztests ist nach Rost (1991) eine verlässliche Methode, um kognitive Leistungsfähigkeit in hoher Ausprägung festzustellen. Holocher-Ertl, Kubinger und Hohensinn (2008) merken kritisch an, dass durch eine reine IQ-Diagnostik einige Kinder mit Hochleistungspotenzial, die zum Beispiel eine separierte, allerdings leicht behandelbare Lernstörung haben, nicht erkannt werden. Sie plädieren dafür, leistungsbezogene Persönlichkeitsmerkmale bei der Identifikation mit zu berücksichtigen. Dieser Ansatz ist sinnvoll, bei einer breit angelegten Studie allerdings aufgrund der Komplexität der Erfassung dieser Merkmale nur schwer realisierbar. Selbst Intelligenztestverfahren an sich sind im Gegensatz zu anderen Verfahren aufwändig in der Durchführung. Es müssen ausreichend Zeit, Geld für Personal und Material sowie Anstrengungsbereitschaft der Kinder gegeben sein. Daher sollten sie in einem diagnostischen Prozess erst nach einem ersten Screening angewendet werden. Heller (2001) empfiehlt ein sequentielles Vorgehen. Zunächst soll mit Checklisten, Nominationsverfahren, Ratingverfahren und Beobachtungstechniken ein Screening zum Zwecke einer Vorauswahl durchgeführt werden. Ein Screeningverfahren ist weniger zeitaufwändig, liefert allerdings auch weniger detaillierte Ergebnisse. Erst in einem zweiten Schritt sollen messgenauere Verfahren (z. B. Intelligenztests) eingesetzt werden. Dies ist allerdings nur möglich, wenn verlässliche Verfahren für die Screeningphase vorhanden sind. LehrerInnen und SchülerInnen verbringen im Lernalltag viel Zeit miteinander. Es liegt nahe, sowohl LehrerInnen als auch SchülerInnen als Informationsquelle heranzuziehen, um die Leistungsfähigkeit der SchülerInnen einzuschätzen (Wild, 1991; Hoge & Coladarci, 1989). LehrerInnen-

einschätzung sowie –nominationsverfahren (Neber, 2004) und auch Peernominationsverfahren (Peers = Gleichaltrige) scheinen hierfür geeignete und mit wenig Aufwand durchführbare Verfahren darzustellen. Die Verlässlichkeit der Anwendung dieser Instrumente bzw. die Einschätzkompetenz der LehrerInnen und SchülerInnen für die Identifikation besonders leistungsfähiger Kinder soll in dieser Arbeit überprüft werden.

Die vorliegende Arbeit orientiert sich dabei an einer ebenfalls in Österreich durchgeführten Studie. Lang untersuchte im Jahr 2008 eine Stichprobe von 13 Volksschulklassen der vierten Schulstufe (233 Kinder) und deren LehrerInnen in Oberösterreich. Die Lehrkräfte der Klassen wiesen jedem Kind der Klasse, jeweils anhand einer persönlichen Einschätzung, einen Prozentrang im mathematischen und sprachlichen Bereich zu. Die Einschätzung sollte gemäß der Leistungsfähigkeit des Kindes im Vergleich zum österreichischen Durchschnitt erfolgen. Außerdem wählten sie anhand einer Nominationsmethode die ein bis drei leistungsfähigsten Kinder im sprachlichen (geteilt in Fähigkeit, Geschichten zu erzählen und Fähigkeit, sich gut auszudrücken) und mathematischen Bereich aus. Die SchülerInnen der Klassen nominierten jeweils ein bis drei MitschülerInnen. Ihnen wurde dabei die Möglichkeit gelassen, sich selbst zu nennen. Vom Kognitiven Fähigkeitstest für 4. bis 12. Klassen, Revision (KFT 4-12+ R von Heller & Perleth, 2000) wurden den SchülerInnen jeweils zwei mathematische und sprachliche Subtests vorgegeben. Die Ergebnisse der Studie zeigen, dass bei vier Kindern im mathematischen Bereich ein Prozentrang von über 90 festgestellt werden konnte. Im verbalen Testteil zeigten drei Kinder eine Leistung, die einem Prozentrang von über 90 zugeordnet werden konnte. Bei solch einer geringen Anzahl an Kindern mit weit überdurchschnittlich guten Leistungen in der Stichprobe war es schwierig, eine verallgemeinernde Aussage über die diagnostische Kompetenz der LehrerInnen in diesem Bereich zu treffen. Um ergänzende Informationen über die diagnostische Kompetenz in Bezug auf die Identifikation von hoher Leistungsfähigkeit bei Kindern von österreichischen LehrerInnen zu erhalten, wurde die vorliegende Studie in Niederösterreich durchgeführt.

Diese Arbeit führt zunächst in die psychologische Diagnostik von Leistungsfähigkeit ein, gibt einen Überblick zum Thema diagnostische Kompetenz, zeigt die Möglichkeiten der Identifikation besonders hoher Leistungsfähigkeit und behandelt mögliche Einflüsse auf die LehrerInnen bei der Einschätzung ihrer SchülerInnen. Die Fragestellungen beziehen sich auf die Kompetenz der Lehrkräfte, die kognitive Leistungsfähigkeit ihrer SchülerInnen korrekt einzuschätzen sowie die Eignung verschiedener Methoden zur

Identifikation besonders hoher Leistungsfähigkeit. Die Durchführung der Studie wird erklärt, beschrieben und dokumentiert. Die Ergebnisse werden präsentiert und zum Abschluss werden die Ergebnisse und Folgerungen, mit Schwerpunkt auf Verbesserungsmöglichkeiten der diagnostischen Kompetenz, diskutiert.

2 Theorie

Das folgende Kapitel befasst sich zunächst mit dem Konstrukt der kognitiven Leistungsfähigkeit und den in der psychologischen Wissenschaft entwickelten Möglichkeiten, dieses Konstrukt zu erfassen. Im Weiteren werden die diagnostische Kompetenz von Lehrkräften und mögliche Einflüsse darauf diskutiert. Schließlich behandelt das Kapitel das Thema der Identifikation von besonders hoher Leistungsfähigkeit.

2.1 Diagnostik von Leistungsfähigkeit

Die psychologische Wissenschaft beschäftigt sich schon lange mit den Konstrukten der menschlichen Intelligenz und Begabung. Schon um 1900 untersuchte Galton „geniale“ Menschen und sammelte Informationen über die Unterschiede zwischen Menschen hinsichtlich ihrer kognitiven Begabung (Klauer & Leutner, 2010). Die Begriffe Begabung und (intellektuelle) Leistungsfähigkeit sollen im Folgenden identisch verwendet werden. Nach einer heute noch immer aktuellen Definition von Stern (1916, S. 110) sind "Begabungen [...] Möglichkeiten zur Leistung, unumgängliche Vorbedingungen; sie bedeuten jedoch nicht Leistung selbst." Auch Mühle (1968) betont in seiner Zusammenfassung bisheriger Definitionen die Ansicht vom Zusammenhang zwischen Höhe der kognitiven Begabung und Höhe der kognitiven Leistung.

Mit dem Konstrukt der Begabung entstanden auch bald erste Ansätze von Intelligenztests. Die Anfänge machten dabei Binet und Simon (1905, zitiert nach Klauer & Leutner, 2010), die kognitiv schwache Kinder erkennen wollten, um sie in Folge gezielter fördern zu können. Sie erstellten einen Test, bestehend aus 30 Aufgaben, mit welchem normalbegabte von minderbegabten Kindern unterschieden werden konnten. Dabei hielten die Forscher folgende Bereiche ausschlaggebend für die Bestimmung der Begabung: Gedächtnis, Vorstellungskraft, Aufmerksamkeit, Willensstärke, Suggestibilität, Motorik, Moral und Verständnis. Die Ergebnisse der Testungen wurden

ausgedrückt in einem Wert, den sie Intelligenzalter nannten. Das Intelligenzalter wurde durch die Anzahl gelöster Aufgaben im Vergleich zur Lösungswahrscheinlichkeit der Aufgaben, die zuvor an einer altersentsprechenden Normstichprobe berechnet wurde, bestimmt (Amelang, Bartussek, Stemmler & Hagemann, 2006).

Um Leistungsfähigkeit zu messen, werden auch heute noch üblicherweise standardisierte Intelligenztestverfahren herangezogen. Die intellektuelle Leistungsfähigkeit wird dabei mit Testkennwerten (meist Intelligenzquotient) dargestellt und mit einer normierten Stichprobe verglichen. Im Gegensatz zu anderen Vorgehensweisen hat sich dies in der Praxis am besten bewährt und scheint am objektivsten zu sein. Sie wird als das nützlichste Messinstrument in diesem Bereich bewertet (Holling & Kanning, 1999, zitiert nach Schweizer, 2006). Die heutigen psychologischen Intelligenztestverfahren können meist als Einzel- oder Gruppenverfahren vorgegeben werden. Die Durchführung in der Gruppe ermöglicht es innerhalb kurzer Zeit eine große Menge an Daten zu erfassen. Dies hat im Besonderen für die Ökonomie bei Forschungsvorhaben einen Vorteil. Die Differenziertheit und Aussagekraft der Ergebnisse ist allerdings eingeschränkt, da während der Testung weniger Verhaltensbeobachtungen vorgenommen werden können und weniger auf besondere Bedürfnisse der Testperson eingegangen werden kann (Kubinger, 2006).

Neben der weit verbreiteten Anwendung von Intelligenztests können auch andere Verfahren eingesetzt werden, um die Leistungsfähigkeit zu bestimmen. So zieht man zum Beispiel Urteile des sozialen Umfeldes der Person heran. In solch einer Beurteilung kann man, im Gegensatz zu einer stichpunktartigen Messung, Beobachtungen über einen längeren Zeitraum mit in die Bewertung einbeziehen (Hagen, 1989). Möchte man die Leistungsfähigkeit von Schulkindern untersuchen, setzt sich das relevante soziale Umfeld aus LehrerInnen (und evtl. anderen BetreuerInnen), MitschülerInnen und Eltern zusammen. LehrerInnen teilen den Lernalltag mit den SchülerInnen und haben häufig die Gelegenheit, sie in leistungsrelevanten Situationen zu beobachten. Zudem können sie in ihrer Berufslaufbahn viel Erfahrung in der Einschätzung von Leistungsfähigkeit sammeln, und durch die große Anzahl der betreuten Kinder bieten sich ihnen viele Vergleichsmöglichkeiten (Rost et al., 2006). LehrerInnen sind außerdem über die öffentlich zugänglichen Schulen leicht zu erreichen und stellen somit eine gut zugängliche Stichprobe dar.

Als Alternative zu der Anwendung von Intelligenztest kann hier exemplarisch die Methode der LehrerInneneinschätzung (Ziegler & Stöger, 2003), bei Wild (1991) zu finden als Lehrerrating, angeführt werden: Die LehrerInnen weisen jedem Schüler und jeder Schülerin einen Wert zu, der den Grad der Begabung widerspiegeln soll. Meist wird die Einschätzung auf einer Skala von 1 bis 100 vorgenommen und liefert somit eine passende Grundlage zum Vergleich der eingeschätzten Werte mit Testwerten aus Intelligenztestverfahren in Prozenträngen, welche ebenfalls von 1 bis 100 angegeben werden. Solche Einschätzungen sind weniger objektiv als standardisierte Testungen. Verzerrende Effekte können durch Urteilsfehler entstehen. Ein typischer Urteilsfehler ist der Halo-Effekt, welcher ursprünglich auf Thorndike (1920, zitiert nach Bortz & Döring, 1995) zurückzuführen ist. Menschen tendieren bei ihren Einschätzungen von Eigenschaften anderer dazu, Ausprägungen eines Bereiches auf andere Bereiche zu übertragen. Dies kann beispielsweise zu einer fälschlichen Generalisierung des Einschätzers von einer hohen sprachlichen Kompetenz des Beobachteten auf eine hohe mathematische Kompetenz führen (Bortz & Döring, 1995).

Zusammenfassend ist zu sagen, dass die Einschätzung kognitiver Leistungsfähigkeit meist mit standardisierten Intelligenztestverfahren vorgenommen wird. Die Personen, die Teil der Lernumfeldes des zu testenden Objektes sind, können ebenfalls Informationen über die Begabung liefern und sollten deshalb in einen diagnostischen Prozess mit einbezogen werden. Um persönliche Einschätzungen vergleichbar zu machen, muss bei der Erhebung auf möglichst standardisierte Bedingungen und klare Anweisungen, was genau und wie eingeschätzt werden soll, geachtet werden.

2.2 Diagnostische Kompetenz

Die Brauchbarkeit der Ergebnisse einer persönlichen Einschätzung der Leistungsfähigkeit durch LehrerInnen, wie sie zum Beispiel bei der erwähnten LehrerInneneinschätzung vorgenommen wird, ist abhängig von der Kompetenz der Lehrperson diesbezüglich. Zum Ende der Volksschule hin gewinnen die Leistungseinschätzungen und Leistungsfähigkeitseinschätzungen der Kinder durch ihre KlassenlehrerInnen immer mehr Relevanz. Die LehrerInnen sind ausschlaggebende Personen bei der Entscheidung, welche weiterführende Schule ein Kind besuchen wird und tragen somit große Verantwortung darüber, ob eine Person den für ihre Leistungsfähigkeit passenden Bildungsweg einschlägt, oder aber auf einen Weg

gelangt, auf dem sie unter- oder überfordert wird. Durchlässigkeit in eine höhere Schule nach der Übergangsentscheidung ist in Österreich kaum gegeben (Spinath, 2010).

Eine falsche Entscheidung kann weit reichende Folgen haben, da eine schlechte Passung von Anforderung und personeninternen Ressourcen zu Motivationsverlust und Resignation bis hin zu einem Ausbildungsabbruch führen kann (van Ophuysen, 2006). Der Typ der weiterführenden Schule und somit die Entscheidung der LehrerInnen haben außerdem Einfluss auf das Fähigkeitsselbstkonzept der SchülerInnen. Das Fähigkeitsselbstkonzept bezeichnet „generalisierte Überzeugungen zu den eigenen Fähigkeiten in einem bestimmten Handlungsbereich“ (Pekrun, Frenzel & Götz, 2010, S.115). So sinkt beispielsweise das Fähigkeitsselbstkonzept bei GymnasiastInnen, da sie sich nach dem Übergang von der Volksschule nun einer Bezugsgruppe gegenüber finden, die in ihren Leistungen im Durchschnitt höher ist als die in der Volksschule (Pekrun, Frenzel & Götz, 2010). Bei einem fälschlicherweise auf ein Gymnasium überwiesenen Kind verstärkte sich dieser Effekt noch, weil das Kind seine Leistungen im Vergleich zur neuen Bezugsgruppe als stark unterdurchschnittlich empfindet.

Außerdem hat die Einschätzung der LehrerInnen Einfluss auf die wahre Leistung des Kindes in Richtung der Einschätzung der Lehrkraft (Hinnant, O'Brien & Ghazarian, 2009). Hinnant et al. nennen als Grundlage dieses Effekts die self-fulfilling prophecy nach Merton (1984), nach welcher die Erwartung über das Eintreffen eines Ereignisses das Ereignis selbst hervorruft. Eine Sonderform dieses Phänomens ist der Pygmalioneffekt, ursprünglich beschrieben von Rosenthal und Jacobson (1971, zitiert nach Hesse & Latzko, 2009). Die Erwartungen der Lehrkraft über die Begabung und damit zusammenhängende Leistung ihrer SchülerInnen beeinflusst ihr Verhalten gegenüber diesen. Dies wiederum ruft bei den SchülerInnen ein der LehrerInnen-erwartung angepasstes Verhalten hervor. Ein schlecht eingeschätzter Schüler oder eine schlecht eingeschätzte Schülerin, von dem bzw. der schlechte Leistung erwartet wird, verhält sich tendenziell auch in Richtung dieser Erwartung.

In der Literatur wird diagnostische Kompetenz als wichtig für eine an die SchülerInnen angepasste Unterrichtsgestaltung angesehen. Hoher Lernerfolg auf Seiten der SchülerInnen ist dann eine Folge von hoher diagnostischer Kompetenz der LehrerInnen, wenn gleichzeitig eine starke Strukturierung des Unterrichts vorhanden ist (Schrader, 1989). Kolb und Jussim (1993) gingen auf die Auswirkungen des Pygmalioneffekts bei kognitiv besonders leistungsfähigen Kindern ein. Diese Kinder fühlten sich teilweise

aufgrund der Diskrepanzen der schulischen Anforderungen und ihrer Begabung im schulischen Kontext unterfordert und reagierten darauf mit auffälligem Verhalten (z. B. Stören des Unterrichts). LehrerInnen interpretierten dies als niedrige soziale Kompetenz und erwarteten dadurch eine niedrige Leistung der Kinder als von unauffälligen SchülerInnen. Diese Erwartung spiegelte sich im Verhalten der LehrerInnen wider und übertrug sich auf das Kind. Das Kind wiederum verhielt sich entsprechend und zeigte niedrigere Leistung als nach dem Potenzial zu erwarten war.

Das bisher Angeführte unterstreicht die Wichtigkeit einer genauen Einschätzung der Leistungsfähigkeit der SchülerInnen. Da das Konstrukt der diagnostischen Kompetenz der LehrerInnen die Güte der Leistungseinschätzungen widerspiegelt, soll im Folgenden näher darauf eingegangen werden. Schrader (1989; 2006) bezeichnet diagnostische Kompetenz von Lehrkräften als die Fähigkeit, urteilende Aufgaben angemessen und zielführend auszuführen, sodass eine optimale pädagogische Entscheidung gefällt werden kann. Dies bedeutet, dass die jeweiligen Ausprägungen der leistungsbezogenen Merkmale der SchülerInnen möglichst genau eingeschätzt werden. Der Autor sieht das Ausmaß der diagnostischen Kompetenz abhängig von vier Teilbereichen: „Allgemeine Fähigkeiten“ der Lehrkraft, worunter in erster Linie Intelligenz verstanden wird; „Bereichsspezifische Fertigkeiten und Wissensstrukturen“, wie etwa didaktisches Wissen; „Spezifische Kenntnisse“ über die Eigenschaften der SchülerInnen und „Spezifische Fertigkeiten“ und Methoden, um an relevante Informationen über die SchülerInnenleistungen zu gelangen. Der Großteil dieser Teilbereiche ist durch Lernen und Erfahrung veränderbar und somit auch förderbar. Diese Definition ist von vielen ForscherInnen im Bereich der diagnostischen Kompetenz im pädagogischen Umfeld übernommen worden, obwohl die Eingrenzung auf die Bereiche empirisch nicht belegt ist. Eine Operationalisierung der einzelnen Teilbereiche und somit eine Möglichkeit, diese zu messen, wurde ebenso wenig unternommen und somit wird sich in der Praxis für die tatsächliche Einschätzung der diagnostischen Kompetenz auf folgendes Konzept gestützt. Schrader und Helmke (1987) haben drei Komponenten zur Berechnung der diagnostischen Kompetenz von Lehrkräften vorgeschlagen. Diese sind in vergleichbaren Studien zur Anwendung gekommen und haben sich als aussagekräftig bewährt. Die Niveauelemente dient zur Einschätzung der absoluten Höhe des Leistungsdurchschnitts der Klasse, die Differenzierungskomponente zur Einschätzung der Streuung der Leistungen in der Klasse, während die Vergleichskomponente zur Einschätzung der Rangreihung, bzw. der Fähigkeitsabstufungen der SchülerInnen

innerhalb der Klasse, herangezogen wird. Letztere wird von den Autoren als eigentliche diagnostische Kompetenz angesehen und auch von den meisten ForscherInnen als einziges Maß zur Bewertung der Urteilsgenauigkeit herangezogen. Beispielsweise setzen Schrader und Helmke (2001) den Schwerpunkt der Beurteilung der diagnostischen Kompetenz auf den klasseninternen Vergleich. Der Einschätzung des Niveaus wird weniger Beachtung geschenkt, obwohl diese bei der Identifikation von besonders begabten Kindern oder bei der weiterführenden Schulwahl eine wichtige Rolle spielen kann. Der Grund für diese Vernachlässigung ist aus der Literatur nicht nachzuvollziehen.

Die diagnostische Kompetenz bezieht sich nicht nur auf die Einschätzung kognitiver Leistungsfähigkeit, sondern kann sich ebenso auf die Einschätzung von Interesse, Selbstkonzept, Attributionen, Aufmerksamkeit, Ängstlichkeit, Lernmotivation, Aufgabenschwierigkeit und andere Konstrukte beziehen. Spinath (2005) berichtete beispielsweise von LehrerInneneinschätzungen der Leistungsmotivation und der leistungsbezogenen Ängstlichkeit der eigenen Klasse. Im Durchschnitt zeigte sich eine Unterschätzung der Motivation und eine Überschätzung der Ängstlichkeit (Niveauelemente). Die Streuung der Leistungsmotivation innerhalb der Klasse wurde überschätzt, während die der Ängstlichkeit leicht unterschätzt wurde (Differenzierungselemente). Auch die Vergleichselemente zeigten für die Einschätzung der Motivation und der Ängstlichkeit geringe Genauigkeit. Zwischen den verschiedenen LehrerInnen ergaben sich allerdings erhebliche individuelle Unterschiede. Die übliche methodische Herangehensweise in den Studien zur diagnostischen Kompetenz von Lehrkräften, die Vergleichselemente zu bestimmen, ist, Korrelationen zwischen den Einschätzungen und den Testwerten zu bilden (Schrader & Helmke, 2001).

Während die diagnostische Kompetenz hinsichtlich der Einschätzung von Leistungsmotivation und leistungsbezogener Ängstlichkeit im Durchschnitt gering ausgeprägt zu sein scheint, zeigen Ergebnisse zur Einschätzung von kognitiver Leistungsfähigkeit leicht positivere Ergebnisse. Hoge und Coladarsi führten schon 1989 eine Metaanalyse zum Zusammenhang zwischen LehrerInneneinschätzung und Testwerten aus psychologischen Tests mit größtenteils amerikanischen Studien durch. Neben mittleren Korrelationen von $r = .62$ und $r = .69$ berichteten sie von starken Unterschieden zwischen den einzelnen LehrerInnen. Auch in neueren Studien werden LehrerInneneinschätzungen in der Literatur als moderat akkurat beurteilt. Beispielsweise

fanden sich bei Hinnant et al. (2009) signifikante Zusammenhänge zwischen den Einschätzungen der LehrerInnen und der Leistungsfähigkeit von SchülerInnen der ersten bis fünften Schulstufe, gemessen mit einem standardisierten Mathematik- und Lesetest. Nach Schrader und Helmke (2001) schätzten LehrerInnen die SchülerInnen der eigenen Klasse in der Rangfolge meist genau ein. Bei der Einschätzung des Gesamtniveaus unterliefen jedoch häufig Fehler. Die Lehrkräfte orientierten sich eher innerhalb der Klasse als an der Leistung eines Durchschnittsschülers oder einer Durchschnittsschülerin. Auffällig wurde dies, wenn zwei Kinder von ihren LehrerInnen sehr unterschiedlich beurteilt wurden, obwohl sie in einem objektiven Leistungstests gleich gut abschnitten. Wild und Rost (1995) berichteten von einem durchschnittlichen Zusammenhang von $r = .67$ (Korrelationskoeffizient) zwischen Intelligenztestergebnissen und LehrerInneneinschätzungen (Vergleichskomponente) und von einer starken Spannweite der Einschätzungen der Lehrkräfte. Zu einem ähnlichen Ergebnis kam Lang (2008) durch ihre Studie in Oberösterreich. Die Durchschnittswerte der Vergleichskomponenten sprachen mit $r = .5$ (mathematischer Bereich) und $r = .53$ (sprachlicher Bereich) für einen mittleren Zusammenhang der Einschätzung und der Leistungsfähigkeit der Kinder. Auch hier zeigten sich starke Unterschiede zwischen den LehrerInnen, die Werte schwankten stark zwischen $r = .01$ und $.65$ sowie $r = .27$ und $.78$. Ebenso führte Spinath (2005) eine Untersuchung der Güte diagnostischer Urteile der Lehrkräfte über u. a. Intelligenzkennwerte von SchülerInnen durch. Die LehrerInnen schätzten dabei die kognitive Leistungsfähigkeit auf einer fünfstufigen Skala ein. Dabei wurden nach dem Konzept von Schrader und Helmke (1987) Niveau-, Differenzierungs- und Vergleichskomponente gebildet. In einer weiteren Studie untersuchten Lorenz und Artelt (2009) die Güte diagnostischer Urteile im mathematischen und verbalen Bereich, die Fachspezifität der Güte und die Stabilität der Güte. Die Fähigkeiten der Schüler wurden in dieser Studie im sprachlichen Bereich mit dem Wortschatztest des CFT-20 (Weiß, 1998) und einem Untertest zum Textverständnis des Lesetests ELFE 1-6 (Lenhard & Schneider, 2006) erfasst. Im mathematischen Bereich wurde ein Arithmetik-Untertest des DEMAT 4 (Gölitz, Roick & Hasselhorn, 2006) verwendet. Die Lehrkräfte schätzten die einzelnen SchülerInnen in den jeweiligen drei getesteten Bereichen auf einer fünfstufigen Skala ein. Um die diagnostische Kompetenz der LehrerInnen zu bestimmen, wurden ebenfalls nach dem Vorbild von Schrader und Helmke (1987) die Vergleichskomponenten anhand von Pearsons Produkt-Moment-Korrelationen zwischen LehrerInnenurteil und den durch die Tests gemessene SchülerInnenleistungsfähigkeit

gebildet. Die gemittelten Zusammenhänge von LehrerInnenurteilen und Testleistungen der Kinder wiesen bei Lorenz und Artelt (2009) Werte um $r = .6$ auf. Es konnte ein signifikanter Zusammenhang bewiesen werden. Darüber hinaus wurde herausgefunden, dass eine hohe diagnostische Kompetenz der LehrerInnen im sprachlichen Bereich nicht zwingend mit einer hohen diagnostischen Kompetenz im mathematischen Bereich einherging. Die Stabilität der Einschätzungen über ein halbes Jahr konnte als zufrieden stellend bewertet werden, insbesondere im sprachlichen Bereich. Die AutorInnen schlossen daraus, dass „diagnostische Kompetenz als bereichsspezifisches und zeitlich stabiles Konstrukt“ angesehen werden könne (S. 220). Ebenso wurden bei Spinath (2005) im Mittel Zusammenhänge zwischen LehrerInnenurteil und Intelligenzkennwert der Kinder von $r = .43$ gemessen, was von der Autorin als treffende Einschätzung bewertet wurde. Sie wies jedoch das Konstrukt der diagnostischen Kompetenz an sich zurück, da zwischen den drei gemessenen Komponenten (Niveau-, Differenzierungs- und Vergleichskomponente) nach ihren Ergebnissen kaum Korrelationen vorhanden sind, und, wie auch bei Lorenz und Artelt (2009), die Urteils Kompetenzen in unterschiedlichen Bereichen nicht konsistent sind. Es muss beachtet werden, dass das Konstrukt der diagnostischen Kompetenz von den ForscherInnen unterschiedlich bewertet wurde. Lorenz und Artelt (2009) hielten es für sinnvoller, die Kompetenz für die einzelnen Domänen spezifisch zu betrachten und sehen eine mangelnde Korrelation zwischen den Einschätzungen in den unterschiedlichen Fächern nicht als Gegenargument zum Bestehen eines Konzepts „diagnostische Kompetenz“.

Die meisten Studien beschäftigen sich ausschließlich mit der Vergleichskomponente. Nun soll auf einige Untersuchungen eingegangen werden, die ebenfalls das Niveau der Einschätzung im Vergleich zum Niveau der Testwerte mit berücksichtigen. Überschätzungstendenzen wurden im Zusammenhang mit PISA 2000 (Programm zur internationalen Schülerbewertung) von Hesse und Latzko (2009) berichtet. Bei einer Befragung von HauptschullehrerInnen, die eine Einschätzung der Lesekompetenzen der SchülerInnen beinhaltete, wurden 90 % der Kinder, die nach der PISA-Testung das niedrigste Kompetenzlevel erreichten, von den LehrerInnen auf ein höheres Level eingeschätzt. Auch in der Studie von Hosenfeld, Helmke und Schrader (2002) wurde das Niveau der SchülerInnenleistungen überschätzt. Lehrkräfte sollten die Leistungen ihrer SchülerInnen in einem Mathematiktest vorhersagen. Die Aufgaben der Testung waren ihnen dabei bekannt und sie sollten einschätzen, wie viele Kinder alle zwölf Aufgaben lösen würden. Die Lehrkräfte überschätzten in 20 von 21 Fällen die Lösungshäufigkeit

der Klassen durchschnittlich um 18 %. In ihrer Studie von 1987 zeigten Helmke und Schrader ebenfalls Niveauüberschätzungen auf. In 35 von 39 Fällen vermuteten LehrerInnen eine höhere Leistung bei ihren SchülerInnen, als sich durch die Testung nachweisen ließ. Die Vergleichskomponente lag bei ihren Untersuchungen bei $r = .64$. Lang (2008) stellte in ihrer oberösterreichischen Untersuchung ebenfalls Überschätzungen des Niveaus fest. Ihre Ergebnisse müssen allerdings relativ zu den generell sehr geringen Intelligenztestleistungen der Kinder gesehen werden, da sehr niedrige Werte eine Überschätzung des Niveaus wahrscheinlicher machen. Die Vergleichskomponenten lagen bei ihren Untersuchungen im Durchschnitt bei $r = .5$ im mathematischen Bereich und $r = .53$ im sprachlichen Bereich. Spinath (2005) hingegen berichtete von spezifischen Fällen der Unterschätzung. Sie kam zu dem Ergebnis, dass LehrerInnen sich bei ihren Urteilen über Leistungsfähigkeit sehr stark an der schulischen Leistung der SchülerInnen orientierten. So wurde zum Beispiel die kognitive Leistungsfähigkeit von Underachievern unterschätzt. Underachievement bedeutet nach Hanses und Rost (1998) schlechte schulische Leistungen ($PR < 50$) (Performanz) bei vorhandener überdurchschnittlicher kognitiver Begabung ($PR > 96$) (Kompetenz). 75 bis 80 % der Varianz, welche bei den Leistungen in der Schule zu beobachten ist, kann nicht anhand von Intelligenzmaßen vorhergesagt werden.

Ergebnisse zu Differenzierungskomponenten finden wir bei Schrader und Helmke (1987) sowie bei Lang (2008). Die teilgenommen LehrerInnen der Studie im Jahre 1987 überschätzten die Streuung der Leistung innerhalb der Klassen im Durchschnitt. Lang (2008) hingegen berichtete von einer Überschätzung der Streuung. In beiden Studien wurden außerdem starke Differenzen der diagnostischen Kompetenz zwischen den LehrerInnen festgestellt.

Zusammenfassend sind ein mittlerer Zusammenhang zwischen LehrerInneneinschätzungen und Testwerten zur kognitiven Leistungsfähigkeit und weitergehend starke interindividuelle Unterschiede zwischen den Lehrkräften festzustellen. Untersuchungen zur Einschätzung des Niveaus und der Streuung sind kaum vorhanden.

Der eher geringe Zusammenhang zwischen objektiven Testergebnissen und Einschätzungen der LehrerInnen geben Anlass zum Zweifel an einer ausreichenden diagnostischen Kompetenz von LehrerInnen in Bezug auf die Leistungsfähigkeit ihrer SchülerInnen. Somit könnten Kinder in ihrer Leistungsfähigkeit nicht genau genug eingeschätzt werden und ihnen in weiterer Folge ein für sie unpassender Bildungsweg

von den LehrerInnen empfohlen werden. Dies könnte den Kindern eine angemessene Förderung vorenthalten.

2.3 Mögliche Einflussfaktoren auf die diagnostische Kompetenz

Die Forschung zeigt starke Unterschiede im Ausmaß der diagnostischen Kompetenz zwischen den LehrerInnen. Systematische Unterschiede in anderen Eigenschaften der LehrerInnen oder systematisch variierende Umweltunterschiede können mögliche Einflussfaktoren auf die diagnostische Kompetenz sein. Karing (2009) erfasste in ihrer Untersuchung von 111 Lehrkräften und ihren Klassen die Streuung der Leistung innerhalb der Klassen und die Kooperation zwischen Eltern und LehrerInnen. Eine geringere Streuung der Leistungsfähigkeit der Klasse hing signifikant mit einer genaueren Einschätzung durch die LehrerInnen zusammen. Ähnliche Ergebnisse berichtete Schrader im Jahre 1989. Bezüglich der Kooperation zwischen LehrerInnen und Eltern konnte Karing keinen signifikanten Zusammenhang feststellen. Unterschiede konnte sie dagegen im Gruppenvergleich von Grundschul- und GymnasiallehrerInnen zum Vorteil der GrundschullehrerInnen feststellen. Die Gruppe der GrundschullehrerInnen schätzte die Leistungsfähigkeit der Kinder in Arithmetik, Wortschatz und Textverstehen akkurater ein als die GymnasiallehrerInnen. Berechnet wurden Korrelationen, die nach Schrader und Helmke (1987) die Vergleichskomponenten darstellen. Im Durchschnitt lagen die Niveauelementen bei den GrundschullehrerInnen bei $r = .6$, bei den GymnasiallehrerInnen bei $r = .45$. In einer Studie von Wild und Rost (1995) wurde der Einfluss von Klassengröße auf die diagnostische Kompetenz untersucht. Es wurden keine Zusammenhänge festgestellt. Schrader (1989) vermutete, die Bezugsnormorientierung der LehrerInnen könne mit der diagnostischen Kompetenz zusammenhängen. Es wurde dabei überprüft, ob sich die LehrerInnen generell eher an einer sozialen oder an einer individuellen Bezugsnorm orientierten. Die soziale Bezugsnorm legt den Fokus auf den Vergleich eines Kindes mit seinen MitschülerInnen; die individuelle auf die Lernfortschritte des Einzelnen. Bei der Untersuchung konnten keine Unterschiede in der diagnostischen Kompetenz der beiden Gruppen festgestellt werden. Zu berücksichtigen ist dabei, dass sich ein sozialer Bezugsrahmen, wie er nach Spinath (2010) normalerweise im schulischen Primärbereich bei der Notengebung verwendet wird, in erster Linie auf die Niveauelemente auswirken würde. Die Bewertung hängt bei solch einer Orientierung immer ausschließlich von der Verteilung

der Begabungen der Kinder innerhalb der Klasse ab. Die Einschätzung der Rangreihung dagegen ist eher unabhängig von der Bezugsnorm, während der ausschließlich klasseninterne Vergleich die Einschätzung der Leistungsfähigkeit einer Klasse verzerren kann. So könnte eine im landesweiten Vergleich leistungsstarke Klasse von der Lehrkraft im Niveau unterschätzt werden, woraus sich ein Nachteil für die Kinder im Bezug auf ihren weiteren Ausbildungsverlauf ergibt. In einem weiteren Bereich konnte Schrader (1989) tendenzielle Unterschiede festmachen: Attribuierten die Lehrkräfte generell schlechte Leistung der SchülerInnen auf unzureichende Begabung und bewerteten diese folglich als wenig beeinflussbar, besaßen sie auch eher eine niedrige diagnostische Kompetenz. Die Attribution der LehrerInnen von schlechten SchülerInnenleistungen auf stabile, personeninterne Eigenschaften der SchülerInnen scheint mit ihrer diagnostischen Kompetenz zusammenzuhängen. Hier könnte eine Variable gefunden worden sein, die zumindest niedrige diagnostische Kompetenz vorhersagen kann. Verschiedene Forscher haben den Zusammenhang von diagnostischer Kompetenz und Unterrichtsexpertise untersucht. Van Ophuysen (2006) geht davon aus, dass sich die Prozesse, welche bei der Urteilsfindung über die Leistungsfähigkeit der SchülerInnen bei der Lehrkraft ablaufen, mit wachsender Unterrichtsexpertise verändern. Es zeigte sich in ihrer Studie, dass LehrerInnen, die schon mehr Erfahrung im Unterrichten haben, genauere Einschätzungen der Leistungsfähigkeit abgaben als Lehramtstudierende. Hoge und Coladarci (1989) bemängelten, dass in den von ihnen verglichenen Studien der Metaanalyse die Unterrichtsexpertise nicht berücksichtigt wurde und forderten ihre Erfassung in der Zukunft, um möglicherweise die starken Schwankungen zwischen der diagnostischen Kompetenz unterschiedlicher Lehrkräfte erklären zu können. Wild (1991) berücksichtigte diese Moderatorvariable in seiner Studie, konnte aber keine Unterschiede zwischen erfahrenen (über fünf Jahre Berufspraxis) und weniger erfahrenen (ein bis fünf Jahre Berufspraxis) feststellen. Hesse und Latzko (2009), Wild und Rost (1995) sowie Schrader (1989) berichteten ebenfalls, dass kein signifikanter Zusammenhang zwischen der Berufserfahrung und diagnostischer Kompetenz bestehe.

Eine weitere Einflussstärke auf die diagnostische Kompetenz könnte die Unterrichtsdauer in der beurteilten Klasse sein. Wild (1991) vermutete hier einen Zusammenhang: Je mehr Zeit mit der Klasse verbracht werde, desto mehr Möglichkeiten gebe es, Informationen zu sammeln, die bei der Einschätzung helfen könnten. In seiner Untersuchung ergaben sich hingegen keine Unterschiede zwischen

LehrerInnen, die ihre Klasse im ersten und denen, die ihre Klasse schon im dritten Jahr unterrichteten. Im Jahre 1995 untersuchten Wild und Rost diese Moderatorvariable erneut und stellten einen tendenziellen Vorteil längerer Unterrichtsdauer für die Genauigkeit der Einschätzung fest, allerdings nur bei Klassen mit einer SchülerInnenanzahl von 15 oder weniger.

Im Speziellen bei der Identifikation besonderer Begabung liegt als mögliche Moderatorvariable das Wissen über Hochbegabung nahe. Brown et al. (2005) zeigten, dass Erfahrungen im Gebiet kognitiver Hochleistungen LehrerInnen dazu veranlasst, mögliche Fehler, die bei der Identifikation von Hochbegabten auftreten könnten, zu berücksichtigen. Es kann demnach vermutet werden, dass Lehrkräfte, die Vorwissen haben, sich der Schwierigkeiten der Identifikation bewusst sind und eher wissen, worauf zu achten ist. Dies könnte einen Vorteil im Identifikationsprozess darstellen. Neben diesem Vorwissen über Hochbegabung scheint auch die Vertrautheit der LehrerInnen mit dem angewendeten Test eine Rolle zu spielen. Hurwitz, Elliott und Braden (2007) fanden heraus, dass die Ergebnisse schulnaher Leistungstests von den LehrerInnen genauer hervorgesagt werden konnten als generelle Leistungstests. Die Auswahl des Tests in einer Studie zur diagnostischen Kompetenz sollte deshalb wohl überlegt und an die Fragestellung angepasst erfolgen. Die Autoren fanden außerdem heraus, dass es Lehrkräften mehr Schwierigkeiten bereitete, SchülerInnen mit körperlichen Behinderungen als SchülerInnen ohne körperliche Behinderungen in ihren kognitiven Leistungen korrekt einzuschätzen. Die nordamerikanische Studie zeigte, dass behinderte Kinder von ihren LehrerInnen häufiger unterschätzt wurden als diejenigen ohne Behinderungen. Sie sprachen sich deswegen für besondere Rücksichtnahme auf diesen Verzerrungseffekt bei der Leistungsdiagnose der Kinder aus.

Zusammenfassend ergeben die Studien zu Einflussfaktoren auf die diagnostische Kompetenz kein einheitliches Bild. Eine wahrscheinliche Ursache dafür könnte sein, dass mehrere Faktoren in komplexer Weise zusammenwirken. Eine systematische Untersuchung stellt sich deshalb als schwierig dar.

Die Ergebnisse legen nahe, dass etliche Einflussgrößen und Merkmale von LehrerInnen bei weitergehenden Interpretationen zu beachten sind. Möglicherweise helfen Erkenntnisse über Einflussgrößen, die potenziell auf die diagnostische Kompetenz wirken, bei einer Einteilung von LehrerInnen in eher gute und weniger gute

DiagnostikerInnen oder bei der Erklärung besonders hoher oder niedriger diagnostischer Kompetenz.

2.4 Identifikation von hoher Leistungsfähigkeit

Ein Teilbereich der diagnostischen Aufgaben, die von LehrerInnen übernommen werden müssen, ist die Identifikation von besonders leistungsfähigen Kindern. Sie sollen möglichst früh erkannt werden, um die passende Förderung zu erhalten und möglichen Defiziten, die durch die besonderen Umstände der Begabung entstehen können, entgegenzuwirken. Sollen aus einer großen Gruppe die Leistungsfähigsten, zum Beispiel zum Zwecke einer spezifischen Förderung, ausfindig gemacht werden, ergibt sich das Problem, dass die Testung mit einem Intelligenztestverfahren jedes Einzelnen zu aufwändig wäre. Es werden deshalb verschiedenste Zugänge gesucht, nicht nur durch Befragung von LehrerInnen, die Identifikation von Kindern mit besonders hoher Leistungsfähigkeit weniger aufwändig zu gestalten. Dies geht einher mit der Billigung von Einbußen in der Genauigkeit. Mit solchen Vorgehensweisen kann Zeit und Geld gespart werden, und, je nach Zweck des diagnostischen Prozesses, kann solch ein Screening unterschiedliche Vorteile mit sich bringen.

Für diese Art von Screening besonderer Leistungsfähigkeit werden, wie noch gezeigt werden wird, verschiedenste Verfahren in Anspruch genommen. Ein Ziel ist dabei nun, Verfahren zu finden, die besonders geeignet für diese Identifikation sind. Um die Eignung verschiedener Methoden für die Identifikation bewerten zu können und miteinander vergleichbar zu machen, haben Pegnato und Birch (1959, zitiert nach Wild, 1991) zwei Formeln aufgestellt. Sie gehen von einem Kriterium und einem Prädiktor aus. Das Kriterium ist dabei das Ergebnis aus einer Intelligenztestung, der Prädiktor die Alternativmethode zur Einschätzung der kognitiven Leistungsfähigkeit. Berechnet werden jeweils ein Effektivitäts- und ein Effizienzindex. Diese stellen den Ausschöpfungsgrad und die Ökonomie des Vorgehens dar. Hohe und ausgeglichene Indizes stehen für eine gute Eignung der Methode.

Im Folgenden werden einige Alternativen zum klassischen Intelligenztest zur Identifikation von besonders hoher kognitiver Begabung vorgestellt und ihre Eignung diskutiert. Nach Hagen (1989) ist es von Vorteil, LehrerInnen und Eltern schon früh in den Identifikationsprozess potenziell besonders Begabter mit einzubeziehen, da sie später, wenn es um Förderungsfragen geht, eine ausschlaggebende Rolle spielen. Fels

(1999) sieht die Identifikation besonders begabter Kinder durch Lehrkräfte in der Schule als „einfachstes und kostengünstigstes Verfahren“ (S. 135) an. Er betont jedoch, dass verschiedene Forschungsergebnisse vermuten lassen, dass LehrerInnen nicht in der Lage sind, die besonders leistungsfähigen Kinder erschöpfend zu erkennen. Eine Studie von Heller, Reimann und Senfter (2005) unterstützt diese Annahme. Ungefähr 650 deutsche Dritt- und Viertklässler wurden mit dem Münchner Hochbegabungs-Testsystem für die Primarstufe (MHBT-P; Heller & Perleth, 2007) auf ihre kognitive Leistungsfähigkeit getestet. Die Aufgabe der LehrerInnen bestand darin, die zehn Prozent der Begabtesten mit Hilfe einer Checkliste, ebenfalls aus dem MHBT-P, zu identifizieren. Die Ergebnisse zeigen, dass nur die Hälfte derjenigen genannt wurde, die im Test die beste Leistung demonstrierten. Dies macht deutlich, dass viele Potenziale übersehen wurden. Auch Ziegler und Stöger (2003) machten deutlich, dass LehrerInnen insbesondere Schwierigkeiten hatten, begabte Kinder zu identifizieren, die in der Schule keine überdurchschnittlichen Leistungen zeigten (i. e. Underachiever). In dieser Studie waren die SchülerInnen selbst am ehesten in der Lage, ihre Leistungsfähigkeit auf einer fünfstufigen Skala, entsprechend den nach einem Intelligenztest ermittelten Werten, einzuschätzen. Die Einschätzung erfolgte hier anhand der Frage, wie talentiert das Kind sei. Die im Vergleich zu anderen Studien geringen Zusammenhänge zwischen den LehrerInneneinschätzungen sowie den Elterneinschätzungen und den Testwerten könnten aus dieser sehr allgemein gehaltenen Frage resultieren. Es liegt nahe, dass die Befragten unter anderem die sprachliche Leistungsfähigkeit bei ihren Überlegungen mit einbezogen, die mit dem in der Studie verwendeten Standard Progressive Matrices (SPM; Raven, Raven & Court, 1999) nicht berücksichtigt wird. In einem Review-Artikel beschäftigten sich Reis und McCoach (2002) mit den Schwierigkeiten des Underachievement bei besonders begabten Kindern. SchülerInnen mit einem besonders hohen kognitiven Potenzial, die aber aufgrund einer Lernstörung Probleme bei der Umsetzung ihres Potenzials haben (i. e. Underachiever) wurden nur schwer erkannt. Die Autorinnen erstellten, basierend auf eigenen Beobachtungen, eine Liste von Charakteristika, anhand derer die Identifikation leichter fallen sollte. Durch eine frühzeitige, korrekte Identifikation könnte eine passende Förderung der Kinder durchgeführt und die Wahrscheinlichkeit des Underachievement verringert werden.

Es zeigte sich, dass nicht nur LehrerInnen und die Kinder selbst, sondern auch Peers eine wichtige Informationsquelle darstellen. Sie kennen ihre MitschülerInnen in den unterschiedlichsten Situationen und u. a. in solchen, zu denen andere keinen Zugang

haben (Banbury & Wellington, 1989, zitiert nach Schweizer, 2006). Auch Hagen (1989) sprach sich für die Einholung von Informationen durch Gleichaltrige zur Identifikation besonderer Begabung aus.

Die schon in der Einleitung erwähnten Nominationsverfahren stellen eine sehr ökonomische Vorgehensweise bei der Identifikation von hoher Leistungsfähigkeit dar. Bei diesen Verfahren muss nicht jeder einzelne Schüler oder jede einzelne Schülerin beurteilt, sondern lediglich die von den Beurteilern als besonders leistungsfähig eingeschätzten Kinder genannt werden. Informationen über die weniger leistungsfähigen SchülerInnen werden in diesem Fall nicht benötigt. Ein weiterer ökonomischer Vorteil besteht in der dichotomen Entscheidung im Gegensatz zu einer komplexen Einschätzung auf einer Skala (Wild, 1991). Meist werden LehrerInnennominationen, Peernominationen oder Selbstnominationen vorgenommen. Einen Überblick über Studien, bei denen die Wirksamkeit von LehrerInnennomination untersucht wurde, gab Wild (1991). Die errechneten Werte schwankten, abhängig von herangezogenem Kriterium, stark. Neun verschiedene Studien wiesen Effektivitätswerte von 0 bis 86 % und Effizienzwerte von 0 bis 78 % auf. Die divergierenden Ergebnisse lassen keine eindeutigen Schlüsse über die Brauchbarkeit von LehrerInnennominationen zu. Eine weitere Methode ist die Peernomination, bei der alle SchülerInnen nach dem begabtesten Kind der Klasse gefragt werden. Wild trug 1991 aus 20 Studien Ergebnisse zusammen und sieht bei einem Großteil der Studien methodische Mängel. Im Durchschnitt konnte er jedoch eindeutige Ergebnisse herausfiltern. Es bestanden signifikante Zusammenhänge zwischen Peernomination und Intelligenztestergebnissen. In seiner eigenen Studie wendete er ebenfalls Peernomination an.

Ab der dritten Klasse der Volksschule besteht laut Jäger und Sitarek (1986, zitiert nach Wild, 1991) ein Zusammenhang zwischen der Selbsteinschätzung der kognitiven Leistungsfähigkeit und Intelligenztestwerten. Nominiert ein Kind der vierten Klasse sich selbst, kann dieser Einschätzung demnach schon Relevanz beigemessen werden. Auch Hagen (1989) bewertete Peernomination ab einem Alter von neun oder zehn Jahren als geeignet. Ziegler und Stöger (2003) zeigen in ihrer schon oben erwähnten Untersuchung exemplarisch auf, dass sich Schüler selbst besser einschätzen konnten als LehrerInnen und Eltern.

Wild (1991) verglich in seiner Studie die Effektivität und Effizienz nach Pegnato und Birch (1959, zitiert nach Wild, 1991) von LehrerInnen-, Peer- und Selbstnomination. Es

sollten möglichst viele der Kinder, die einen IQ > 130 hatten, im Screening erkannt werden. LehrerInnen und Peers konnten 75 bis 85 % der Kinder identifizieren, während nur die Hälfte der besagten Kinder sich selbst nominierte. Zwischen LehrerInnen- und Peernomination konnte kein Unterschied festgestellt werden. Es könnte vermutet werden, dass die niedrige Erfolgsrate der Kinder mit dem niedrigen Durchschnittsalter der Stichprobe zusammen hängt. Der Großteil der Kinder war zum Zeitpunkt der Erhebung acht oder neun Jahre alt. Im Gegensatz dazu setzte sich die Stichprobe bei Ziegler und Stöger (2003), in deren Studie eine hohe Erfolgsquote der Selbsteinschätzung berichtet wurde, ausschließlich aus ViertklässlerInnen zusammen. Der Altersdurchschnitt ist in der Studie nicht angegeben. Es kann jedoch vermutet werden, dass er in deutschen Grundschulklassen eher im späten neunten oder Mitte des zehnten Lebensjahres und somit höher als in Wilds Untersuchung liegt. Allerdings wäre nach Nicholls (1978) Aussagen zu erwarten gewesen, dass sich die jüngeren Kinder in ihrer Leistungsfähigkeit eher überschätzen als unterschätzen. Der Forscher schloss aus seinen Untersuchungen, dass Kinder ab einem Alter von zehn Jahren verlässlich in der Lage sind, zwischen Anstrengung und Fähigkeit zu unterscheiden und somit ihre eigene Leistungsfähigkeit realistischer einschätzen als in einem jüngeren Alter. Die Ergebnisse zur Selbstnomination von Wild (1995) erscheinen deshalb unerwartet. Es muss allerdings beachtet werden, dass in den Studien von Ziegler und Stöger (2003) und Wild (1995) unterschiedliche Erhebungsverfahren verwendet wurden und dass es auch um die Erfassung unterschiedlicher Dinge ging. Während bei Wild den Kindern eine bipolare Entscheidung zur Selbstnomination überlassen wurde, ob sie besonders begabt seien oder nicht, mussten sich die Kinder bei Ziegler und Stöger (2003) in ihrer Leistungsfähigkeit auf einer fünfstufigen Skala einschätzen. Im einen Fall wurde ausdrücklich nach besonders Begabten „gesucht“, andererseits ging es um eine generelle Einschätzung der Begabung.

Nominationsverfahren müssen kritisch betrachtet werden. Die Aufforderung, die besten SchülerInnen der Klasse zu nennen, führt zur Verwendung einer klasseninternen Referenz. Die Befragten orientieren sich bei der Entscheidung, ob sie eine Person als „besonders leistungsfähig“ angeben, ausschließlich an der Leistungsfähigkeit der SchülerInnen innerhalb der Klasse. Ob ein Kind im Vergleich zur Gesamtpopulation herausragende Leistungen zeigt oder nur innerhalb der Klasse, kann in manchen Fällen einen großen Unterschied ausmachen. Wenn es in einer Klasse keine, im Vergleich zur Normstichprobe, überdurchschnittlich Leistungsfähigen gibt, können auch keine

identifiziert werden. Man wird durch das Verfahren jedoch gezwungen, mindestens eine Person zu nennen. Dies kann zu einer Verzerrung der Ergebnisse führen. Fels (1999) führte einige Vor- und Nachteile von Peernomination an. Forschung aus den USA zeigt, dass überdurchschnittlich begabte Kinder auch besonders beliebt waren und auch deshalb häufiger nominiert wurden, was positiv zu einer höheren Effektivität des Verfahrens beitrug (Pasternak & Silvey, 1969, zitiert nach Fels, 1999). Der Autor stellte allerdings in Frage, ob diese Ergebnisse auf den deutschsprachigen Raum übertragbar seien, wo „möglicherweise eine andere Einstellung zu Leistungsdruck und Konkurrenzdenken besteht“ (S. 140). Die Auswahl von besonders Leistungsfähigen sei zumindest in Deutschland selten und daher ungewöhnlich und irritierend für die TeilnehmerInnen. Fels vermutete, im Gegensatz zu oben berichteten Ergebnissen aus den USA, dass besonders leistungsfähige Kinder durch ihr häufig aus der Norm fallendes Verhalten, in der Klasse abgelehnt werden könnten und dies Einfluss auf die Nomination habe. Hany (1987, zitiert nach Fels, 1999) sah aufgrund mangelnder Vergleichsmöglichkeiten auf Seiten der Kinder insbesondere bei der Identifikation extrem Begabter Schwierigkeiten bei der Peernomination. Insgesamt riet Fels zur Peernomination, allerdings nur als Zusatz zu anderen Verfahren.

Mögliche Effektivitäts- und Effizienzgewinne durch eine Kombination mehrerer Methoden (LehrerInnenurteil, -nomination, Peer- und Selbstnomination) hat Wild (1991) untersucht. Er kam zu dem Schluss, dass das Vorgehen keine nennenswerte Verbesserung mit sich bringe. Von den untersuchten Methoden solle nach Meinung des Autors das LehrerInnenurteil angewendet werden, da es den Nominationsmethoden leicht überlegen sei.

Neben den diskutierten und auch in der vorliegenden Studie verwendeten LehrerInnen- und Peernominationen, gibt es noch weitere Methoden zur Identifikation besonders begabter Kinder. Häufig werden in der Praxis Checklisten zur Identifikation von besonderer Begabung verwendet (vgl. weiter oben: Studie von Heller et al., 2005). Diese sind meist eher willkürlich aus verschiedensten Quellen gewonnene Erkenntnisse über (vermeintliche) Unterschiede von besonders Begabten im Vergleich zu anderen Kindern zur Verbesserung der Urteile von Ratern (z. B. Eltern, LehrerInnen...) und können nicht erschöpfend sein (Urban, 1992). Rost kritisierte schon 1991, dass Checklisten vielen wissenschaftlichen Anforderungen nicht genügen. Sie wiesen etwa mangelnde Validität der Items und keine zufrieden stellende Ergebnisse bei der Evaluation auf. Checklisten stellten eher Gelegenheitsbeobachtungen statt systematischerer Vorgehensweisen dar

und seien allenfalls eine Ansammlung von Einzelbefunden der Forschung, die auf die Gruppe der besonders Begabten generalisiert werden. Auf Grund der mangelhaften Erfüllung wissenschaftlicher Ansprüche führte Rost (1991) an, dass häufig „aufgeweckte“ (S. 217) Kinder durch Checklisten fälschlicherweise als hochbegabt eingestuft werden (Fehler erster Art, s. S. 40), die Items sehr allgemein formuliert seien und häufig große (per se schon fehlerbehaftete) Erinnerungsleistungen aus dem Langzeitgedächtnis von z. B. Eltern abverlangt werden. Außerdem beinhalten Checklisten Fragen, deren Beantwortung nicht durch reine Beobachtung möglich sei. Eine sehr neue, auf psychometrische Güte geprüfte Checkliste für GrundschülerInnen findet sich in Sommer, Fink und Neubauer (2008). Alle Items sind validiert und mit Angaben von Trennschärfeindizes versehen. In der deutschen Studie von Heller et al. (2005) kam eine Checkliste im wissenschaftlichen Kontext zum Einsatz. Bei der Untersuchung wurden einerseits Testurteile aus einer Intelligenztestung mit einer spezifischen Hochbegabungs-Testbatterie, andererseits LehrerInnenurteile in Form von Checklisten zur Hochbegabungseinschätzung erhoben. Die Lehrkräfte konnten anhand der Checklisten etwas mehr als die Hälfte der Kinder, die nach den Ergebnissen der Intelligenztestung zu den 10 % Leistungsfähigsten gehörten, identifizieren. Die AutorInnen bewerten dies als nicht zufrieden stellend.

Darüber hinaus gibt es einige wenige brauchbare Beobachtungsbögen, welche zur Identifikation besonders leistungsfähiger Kinder verwendet werden können. Exemplarisch wird der Beobachtungsbogen zur Erkennung von Kindern mit hohen Fähigkeiten von Huser (1999) vorgestellt. Er soll von dem oder der betreuenden LehrerIn und den Eltern nach einer ein- bis zweimonatigen Beobachtungszeit ausgefüllt werden. Der Fragebogen ist unterteilt in die Bereiche „Allgemeine Merkmale“, „Merkmale von schulisch unterforderten Kindern“, „Sprachliche Intelligenz“, „Mathematische Intelligenz“, „Inter- und Intrapersonale Intelligenz“, „Naturalistische Intelligenz“ und „Kreatives Denken und Tun“ und deckt somit ein breiteres Spektrum als ausschließlich hohe kognitive Begabung ab. Im Laufe der Beobachtung werden Punkte, je nach Entscheidung des Beobachters über die unterschiedlichen Items, vergeben. Die Punktezahl indiziert die Höhe der Wahrscheinlichkeit einer besonderen Begabung. Huser (1999) betont, dass der Beobachtungsbogen nicht als einziges Diagnostikum bei der Entscheidung über eine besondere Begabung herangezogen werden sollte, sondern vielmehr als Gesprächsgrundlage diene.

Zusammenfassend kann gesagt werden, dass in der Forschung bei der Identifikation von besonders hoher Leistungsfähigkeit als Alternative zu Intelligenztests häufig LehrerInnenurteile herangezogen werden. Direkte Einschätzungen der LehrerInnen über die Leistungsfähigkeit der Kinder sind dabei häufig Nominationsmethoden, bei denen die Leistungsfähigsten benannt werden, überlegen. Die Ergebnisse zur Tauglichkeit von Nominationsmethoden sind allerdings sehr heterogen.

Mittels der angeführten Literatur konnte gezeigt werden, dass Alternativen zu Intelligenztestungen bei der Identifikation von besonders begabten Kindern in der Schule bestehen. Eindeutige Aussagen über Vor- und Nachteile sowie eine Festlegung einer qualitativ hervorstechenden Methode ist aufgrund der unklaren Forschungslage allerdings nicht möglich.

2.5 Forschungsanliegen

Ziel der Arbeit war es, querschnittlich an einer Stichprobe von österreichischen Volksschulkindern der vierten Klasse zu untersuchen, mit welchen Screeningverfahren besonders leistungsfähige Kinder am besten identifiziert werden können. Ein besonderer Schwerpunkt wurde dabei auf die Einschätzung der diagnostischen Kompetenz der LehrerInnen gelegt. Grundlage der Arbeit stellte dabei eine Studie von Lang (2008) dar, in der 13 oberösterreichische Volksschulklassen als Stichprobe dienten. Als Untersuchungsgegenstand wurden in vorliegender Studie SchülerInnen der vierten Klasse Volksschule ausgewählt, weil gerade in dieser Zeit die Einschätzung der Leistungsfähigkeit in Hinsicht auf den bevorstehenden Schulübergang eine große Rolle spielt. Viele Studien zur diagnostischen Kompetenz wurden in dieser Altersstufe durchgeführt, was zu einer hohen Vergleichbarkeit untereinander führt (z. B. Lorenz & Artelt, 2009). Da die Entfaltung besonderer Leistungsfähigkeit von den spezifischen kulturellen, sozialen und schulischen Gegebenheiten eines Landes abhängen, wurde die skizzierte Fragestellung, zu der es bereits eine umfangreiche internationale Literatur gibt, wie schon bei Lang (2008), unter den konkreten Bedingungen des österreichischen Schulsystems und der damit verbundenen LehrerInnenausbildung, bearbeitet.

Als objektive Referenz für die Leistungsfähigkeit der SchülerInnen wurden Ergebnisse einer psychologischen Testung mit dem KFT 4-12+ R (Heller & Perleth, 2000) verwendet. Aus Gründen der Praktikabilität wurde ein Gruppenverfahren angewandt. Ein weiterer wesentlicher Grund für die Anwendung eines Gruppenverfahrens war das Bestreben, die Hemmschwelle zur Teilnahmebereitschaft auf Seiten der Kinder möglichst gering zu halten. Somit mussten sie sich keiner persönlichen Testung unterziehen, in der sie sich, noch stärker als bei einer Testung in einer großen Gruppe, direkt beobachtet und bewertet fühlen könnten.

Für die Beurteilung der diagnostischen Kompetenz der Lehrkräfte wurde eine LehrerInneneinschätzungsmethode als Datenquelle herangezogen. Die LehrerInnen schätzten hierbei jedes einzelne Kind bezüglich seiner mathematischen und sprachlichen Leistungsfähigkeit mit einem Prozentrang von 1 bis 100 ein. Die Bewertung der LehrerInneneinschätzung orientierte sich an den Komponenten (Niveauekomponente, Differenzierungskomponente und Vergleichskomponente) von Schrader und Helmke (1987). Weitere Teilbereiche der diagnostischen Kompetenz, die in dem Komponentenkonzept keine Berücksichtigung finden (wie z. B. didaktisches Wissen),

konnten in dieser Arbeit nicht beachtet werden, da eine Möglichkeit der Operationalisierung und Bewertungsansätze nicht gegeben sind. Hervorzuheben ist dabei, dass, im Unterschied zu vielen anderen Arbeiten zur diagnostischen Kompetenz, allen drei Komponenten Beachtung geschenkt wurde. In der Literatur konnte kein überzeugendes Argument gefunden werden, welches einen Ausschluss der Differenzierungs- und Niveauelemente hinlänglich begründete. Daher soll in dieser Arbeit überprüft werden, ob die beiden Komponenten für die Güte der LehrerInnenurteile neben der Vergleichskomponente zusätzliche Information liefern.

In einem weiteren Teil der vorliegenden Untersuchung wurden verschiedene Screeningmethoden zur Identifikation von besonderer Leistungsfähigkeit auf ihre Tauglichkeit überprüft und miteinander verglichen. Hierfür wurde der Ansatz von Pegnato und Birch (1959, zitiert nach Wild, 1991) herangezogen, da dieser in vielen ähnlichen Studien zur Anwendung kam und somit die Möglichkeit des Vergleichs der Ergebnisse liefert. Das Kriterium für die Effizienz- und Effektivitätsberechnung war im Falle der vorliegenden Arbeit das Ergebnis aus der Testung mit dem KFT 4-12+ R; der Prädiktor wurde durch die Einschätzung der LehrerIn bzw. der Kinder repräsentiert. Außerdem sollte, wie auch Wild (1991) empfiehlt, überprüft werden, ob eine Kombination verschiedener Methoden eine Optimierung der Ergebnisse mit sich bringe.

Die Nominationsverfahren wurden nach dem Vorbild von Lang (2008) übernommen, um gleiche Bedingungen und damit Vergleichbarkeit zu gewährleisten. Die Ergebnisse von Ziegler und Stöger (2003) zeigten, dass die Selbstnomination der SchülerInnen eine gute Einschätzung hoher Leistungsfähigkeit mit sich brachte. Daher wurde in der vorliegenden Studie die Selbstnomination im Rahmen der Peernomination mit berücksichtigt. Diese war im Rahmen der Peernomination möglich.

Einschlägige Forschungsergebnisse machen es wahrscheinlich, dass verschiedene Größen Einfluss auf die diagnostische Kompetenz von Lehrkräften haben. Um mehr über die Bedingungen hoher diagnostischer Kompetenz zu erfahren und um gezieltere Förderung zu ermöglichen, wurden in der vorliegenden Arbeit der Einfluss von Moderatorvariablen und ihre Wechselwirkungen überprüft.

Eine mögliche Moderatorvariable stellt die Berufserfahrung dar. Van Ophuysen (2006) fand im Gegensatz zu anderen ForscherInnen einen Zusammenhang von Unterrichtsexpertise (im Folgenden als Berufserfahrung bezeichnet) als Moderatorvariable und diagnostischer Kompetenz. Da die Forschungsergebnisse

diesbezüglich nicht eindeutig sind, wurde die Variable in dieser Studie zusätzlich beachtet. Auf Grundlage der Daten wurden zwei Gruppen der Berufserfahrung gebildet. Die Stichprobe umfasste vier LehrerInnen, die Berufserfahrung zwischen einem und fünf Jahren aufwiesen. Diese wurden der Gruppe „wenig Berufserfahrung“ zugewiesen. Im Gegensatz dazu hatten sieben LehrerInnen Berufserfahrung, die schon mindestens neun Jahre ausmachte. Diese Gruppe wurde mit „viel Berufserfahrung“ bezeichnet.

Wild und Rost (1995) haben tendenzielle Zusammenhänge zwischen kurzer und langer Unterrichtsdauer in der für die Untersuchung relevanten Klasse und der diagnostischen Kompetenz feststellen können. Andere Studien zeigten keine Unterschiede zwischen der diagnostischen Kompetenz von Lehrkräften, die schon seit Beginn der Volksschulzeit die Klasse unterrichten und solchen, die die Klasse erst später übernommen haben. Aufgrund der heterogenen Ergebnisse sollte der Einfluss der Unterrichtsdauer in der vorliegenden Studie erneut überprüft werden. Die Aufteilung in zwei Gruppen (kurze vs. lange Unterrichtsdauer) wurde auf empirischer Grundlage vorgenommen. Es befanden sich in der Stichprobe vier LehrerInnen, die ihre aktuelle Klasse seit der dritten oder erst seit der vierten Klasse unterrichteten. Diese wurden der Gruppe „kurze Unterrichtsdauer“ zugeteilt. Sieben LehrerInnen unterrichteten dagegen schon seit der ersten Klasse (Gruppe „lange Unterrichtsdauer“) und hatten somit mindestens doppelt so lange Zeit, Erfahrungen mit den Kindern zu sammeln.

Als dritte mögliche Moderatorvariable sollte das Wissen über Hochbegabung, bzw. die Teilnahme an einer Fortbildung zum Thema Hochbegabung dienen. Brown et al. (2005) vermuteten, dass ein positiver Zusammenhang zwischen Wissen über Hochbegabung und Höhe der diagnostischen Kompetenz bestehe. Dies wurde als Ausgangspunkt dafür genommen, diese Variable in der vorliegenden Arbeit zu überprüfen. Die LehrerInnen der Stichprobe wurden in zwei Gruppen unterteilt. Einige LehrerInnen hatten mindestens eine Fortbildung zu Hochbegabung besucht (Gruppe „Fortbildung HB besucht“), andere wiesen keine Teilnahme an einer vergleichbaren Fortbildung auf (Gruppe „Fortbildung HB nicht besucht“). Es wurden darüber hinaus mögliche Wechselwirkungen zwischen den potenziellen Moderatorvariablen überprüft.

Im Einzelnen wurden folgende Fragestellungen bearbeitet.

Diagnostische Kompetenz

1. Wie ist die diagnostische Kompetenz der LehrerInnen in Bezug auf die Einschätzung der mathematischen bzw. sprachlichen Leistungsfähigkeit ihrer SchülerInnen (mit Hilfe der Einschätzungsmethode: Zuweisung eines Prozentranges von 1 bis 100 zu jedem Kind der Klasse) zu bewerten?

Identifikation der klassenintern leistungsfähigsten Kinder

2. Wie gut sind die LehrerInnen in der Lage, die drei besten SchülerInnen der Klasse im mathematischen bzw. sprachlichen Leistungsbereich mit Hilfe der Einschätzungsmethode zu identifizieren?
3. Wie gut sind die LehrerInnen in der Lage, die drei besten SchülerInnen der Klasse im mathematischen bzw. sprachlichen Leistungsbereich mit Hilfe der Nominationsmethode (Nennung der drei leistungsfähigsten Kinder der Klasse) zu identifizieren?
4. Wie gut sind die SchülerInnen in der Lage, die drei besten SchülerInnen der Klasse im mathematischen bzw. sprachlichen Leistungsbereich mit Hilfe der Nominationsmethode zu identifizieren?
5. Kann durch eine Kombination der Einschätzungsmethode, der LehrerInnen- und der Peernomination, eine höhere Trefferanzahl der besten SchülerInnen erreicht werden?

Identifikation von hoher Leistungsfähigkeit, bezogen auf die österreichische Gesamtpopulation

6. Wie effektiv und effizient können mit Hilfe der Einschätzungsmethode als Screeningmethode besonders leistungsfähige SchülerInnen (PR > 85) im mathematischen bzw. sprachlichen Bereich identifiziert werden?
7. Wie effektiv und effizient können mit Hilfe der LehrerInnennomination als Screeningmethode besonders leistungsfähige SchülerInnen (PR > 85) im mathematischen bzw. sprachlichen Bereich identifiziert werden?
8. Wie effektiv und effizient können mit Hilfe der Peernomination als Screeningmethode besonders leistungsfähige SchülerInnen (PR > 85) im mathematischen bzw. sprachlichen Bereich identifiziert werden?

9. Kann durch eine Kombination der Einschätzungsmethode, der LehrerInnen- und der Peernomination die Effektivität und Effizienz der Identifikation besonders leistungsfähiger SchülerInnen im mathematischen bzw. sprachlichen Bereich erhöht werden?

Merkmale, die Einfluss auf die Qualität der Einschätzung haben könnten

10. Wird die diagnostische Kompetenz des Lehrers/ der Lehrerin durch die Berufserfahrung des Lehrers/ der Lehrerin moderiert?
11. Wird die diagnostische Kompetenz des Lehrers/ der Lehrerin durch die Unterrichtsdauer in der beurteilten Klasse moderiert?
12. Wird die diagnostische Kompetenz Lehrers/ der Lehrerin durch Wissen des Lehrers/ der Lehrerin über Hochbegabung moderiert?
13. Wird die diagnostische Kompetenz des Lehrers/ der Lehrerin durch eine Kombination der genannten Einflussgrößen moderiert?

3 Empirie

Im folgenden Kapitel wird zunächst auf die Zusammensetzung der untersuchten Stichprobe im Vergleich zu derjenigen, die Lang (2008) untersuchte, eingegangen. Im Anschluss werden die eingesetzten Methoden, die Verfahren der Auswertung der Daten und die Ergebnisse der Studie diskutiert und beschrieben.

3.1 Methode

3.1.1 Geplante Stichprobe

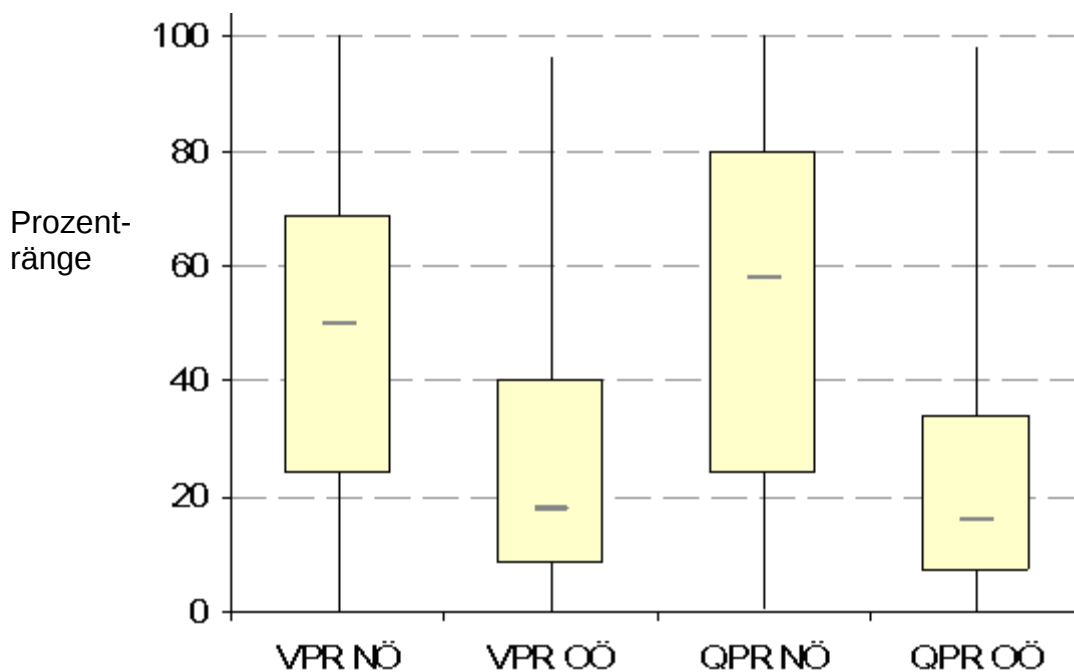
Anknüpfend an die Ergebnisse aus Langs (2008) österreichischer Studie, die aufgrund der geringen Anzahl an besonders leistungsfähigen Kindern in der Stichprobe als wenig aussagekräftig bewertet wurde, sollte diese erweitert und eine Analyse der Gesamtdaten durchgeführt werden.

Die Stichprobe aus der Untersuchung von Lang (2008) besteht aus 13 Volksschulklassen der vierten Schulstufe (233 Kinder) und deren KlassenlehrerInnen aus dem Großraum Linz in Oberösterreich. Die Daten wurden an acht unterschiedlichen Schulen erfasst. Geplant war eine Erweiterung um weitere elf bis 13 vierte Klassen (ca. 230 Kinder) im Raum Niederösterreich, was zu einer Gesamtstichprobe von ca. 460 Kindern führen sollte.

3.1.2 Tatsächliche Stichprobe

Die im Rahmen dieser Arbeit erhobene Stichprobe setzt sich aus elf Volksschulklassen der vierten Schulstufe (219 Kinder) und ihren KlassenlehrerInnen aus Niederösterreich zusammen. Die Stichprobe umfasst 108 Mädchen und 111 Jungen. Das Alter der Kinder liegt zwischen neun und zwölf Jahren, das Durchschnittsalter beträgt 10,34 Jahre mit einer Standardabweichung von 0,49. Es wurde an fünf verschiedenen Volksschulen erhoben. Eine Zufallsstichprobe war aufgrund der Abhängigkeit von Bereitschaft der Lehrkräfte und SchuldirektorInnen und aus ökonomischen Gründen nicht möglich, ist aber im Hinblick auf den Untersuchungszweck nicht unabdingbar.

Bei der Analyse der Daten zeigte sich, dass große Unterschiede in den Ergebnissen der Intelligenztestung zwischen der Stichprobe von Lang (2008) und der vorliegenden Stichprobe bestehen. T-Tests wiesen signifikante Unterschiede (2-seitige Signifikanz $p < .05$) sowohl zwischen den Mittelwerten des mathematischen Testteils (QPR) als auch zwischen den Mittelwerten des verbalen Testteils (VPR) auf (siehe Abb. 1). Die Freiheitsgrade lagen bei 450 (QPR und VPR). Die T-Werte bei -12,47 (QPR) und -10,27 (VPR). Die Voraussetzungen der Homogenität der Varianzen waren erfüllt.



Legende: VPR = Verbaltestergebnisse, QPR = Quantitativtestergebnisse.

Abb. 1: Verteilung der Quantitativtestwerte und Verbaltestwerte, Stichprobe NÖ (Niederösterreich) . N = 219 und Stichprobe OÖ (Oberösterreich) aus Lang (2008). N = 233

Weiter zeigte sich, dass der Anteil der Kinder, die Deutsch als Muttersprache haben, in der alten und der neuen Stichprobe unterschiedlich groß ist. Während in der neuen Stichprobe 50 (von 219) Kinder (22,8 %) eine andere Sprache als Deutsch als Muttersprache angaben, war dies in der alten Stichprobe bei 119 (von 233) Kindern (51,1 %) der Fall. Vergleich man im zusammengenommenen Datensatz die Testergebnisse zwischen den Gruppen „Deutsch als Muttersprache“ (N = 283) und „Andere Muttersprache“ (N = 169), ergaben sich erheblich signifikante Unterschiede sowohl im mathematischen als auch im sprachlichen Bereich zugunsten der Kinder mit Deutsch als Muttersprache (T-Test, 2-seitige Signifikanz, $p < .05$). Die Voraussetzung der Homogenität der Varianzen war gegeben. Es ergaben sich bei 450 Freiheitsgraden

T-Werte von 8,46 (VPR) und 6,39 (QPR). Der systematische Faktor Muttersprache trug somit erheblich zur großen Heterogenität der ursprünglich geplanten, aus zwei Untersuchungen zusammen gesetzten Stichprobe bei.

Im Manual des KFT 4-12+ R (Heller & Perleth, 2000) wurde auf mögliche Unterschiede in den Testergebnissen zwischen türkischsprachigen Kindern und deutschsprachigen Kindern zugunsten der deutschsprachigen aufmerksam gemacht. Die Stichprobe dieser im Handbuch angeführten Untersuchung kann allerdings kaum als repräsentativ bewertet werden, da sie von einer einzigen Schule in München stammt. Aufgrund dieser Ergebnisse der Testentwickler und der vorhandenen Heterogenität der zwei Stichproben (vorliegende Studie und Lang, 2008), wurde auf eine kombinierte Auswertung der oberösterreichischen und der niederösterreichischen Stichproben verzichtet. Bei einer Auswertung der zusammengelegten Stichproben wären die Ergebnisse stark verzerrt, unnatürlichen Charakters und damit nicht aussagekräftig. Es ist darum sinnvoller, die zwei Stichproben getrennt voneinander zu betrachten.

3.1.3 Untersuchungsplan

Für die Erhebung der neuen Daten wurden aus der Liste aller öffentlichen Schulen in Österreich diverse Volksschulen, aus ökonomischen Gründen der Erhebung, im näheren Umkreis von Wien (NÖ) kontaktiert, bis eine mündliche Zusage von fünf Volksschulen vorlag. Die Genehmigung vom Landesschulrat für NÖ wurde eingeholt. Die Erhebungen fanden einmalig vormittags im Mai und Juni 2010 statt. Pro Klasse betrug das gesamte Vorgehen ungefähr 1,5 Schulstunden.

Zu Beginn der Stunde begrüßte die Testleiterin die Kinder und die Lehrerin oder den Lehrer und erklärte den Ablauf der Testung. Den Kindern wurde eine kleine Belohnung für ihre Teilnahme versprochen. Daraufhin wurde die Kurzform des KFT 4-12+ R nach den Anweisungen aus dem Handbuch als Gruppentest vorgegeben. Während der Instruktionen sollten die Kinder still mitlesen, damit ein besseres Verständnis gewährleistet war. Die Vorgabe dauerte etwas über eine Schulstunde. Die LehrerInnen bearbeiteten zeitgleich einen LehrerInnenfragebogen, welcher aus einer LehrerInneneinschätzung und einer LehrerInnennomination bestand (siehe Anhang 7). Nach der Testung beantworteten die anwesenden Kinder einen SchülerInnenfragebogen (bzw. eine Peernomination) (siehe Anhang 9). Das Design der Fragebögen wurde von Lang (2008) zur Durchführung der eigenen Studie erstellt und in der vorliegenden Studie

übernommen. Die Testleiterin erklärte das Vorgehen der Peernomination und erläuterte dabei den Kindern jede Frage einzeln, bevor sie bearbeitet wurde (siehe Anhang 8). Zuletzt wurde die Belohnung an die Kinder und die Lehrkraft verteilt.

3.1.4 Erhebungsinstrumente

Kognitiver Fähigkeitstest

Als Referenzwert für die kognitive Leistungsfähigkeit der Kinder wurden Ergebnisse eines Intelligenztests herangezogen. Es wurde die revidierte Form des Kognitiven Fähigkeitstests für 4. bis 12. Klassen (KFT 4-12+ R von Heller & Perleth, 2000) verwendet, da er sich als Gruppentest eignet und auch im oberen Leistungsbereich noch gut differenziert, einfach in der Durchführung ist und wenig Vorbereitungen bedarf. Der KFT 4-12+ R stellt die deutsche Version des Cognitive Abilities Tests (CAT) von Thorndike und Hagen (1971) dar und stützt sich auf das Berliner Intelligenzstrukturmodell von Jäger (1984). Dieses teilt die Intelligenz in operative und inhaltsgebundene Fähigkeiten ein.

Im Testmanual existieren differenzierte Normen. Für die Auswertung wurde die Normstichprobe der ViertklässlerInnen herangezogen. Die Objektivität ist bei einer Durchführung nach dem Manual, das das erwünschte Vorgehen sehr ausführlich beschreibt, gegeben. Es werden im Manual Werte der inneren Konsistenz für die verwendeten Subtests in der Altersstufe der vierten Klasse zwischen $r = .6$ und $.9$ sowie Werte der Stabilitätsreliabilität nach ein und zwei Jahren von $r = .8$ bis $.9$ angeführt. Vergleiche mit Verbal-Comprehension- und Reasoning-Tests sprechen den verwendeten Subtests eine Konstruktvalidität von $r = .4$ bis $.6$ zu. Reliabilität und Validität sind zufrieden stellend.

Da die Testung möglichst kurz und wenig belastend für die Kinder sein sollte und sich die Fragestellung lediglich auf zwei Teilbereiche der Intelligenz beschränkt, wurde eine Kurzform des Tests, bestehend aus vier Untertests, herangezogen. Die Kinder bekamen Bögen mit Aufgaben und jeweils einen separaten Antwortbogen, auf dem sie ihre Antwort in Form von einem Kreuz pro Item anbringen sollten. Abbildung 2 zeigt das Antwortformat. Auf den Aufgabebögen waren die Items pro Subtest jeweils auf ein bis drei Seiten abgedruckt.

Aufgaben					
	A	B	C	D	E
1	☐	☐	☐	☐	☐
2	☐	☐	☐	☐	☐
3	☐	☐	☐	☐	☐
4	☐	☐	☐	☐	☐
5	☐	☐	☐	☐	☐

Abb. 2: Ausschnitt des Antwortbogens des KFT 4-12+ R

Es werden nun die verwendeten Subtest (Verbaltests und Quantitativtests), die in der Kurzform des Tests bei der Untersuchung zur Anwendung kamen vorgestellt.

- V-Test1: Wortschatz

Es muss zu einem Wort sein Synonym oder Oberbegriff aus einer Liste von fünf Wörtern gesucht werden. Der Untertest besteht aus 25 Items. Hierfür stehen sieben Minuten zur Verfügung.

Itembeispiel:

B1 Rose	A Musik	B Blume	C Speise	D Aussicht	E Lasso
----------------	----------------	----------------	-----------------	-------------------	----------------

Entschied man sich beispielsweise bei diesem Item für Antwort **B Blume**, müsste auf dem Antwortbogen in der Zeile der bearbeiteten Aufgabe ein Kreuz im Kästchen, das in Spalte B abgedruckt ist, gemacht werden.

- V-Test3: Wortanalogien

Ein Wortpaar ist gegeben. Einem dritten Wort muss aus einer Liste von fünf Wörtern jenes zugeordnet werden, das zum dritten Wort im gleichen Verhältnis steht wie das zweite zum ersten. Dieser Teil des Tests beinhaltet 20 Items. Die Kinder haben auch hierfür sieben Minuten Zeit.

Itembeispiel:

B1 groß : riesig → klein :				
A Junge	B erwachsen	C winzig	D wenig	E mehr

- Q-Test1: Mengenvergleiche

Es sind in zwei nebeneinander liegenden Feldern Mengen abgebildet. Die Testpersonen sollen entscheiden, ob die rechte Menge kleiner, größer oder gleich groß als die linke Menge ist. Dieser Subtest umfasst 25 Items. Es sind zehn Minuten dafür angesetzt.

Itembeispiel:

B1	
	A (Wert im linken Feld größer) B (Werte gleich groß) C (Wert im rechten Feld größer)

- Q-Test2: Zahlenreihen

Eine Zahlenreihe aus fünf Ziffern muss um eine weitere, zu wählen aus fünf Möglichkeiten, ergänzt werden. Für die 20 verschiedenen Aufgaben sind neun Minuten angedacht.

Itembeispiel:

B1 10 9 8 7 6	→ A 3	B 4	C 5	D 6	E 7
----------------------	--------------	------------	------------	------------	------------

Mit der Kurzform des KFT 4-12+ R kann das kognitive Fähigkeitsniveau von Kindern ab der vierten Schulstufe gemessen werden. Die V-Tests erfassen dabei sprachliches, die Q-Tests quantitatives Denken. Die Aufgaben sind je Subtest in ihrem Schwierigkeitsgrad ansteigend.

LehrerInneneinschätzung

Die LehrerInneneinschätzung wurde von Lang (2008) auf Grundlage einer Studie von Stöger (2003) erstellt und zum Zwecke der Vergleichbarkeit der Daten auch in vorliegender Arbeit verwendet. Die LehrerInnen wurden anhand einer schriftlichen Information in die Aufgabe eingeführt (siehe Anhang 7) und aufgefordert, jedes anwesende Kind der Klasse auf einer Skala von 1 bis 100 in Bezug auf a) die sprachliche und b) die mathematische Fähigkeit einzuschätzen. In einer Tabelle sollte

links der Name des Kindes eingetragen werden und rechts daneben der zugehörige Prozentrang. Der Begriff „Leistungsfähigkeit“ wurde in der vorliegenden Arbeit durch den Begriff „Fähigkeit“ ersetzt, da sich in einer Vorbefragung von GrundschullehrerInnen aus dem persönlichen Umfeld der Forscherin ergab, dass auf diese Art weniger wahrscheinlich eine Vermischung zwischen dem Konzept der Begabung und dem Konzept der Leistung stattfindet. Es wurde erklärt, was unter sprachlicher und mathematischer Fähigkeit verstanden wird und dass es sich um eine klassenübergreifende Einschätzung handelt. Als Vergleichsgruppe der einzelnen SchülerInnen sollte ein österreichischer Durchschnittsschüler bzw. eine österreichische Durchschnittsschülerin angesehen werden. Die sich ergebenden Daten in Form von Prozenträngen konnten mit den in Prozenträge umgerechneten T-Werten des KFT 4-12+ R verglichen werden.

Nomination

Bei der Nomination sollten die LehrerInnen beziehungsweise die SchülerInnen jeweils ein bis drei MitschülerInnen aufschreiben, die „besonders gut rechnen“, „besonders gut Geschichten erzählen oder schreiben“ oder „sich besonders gut ausdrücken“ konnten (siehe Anhang 9). Es bestand für die SchülerInnen die Möglichkeit, sich selbst zu nennen. Lang (2008) entwickelte für den sprachlichen Bereich zwei verschiedene Fragestellungen, da Unsicherheit bezüglich der Eignung bestand. Sie vermutete, dass die Frage nach dem „Geschichten erzählen“ zwar leichter verständlich für die Kinder sei, jedoch die Gefahr eine Vermischung mit dem Konstrukt der Kreativität bestehe. Die LehrerInnen hatten die Möglichkeit, ihrer Meinung nach besonders herausragende SchülerInnen mit einem Stern zu kennzeichnen.

Sonstige Daten

Neben den demographischen Daten wurden den LehrerInnen zusätzlich folgende Fragen gestellt: „Seit wann unterrichten Sie diese Klasse?“ (Antwortmodus: Seit der 1., 2., 3. oder 4. Klasse), „Seit wie vielen Jahren üben Sie den Beruf der VS-Lehrerin aus?“ (Antwortmodus: Seit ____ Jahr/en), „Haben Sie Fortbildungen im Bereich der Hochbegabung besucht?“ (Antwortmodus: Nein/Ja) und „Wenn ja, im Ausmaß von ca. ____ Fortbildungstagen“, „Haben Sie im Rahmen Ihrer Ausbildung zum Thema Hochbegabung etwas gehört?“ (Antwortmodus: fünfstufige Skala von „gar nichts,“ bis

„sehr viel“), „Haben Sie sich privat näher mit dem Thema Hochbegabung beschäftigt?“ (Antwortmodus: Nein/Ja) und „Wenn ja (in Ausbildung oder privat), geben Sie bitte kurz an, mit was genau und in welchem Rahmen Sie sich mit dem Thema beschäftigt haben.“ (Antwortmodus: offen) (siehe Anhang 7).

Die Kinder wurden im Anschluss an die Nomination nach ihrer Muttersprache gefragt.

3.2 Auswertungsverfahren

Um die Ergebnisse des Intelligenztestverfahrens (Kurzform des KFT 4-12+ R) mit denen der LehrerInneneinschätzung zu vergleichen, wurden die Rohwerte der einzelnen Kinder pro Testteil zuerst in T-Werte transformiert und dann in Prozentränge (PR) umgerechnet. Es ergaben sich zwei verschiedene Werte in Prozenträngen pro Kind: einer für den mathematischen Testteil (QPR) und einen für den sprachlichen Testteil (VPR).

Die Intelligenztestwerte wurden mit den Nominationsmethoden vergleichbar, indem, nach dem Vorbild von Lang (2008), klasseninterne Rangreihungen und eine klassenübergreifende Reihung der Intelligenztestergebnisse vorgenommen wurde: Die Ränge (von 1 bis x ; x = Anzahl der anwesenden Kinder in der Klasse) wurden bei gleichem PR mehrfach vergeben, die Folgeränge wurden dann jeweils übersprungen.

Damit die Einschätzungsmethode mit den Nominationsmethoden verglichen werden konnte, wurde eine Rangreihung der Einschätzungswerte pro Klasse vorgenommen. Auch hier wurden die Ränge (von 1 bis x ; x = Anzahl der anwesenden Kinder in der Klasse) bei gleichem PR mehrfach vergeben, die Folgeränge wurden dann jeweils übersprungen.

Mit den Daten der Peernomination wurde eine Rangreihung pro Klasse erstellt. Die Ränge wurden nach Häufigkeit der Nennungen in der Klasse vergeben. Auch hier konnten mehrere Kinder die gleichen Ränge belegen. Für Vergleiche zwischen den Klassen wurde die Anzahl der Nennungen durch die Anzahl der Kinder in der Klasse dividiert.

Für jedes Kind lagen nun neben den Intelligenztestergebnissen (mathematischer und sprachlicher Bereich) und den LehrerInneneinschätzungen (mathematischer und sprachlicher Bereich) in Prozenträngen auch sechs Rangplätze innerhalb der Klasse, zwei abhängig vom Intelligenztestergebnis (mathematischer und sprachlicher Bereich), zwei abhängig von der LehrerInneneinschätzung (mathematischer und sprachlicher

Bereich) und zwei abhängig von der Peernomination (mathematischer und sprachlicher Bereich) vor. Weitere zwei Rangplätze wurden dem Kind durch seine Intelligenz-testergebnisse (mathematischer und sprachlicher Bereich) im Vergleich zur Gesamtstichprobe zuteil.

Fragestellung: Diagnostische Kompetenz

Die Fragestellung nach der diagnostischen Kompetenz der LehrerInnen wurde einerseits durch einen Vergleich der LehrerInneneinschätzung und den Testergebnissen des KFT 4-12+ R beantwortet. Eine Kreuztabelle über die Prozentränge, unterteilt in fünf Intervalle (orientiert an Ziegler & Stöger, 2003) mit den Bereichen $PR < 15$; $15 \leq PR < 25$; $25 \leq PR < 75$; $75 \leq PR < 85$ und $85 < PR$ wurde berechnet. Andererseits wurden pro LehrerIn drei Komponenten nach Schrader und Helmke (1987) bestimmt, jeweils für den mathematischen und den verbalen Teil getrennt.

Niveaukomponente

$NK = M(\text{Urteile einer Lehrerin}) - M(\text{Testwerte des KFT 4-12+ R einer Klasse})$

0 bedeutet akkurat

Differenzierungskomponente

$DK = SD(\text{Urteile einer Lehrerin}) / SD(\text{Testwerte einer Klasse})$

1 bedeutet akkurat

Vergleichskomponente

$VK = \text{Produktmomentkorrelation zwischen Urteilen der Lehrerin über die einzelnen SchülerInnen und den Testwerten der einzelnen SchülerInnen}$

1 bedeutet akkurat

Zusätzlich wurde zur Abwägung des Zusammenhangs der Einschätzungen und Testergebnisse über die Klassen hinweg eine Spearman-Rangkorrelation gerechnet.

Fragestellung: Identifikation der klassenintern leistungsfähigsten Kinder

Die Fragestellungen zur Identifikation der klassenintern leistungsfähigsten Kinder, wurden zum einen durch die Bestimmung der „Treffer“ beantwortet. Jene Personen, die durch die Einschätz- oder Nominationsverfahren auf die ersten drei Ränge gestellt wurden und auch im Test auf den ersten drei Rängen zu finden waren, wurden dabei als „Treffer“ bezeichnet. Diese Vorgehensweise orientierte sich zwecks Vergleichbarkeit der Daten an Lang (2008). Die mangelhafte Messgenauigkeit des Intelligenztests bringt methodische Mängel mit sich, da mit einem großen Intervall, auf dem der wahre Wert der getesteten Person liegen kann, gerechnet werden muss. Dies führt zu Ungenauigkeiten bezüglich der Bestimmung der drei „besten“ Kinder. Das oben beschriebene Vorgehen stellt somit einen Versuch dar, sich methodisch der Fragestellung zu nähern; Messfehler sind allerdings nicht auszuschließen. Um die geeignetsten Verfahren zur Identifikation zu bestimmen, wurde zum anderen für die verschiedenen Verfahren die Effizienz und Effektivität nach Pegnato und Birch (1959, zitiert nach Wild, 1991) bestimmt

Wie viele der wirklich überdurchschnittlich Leistungsfähigen werden identifiziert?

Die Effektivität indiziert, wie erfolgreich die LehrerInnen und SchülerInnen mit Hilfe der jeweiligen Methode die Personen in der Stichprobe identifizieren können, die einem vorher festgelegten Kriterium entsprechen. Eine Effektivität von 100 % bedeutet demnach, dass alle diese Personen identifiziert werden konnten.

Effektivität = Anzahl richtig nominiertes „(Hoch-)begabter“ / Anzahl „(hoch-)begabter“ Personen

Wie viele der Nominierten sind überdurchschnittlich leistungsfähig?

Die Effizienz bezeichnet, wie ökonomisch die Methode ist. Nimmt man das Beispiel eines Screeningverfahrens, bei dem in einer zweiten Runde eine aufwändigere Intelligenztestung durchgeführt werden soll, ist eine möglichst genaue Vorauswahl an potenziell besonders leistungsfähigen Kindern von Vorteil. Werden in einer Screeningrunde sehr viele Kinder nominiert, ist der Aufwand der zweiten Runde groß. Eine Effizienz von 100 % bedeutet, dass alle Nominierten auch tatsächlich besonders leistungsfähig sind.

Effizienz = Anzahl richtig nommierter „(Hoch-)begabter“ / Anzahl nommierter Personen

Mit der Analyse der ersten drei Ränge der Urteile und der Testergebnisse wurde untersucht, wie effektiv und effizient die Methoden sind, um die SchülerInnen der Klasse zu identifizieren, die die ersten drei Ränge in den Testergebnissen belegten. Dies bedeutet bei folgender Analyse nicht zwangsweise, dass diese Personen überdurchschnittlich leistungsfähig sind, sondern lediglich, dass sie die leistungsfähigsten der Klasse sind.

Wild (1999) weist darauf hin, dass die Bewertung der Höhe der Indizes abhängig von der Fragestellung sein muss. Möchte man z. B. zu Forschungszwecken möglichst viele besonders leistungsfähige Kinder in einer Population identifizieren, sollte der Effektivitätswert gegen 100 % streben. Umso kleiner wird dann jedoch der Effizienzwert, was dazu führt, dass nach einer sehr effektiven Screeningphase solch eine geringe Effizienz gegeben ist, dass fast alle Kinder der Population noch einmal getestet werden müssten. Ein besonders effizientes und damit weniger aufwändiges Vorgehen führt dagegen zu einer geringeren Effektivität, was das Risiko eines Fehlers zweiter Art wachsen lässt und somit viele besonders Leistungsfähige unentdeckt bleiben. Ein Fehler erster Art bezeichnet gemäß Zöfel (2003) in der Wissenschaft die Ablehnung einer Nullhypothese („Das Kind ist nicht besonders begabt“) zugunsten der Alternativhypothese („Das Kind ist besonders begabt“), obwohl die Nullhypothese gilt, während ein Fehler zweiter Art die Annahme der Nullhypothese trotz gültiger Alternativhypothese bezeichnet. Fehler erster und zweiter Art wirken reziprok aufeinander ein. So führt beispielsweise eine Verringerung des Fehlers erster Art zu einer Erhöhung des Fehlers zweiter Art und umgekehrt.

Des Weiteren wurde ein möglicher Informationszuwachs durch die Kombination der verschiedenen Methoden überprüft. Dies wurde ebenfalls anhand der „Treffer“ und der Effektivitäts- sowie Effizienzindizes bewertet.

Fragestellung: Identifikation besonders leistungsfähiger Kinder

Für die Bewertung der verschiedenen Methoden als Screeningverfahren für die Identifikation von hoher Leistungsfähigkeit, bezogen auf die österreichische Gesamtpopulation, wurden ebenfalls „Treffer“ und Effektivitäts- sowie Effizienzindizes herangezogen. Die Kombination der Methoden wurde in derselben Weise bewertet.

Fragestellung Einflussgrößen auf die diagnostische Kompetenz

Die Überprüfung der Moderatorvariablen Unterrichtsexpertise, Unterrichtsdauer und Wissen über Hochbegabung wurde mittels multifaktorieller Varianzanalyse vorgenommen. Die LehrerInnen wurden dazu in jeweils zwei Gruppen geteilt. Als abhängige Variablen dienten die Niveauelementen, die Differenzierungselementen und die Vergleichselementen. Bei der unabhängigen Variable Unterrichtsexpertise wurde unterschieden zwischen BerufsanfängerInnen (unter fünf Jahren Berufserfahrung) und Berufserfahrenen (über neun Jahre Berufserfahrung). Bei der Unterrichtsdauer wurde unterschieden zwischen LehrerInnen, die ihr erstes Jahr mit der Klasse verbrachten und LehrerInnen, die länger als ein Jahr in der Klasse unterrichteten. Beim Wissen über Hochbegabung wurde eingeteilt in kein bis kaum Wissen über Hochbegabung (keine Fortbildung über Hochbegabung besucht) und Wissen über Hochbegabung (Fortbildung über Hochbegabung besucht) vorhanden.

3.3 Ergebnisse

3.3.1 Deskriptive Ergebnisse der Stichprobe

Die Stichprobe setzte sich aus 108 Mädchen und 111 Jungen zusammen. Eine Gleichverteilung der Geschlechter war gegeben. Die übrigen statistischen Kennwerte der Stichprobe sind Tabelle 1 zu entnehmen.

Tab. 1: Statistische Kennwerte der Stichprobe. N = 219

Variable	Mittelwert	Standardabweichung
w/m	108/111	
QPR	54,71	30,81
VPR	50,34	28,17
LEM	64,39	20,62
LES	60,81	20,60

Legende: QPR = Quantitativtestergebnisse, VPR = Verbaltestergebnisse, LEM = mathematische LehrerInneneinschätzung, LES = sprachliche LehrerInneneinschätzung.

Die Mittelwerte der Intelligenztestwerte lagen für den mathematischen als auch den sprachlichen Bereich im Durchschnittsbereich. Die dazugehörigen Standardabweichungen waren im Vergleich zur Normstichprobe eher groß. Es muss allerdings berücksichtigt werden, dass hier nur mit Normen der fünften Klasse verglichen werden konnte, da für die Normen der vierten Klassen im Manual des KFT 4-12+ R keine Standardabweichungen angegeben sind. Die Mittelwerte der mathematischen und sprachlichen LehrerInneneinschätzung lagen im Durchschnittsbereich. Die Standardabweichungen waren kleiner als die der Intelligenztestwerte und gering kleiner als die bei Lang (2008).

Die Prüfung auf Normalverteilung der Testkennwerte und der LehrerInneneinschätzung wurde mit dem Kolmogorov-Smirnov Test durchgeführt (Signifikanzniveau $p \geq .01$). Es waren Normalverteilungen bei den Einschätzungsvariablen sowie bei den Testkennwerten des sprachlichen Testteils gegeben. Die Ergebnisse des mathematischen Testteils waren nicht normalverteilt (siehe Anhang 2).

Die statistischen Kennwerte der einzelnen Klassen sind im Anhang zu finden (siehe Anhang 3). Berichtet werden Stichprobengröße, Mittelwert der mathematischen LehrerInneneinschätzung (LEM), der sprachlichen LehrerInneneinschätzung (LES), von den Quantitativtestergebnissen, bzw. den mathematischen Intelligenztestergebnissen (QPR) und den Verbaltestergebnissen, bzw. den sprachlichen Intelligenztestergebnissen (VPR) sowie dazugehörige Standardabweichungen. Außerdem werden die Minimal- und Maximalwerte pro Klasse angegeben. Es befanden sich alle Intelligenztestmittelwerte der Klassen im Durchschnittsbereich. Die Einschätzungen der LehrerInnen bewegten

sich ebenfalls alle im Durchschnittsbereich. Zwischen den Schulen lagen teilweise erhebliche Unterschiede vor. Im mathematischen Bereich betrug der niedrigste Durchschnittswert einer Klasse $PR = 32$, während der Maximalwert bei $PR = 74$ lag. Der Minimalwert der durchschnittlichen Verbaltestergebnisse einer Klasse lag bei $PR = 32$, während der höchste bei $PR = 77$ lag. Die Einschätzungswerte schwankten weniger stark. Hier zeigte sich im mathematischen Bereich ein niedrigster Durchschnittswert von $PR = 54$ und ein höchster von $PR = 75$. Im sprachlichen Bereich fand sich der Minimalwert bei $PR = 52$ und der Maximalwert bei $PR = 73$.

3.3.2 Ergebnisse zur diagnostischen Kompetenz

Wie ist die diagnostische Kompetenz der LehrerInnen zu bewerten?

Das folgende Kapitel soll darstellen, wie gut LehrerInnen in der Lage sind, a) mathematische Leistungsfähigkeit und b) sprachliche Leistungsfähigkeit ihrer SchülerInnen einzuschätzen. Dies wurde u. a. anhand von Kreuztabellen berechnet. Die Einschätzungen und die Testergebnisse wurden dazu in fünf Intervalle unterteilt (orientiert an Ziegler & Stöger, 2003).

Die Tabellen geben einen Überblick über die Verteilung der LehrerInneneinschätzung der SchülerInnen im mathematischen bzw. sprachlichen Bereich und der Testergebnisse der SchülerInnen im mathematischen bzw. sprachlichen Teil des KFT 4-12+ R sowie deren Übereinstimmungen.

a) Mathematischer Bereich

Tab. 2: Kreuztabelle der LehrerInneneinschätzung im mathematischen Bereich und der Testergebnisse im mathematischen Bereich in Prozenträngen, gestaffelt in fünf Intervalle. $N = 219$

LE der mathematischen Leistungsfähigkeit		PR < 15	15 ≤ PR < 25	25 ≤ PR ≤ 75	75 < PR ≤ 85	85 > PR	Total
QPR	PR < 15	1	4	24	1		30
	15 ≤ PR < 25		0	24		1	25
	25 ≤ PR ≤ 75			62	5	21	88
	75 < PR ≤ 85			12	2	9	23
	PR > 85		1	26	5	21	53
Total		1	5	148	13	52	219

Legende: LE = LehrerInneneinschätzung, PR = Prozentrang, QPR = Quantitativtestergebnisse.

Betrachtet man die Prozentränge der Einschätzungen und der Testergebnisse in fünf Intervallen ($PR < 15$; $15 \leq PR < 25$; $25 \leq PR < 75$; $75 \leq PR < 85$; $85 < PR$), zeigt sich im mathematischen Bereich eine Übereinstimmung der Werte in 86 von 219 Fällen (39,3 %). In Tabelle 2 ist dies abzulesen als die Summe aller fettgedruckten Werte. Eine Unterschätzung trat in 49 Fällen (22,4 %) auf. Die Tabelle veranschaulicht dies als Summe aller Werte, die oberhalb der fettgedruckten Diagonale der übereinstimmenden Werte liegen. Die Summe der Überschätzungen beträgt 84 von 219 (38,4 %). Diese befinden sich in der Tabelle unterhalb der fettgedruckten Diagonale.

b) Sprachlicher Bereich

Tab. 3: Kreuztabelle der LehrerInneneinschätzung im sprachlichen Bereich und der Testergebnisse im sprachlichen Bereich in Prozenträngen, gestaffelt in fünf Intervalle. N= 219

LE der verbalen Leistungsfähigkeit		PR < 10	10 ≤ PR < 25	25 ≤ PR < 75	75 ≤ PR < 90	90 < PR	Total
VPR	PR < 10	0	3	18			21
	10 ≤ PR < 25		1	28	4	1	34
	25 ≤ PR < 75		2	80	19	9	110
	75 ≤ PR < 90		1	20	6	7	34
	90 < PR			13	3	4	20
Total		0	7	159	32	21	219

Legende: LE = LehrerInneneinschätzung, PR = Prozentrang, VPR = Verbaltestergebnisse.

Betrachtet man die Prozentränge der Einschätzungen und der Testergebnisse in fünf Intervallen, zeigte sich im sprachlichen Bereich eine Übereinstimmung der Werte in 91 Fällen (41,6 %), eine Unterschätzung in 49 Fällen (22,4 %) und eine Überschätzung in 89 Fällen (40,6 %). Die Werte sind aus Tabelle 3 ebenso wie im mathematischen Bereich (s. o.) abzulesen.

Zusammenfassend ist in Bezug auf die diagnostische Kompetenz der LehrerInnen festzustellen, dass der Großteil der LehrerInnen die Leistungsfähigkeit der Kinder entweder über- oder unterschätzte. Lediglich c. a. 40 % der Kinder wurden korrekt dem Durchschnittsintervall zwischen Prozentrang 25 und 75 zugeordnet.

Wie genau schätzten die LehrerInnen das Klassenniveau ein?

Die Bestimmung drei verschiedener Komponenten nach Schrader und Helmke (1987) soll ein differenziertes Bild der diagnostischen Kompetenz der LehrerInnen liefern. Die

Genauigkeit der Einschätzungen der Klassenlehrer und Klassenlehrerinnen aller Klassen wird dargestellt.

Die Voraussetzung einer Normalverteilung der QPR, VPR, LEM und LES pro Klasse sind nach dem Kolmogorov-Smirnov Test gegeben (Signifikanzniveau $p \geq .01$) (siehe Anhang 2).

a) Mathematischer Bereich

Tab. 4: Mittelwerte der LehrerInneneinschätzung (Mathe), Mittelwerte der Testergebnisse (Mathe) und Niveauelemente pro Lehrkraft bzw. Klasse (N bis X). N= 219

LK	N	O	P	Q	R	S	T	U	V	W	X
M _{LEM}	70,71	60,75	64,25	59,30	68,61	54,29	60,86	62,25	74,76	74,50	55,25
M _Q	65,88	63,43	72,09	52,15	47,74	41,64	44,91	73,88	60,86	42,56	32,30
NK	4,83	-2,68	-16,75	7,15	20,87	12,65	15,95	-11,63	13,90	31,94	22,95
Ü in %	7			14	44	30	36		23	75	71
U in %		4	23					16			
M _{NK} : 9,02											

Legende: LK = Lehrkraft (kodiert mit N bis X), M= Mittelwert, LEM = LehrerInneneinschätzung der mathematischen Fähigkeit, Q = Quantitativer Testergebnisse, NK = Niveauelemente, Ü = Überschätzung, U = Unterschätzung, Markierung = Extremwerte.

Die Differenz der LehrerInneneinschätzung der mathematischen oder sprachlichen Fähigkeiten und des Klassendurchschnitts der jeweiligen Testteilergebnisse stellt die Niveauelemente der einzelnen LehrerInnen dar. Eine Niveauelemente mit dem Wert null ist dabei die genauest mögliche Einschätzung.

Die Werte schwankten im mathematischen Bereich zwischen einer sehr niedrigen Niveauelemente von -16,75 (siehe Tabelle 4, Zeile NK, Spalte P), wobei es sich um eine Unterschätzung des Niveaus handelte und 31,94 (siehe Tabelle 4, Zeile NK, Spalte W), wobei hier eine starke Überschätzung vorlag. Der Durchschnitt aller Niveauelemente betrug 9,02, was in Tabelle 4 in der letzten Zeile dargestellt ist.

b) Sprachlicher Bereich

Tab. 5: Mittelwerte der LehrerInneneinschätzung (Sprache), Mittelwerte der Testergebnisse (Sprache) und Niveauelemente pro Lehrkraft bzw. Klasse (N bis X). N= 219

LK	N	O	P	Q	R	S	T	U	V	W	X
M _{LES}	56,90	55,15	58,19	55,35	67,26	55,00	63,45	57,17	72,76	74,11	51,75
M _V	49,57	63,50	69,72	47,76	34,27	31,93	38,10	76,69	65,67	34,25	36,68
NK	7,33	-8,35	-11,53	7,59	32,99	23,07	25,35	-19,52	7,09	39,86	15,07
Ü in %	15			16	96	72	67		11	116	41
U in %		13	17					25			
M _{NK} : 10,81											

Legende: LK = Lehrkraft (kodiert mit N bis X), M= Mittelwert, LES = LehrerInneneinschätzung der sprachlichen Fähigkeit, V = Verbaltestergebnisse, NK = Niveauelemente, Ü = Überschätzung, U = Unterschätzung, Markierung = Extremwerte.

Im sprachlichen Bereich lag die niedrigste Niveauelemente bei -19,52 (Tabelle 5, Zeile NK, Spalte U) und die höchste bei 39,86 (Tabelle 5, Zeile NK, Spalte W). Hier betrug der Durchschnitt 10,81 (siehe letzte Zeile der Tabelle). Insgesamt wurde das Niveau der Klasse in acht von elf Fällen (72,7 %) überschätzt, wie die Tabelle in der Zeile „Ü in %“ darstellt. Die Einschätzungen der LehrerInnen waren von der Tendenz über die Fächer hinweg konsistent. Alle Lehrkräfte, die in einem Bereich überschätzten bzw. unterschätzten, überschätzten bzw. unterschätzten ebenfalls im jeweils anderen Bereich.

Insgesamt war bei der Genauigkeit der LehrerInneneinschätzungen des Klassenniveaus in den meisten Fällen eine Überschätzung des Leistungsniveaus im sprachlichen sowie im mathematischen Bereich zu beobachten.

Wie genau schätzten die LehrerInnen die Streuung der Leistungsfähigkeit ein?

a) Mathematischer Bereich

Tab. 6: Standardabweichungen der LehrerInneneinschätzung (Mathe), Standardabweichungen der Testergebnisse (Mathe) und Differenzierungskomponente pro Lehrkraft bzw. Klasse (N bis X). N= 219

LK	N	O	P	Q	R	S	T	U	V	W	X
SD _{LEM}	18,66	13,60	16,85	18,43	22,61	30,10	19,60	17,92	22,61	18,58	19,70
SD _Q	26,83	29,90	26,87	34,02	30,24	31,32	31,10	19,02	30,91	22,51	29,00
DK	0,70	0,45	0,63	0,54	0,75	0,96	0,63	0,94	0,73	0,83	0,68
	U	U	U	U	U	1	U	1	U	U	U
M _{DK} : 0,71 (U)											

Legende: LK = Lehrkraft (kodiert mit N bis X), SD = Standardabweichung, LEM = LehrerInneneinschätzung der mathematischen Fähigkeit, Q = Quantitativer Testergebnisse, DK = Differenzierungskomponente, Ü = Überschätzung, U = Unterschätzung, Markierung = Extremwerte.

Der Quotient der Standardabweichung der LehrerInneneinschätzung der mathematischen oder sprachlichen Fähigkeiten und der Standardabweichung des Klassendurchschnitts der jeweiligen Testteilergebnisse ergibt die Differenzierungskomponente der einzelnen KlassenlehrerInnen. Eine Differenzierungskomponente von 1 ist dabei die genauest mögliche Einschätzung. Zwischen 0 und 0,91 wird von einer Unterschätzung der Streuung der Fähigkeiten gesprochen, Werte über 1,09 stellen eine Überschätzung dar.

Die Werte schwankten im mathematischen Bereich zwischen 0,45 und 0,94, was in Tabelle 6 in der Zeile DK dargestellt ist. Der Durchschnitt von 0,71 (letzte Zeile der Tabelle 6) bezeichnet eine Unterschätzung der Streuung durch die LehrerInnen. Dabei ist anzumerken, dass die Streuung im mathematischen Bereich von neun der elf LehrerInnen (81,8 %) unterschätzt wurde, die restlichen LehrerInnen schätzten die Streuung genau ein.

b) Sprachlicher Bereich

Tab. 7: Standardabweichungen der LehrerInneneinschätzung (Sprache), Standardabweichungen der Testergebnisse (Sprache) und Differenzierungskomponente pro Lehrkraft bzw. Klasse (N bis X). N= 219

LK	N	O	P	Q	R	S	T	U	V	W	X
SD _{LES}	20,34	12,61	15,92	16,35	22,88	26,75	19,53	19,67	20,44	17,95	22,20
SD _V	27,64	24,36	18,81	24,44	26,40	23,35	31,10	15,42	21,45	21,09	24,27
DK	0,74	0,52	0,85	0,67	0,87	1,15	0,63	1,28	0,95	0,85	0,91
	U	U	U	U	U	Ü	U	Ü	1	U	U

M_{DK}: 0,86 (U)

Legende: LK = Lehrkraft (kodiert mit N bis X), SD = Standardabweichung, LES = LehrerInneneinschätzung der sprachlichen Fähigkeit, V = Verbaltestergebnisse, DK = Differenzierungskomponente, Ü = Überschätzung, U = Unterschätzung, Markierung = Extremwerte.

Im sprachlichen Bereich lagen die Werte der Differenzierungskomponente zwischen 0,52 und 1,28 (Tabelle 7, Zeile DK). Der Durchschnitt von 0,86 (letzte Zeile der Tabelle) weist ebenfalls auf eine tendenzielle Unterschätzung der Streuung der Leistungsfähigkeit hin. Die Streuung wurde von acht der elf Lehrkräfte (72,7 %) unterschätzt, zwei LehrerInnen überschätzen die Streuung und eine Lehrkraft schätzte sie richtig ein. Zusammenfassend ist zu sagen, dass der Großteil der LehrerInnen im mathematischen und im sprachlichen Bereich die Streuung der Leistungsfähigkeit innerhalb der Klassen unterschätzte.

Wie genau schätzen die LehrerInnen die Rangreihung der Leistungsfähigkeit ein?

a) Mathematischer Bereich

Tab. 8: Vergleichskomponenten des mathematischen Bereichs pro Lehrkraft. N = 219

LK	N	O	P	Q	R	S	T	U	V	W	X
VK _{Q & LEM}	.65**	.67**	.40	.63**	.74**	.58*	.71**	.34	.49*	.51*	.84**

Legende: LK = Lehrkraft (kodiert mit N bis X), VK = Vergleichskomponente, LEM = LehrerInneneinschätzung der mathematischen, Q = Quantitativtestergebnisse.

** .01 Niveau signifikant (zweiseitig).

* .05 Niveau signifikant (zweiseitig).

Die jeweilige Produkt-Moment-Korrelation zwischen den LehrerInneneinschätzungen der mathematischen oder sprachlichen Fähigkeiten und den Testteilergebnissen wird als Vergleichskomponente bezeichnet. Je höher die Korrelation, desto genauer ist die Einschätzung der Rangreihung der Fähigkeiten der Kinder in der jeweiligen Klasse durch die KlassenlehrerIn. Diese Korrelationen pro Lehrkraft werden in Tabelle 8 und 9 gezeigt.

Eine Effektstärke unter $r = .3$ gilt nach Bortz und Döring (1995) als klein, zwischen $r = .3$ und $r = .5$ als mäßig und größer als $r = .5$ als groß.

Es konnten im mathematischen Bereich bei acht der elf KlassenlehrerInnen (72,7 %) große Effektstärken, d. h. genaue Einschätzungen der Rangreihung festgestellt werden. Neun Korrelationen waren signifikant. Die einzelnen Korrelationen sind in Tabelle 8 aufgeführt. Der schwächste Zusammenhang lag bei $r = .34$ in Klasse U. Den stärksten Zusammenhang fand man mit $r = .84$ in Klasse X. Der Durchschnitt lag bei $r = .6$. Es zeigte sich ein mittlerer klassenübergreifender Zusammenhang bei der Berechnung einer Spearman-Rangkorrelation für den mathematischen Bereich ($r = .43^{**}$).

b) Sprachlicher Bereich

Tab. 9: Vergleichskomponenten des sprachlichen Bereichs pro Lehrkraft. N = 219

LK	N	O	P	Q	R	S	T	U	V	W	X
VK _{V & LES}	.67**	.08	.80**	.37	.67**	.62*	.74**	.53**	.62**	.68**	.75**

Legende: LK = Lehrkraft (kodiert mit N bis X), VK = Vergleichskomponente, LES = LehrerInneneinschätzung der sprachlichen Fähigkeit, V = Verbaltestergebnisse.

** .01 Niveau signifikant (zweiseitig).

* .05 Niveau signifikant (zweiseitig).

Im sprachlichen Bereich wurden große Effektstärken bei neun KlassenlehrerInnen festgestellt, wie in Tabelle 9 angeführt ist. Diese neun Korrelationen fielen alle signifikant

aus. Die Einschätzung der Lehrkraft der Klasse O wies einen sehr niedrigen Zusammenhang mit den Ergebnissen der Kinder in der Intelligenztestung auf. Der höchste Zusammenhang fand sich in Klasse P mit $r = .8$. Der Durchschnitt aller Vergleichskomponenten betrug für den sprachlichen Bereich $r = .59$.

Auch für den sprachlichen Bereich wurde eine Spearman-Rangkorrelation ($r = .54^{**}$) berechnet. Es zeigte sich ein mittlerer klassenübergreifender Zusammenhang.

Über die Klassen und die Bereiche hinweg ergab sich für den Großteil der Lehrkräfte ein mittlerer bis hoher Zusammenhang zwischen den Einschätzungen und den Testergebnissen der Kinder. Das bedeutet, dass die meisten LehrerInnen die Rangreihung der Leistungsfähigkeit innerhalb ihrer Klasse gut einschätzten.

3.3.4 Ergebnisse zur Identifikation der klassenintern leistungsfähigsten Kinder

In diesem Kapitel soll erörtert werden, wie gut LehrerInnen und SchülerInnen unter Verwendung der Einschätzungsmethode und der Nominationsmethode in der Lage sind, die SchülerInnen der Klasse zu bestimmen, die die ersten drei Ränge innerhalb der Klassen bei den Leistungstests erreicht haben.

Deskriptive Ergebnisse der verschiedenen Methoden

Bei der Einschätzungsmethode bewegten sich die durchschnittlichen Urteile der LehrerInnen pro Klasse zwischen $PR = 54,29$ und $PR = 74,76$ im mathematischen Bereich. Die niedrigste Standardabweichung betrug dabei $SD = 13,60$, die höchste $SD = 30,06$. Im sprachlichen Bereich war der niedrigste Durchschnittswert $PR = 51,75$ und der höchste $PR = 74,11$. Die durchschnittlichen Standardabweichungen befanden sich zwischen $SD = 12,61$ und $SD = 22,61$.

Bei allen drei Fragen der LehrerInnennominationsmethode wurden jeweils 15,1 % der Kinder durch die LehrerInnen nominiert. Frage 1 (Mathe: „Welche drei Kinder aus Ihrer Klasse, die an dieser Untersuchung teilnehmen, können am besten rechnen?“) betrifft die mathematische Leistungsfähigkeit, während es bei den Fragen 2 (Erzählen: „Welche drei Kinder aus Ihrer Klasse, die an dieser Untersuchung teilnehmen, können am besten interessante Geschichten erzählen/schreiben?“) und 3 (Ausdruck: „Welche drei Kinder

aus Ihrer Klasse, die an dieser Untersuchung teilnehmen, können sich am besten ausdrücken (Wortschatz)?“ um die sprachliche Leistungsfähigkeit geht.

Die Ergebnisse der Peernomination gestalteten sich folgendermaßen: Bei Frage 1 (Mathe) wurden die Kinder durchschnittlich von 13,2 % ihrer MitschülerInnen genannt. Insgesamt wurden 50,7 % der Kinder von mindestens einem Mitschüler oder einer Mitschülerin nominiert. Bei Frage 2 (Erzählen) wurden die SchülerInnen im Durchschnitt von 12 % der MitschülerInnen genannt. Es wurden 61,6 % der Kinder von mindestens einem Mitschüler oder einer Mitschülerin nominiert. Bei Frage 3 (Ausdruck) wurden die Kinder durchschnittlich von 12,1 % ihrer KlassenkameradInnen genannt. Insgesamt wurden 69,4 % aller Kinder bei Frage 3 von ihren MitschülerInnen nominiert.

Eine weitere Möglichkeit der Identifizierung besonders leistungsfähiger Kinder ist die Selbstnomination. Kinder schätzen bei dieser Methode selbst ein, ob sie zu der Gruppe der besonders Leistungsfähigen gehören. In der durchgeführten Untersuchung bestand im Rahmen der Peernomination die Möglichkeit, sich selbst zu nominieren. Die Selbstnomination ist in vorliegender Studie nicht als eigene Methode anzusehen, sondern wird als Zusatzinformation aus der Peernomination gewonnen. Bei Frage 1 (Mathe) nominierten sich 27,4 % der Kinder selbst. Bei Frage 2 (Erklären) nominierten sich 35,2 % und bei Frage 3 (Ausdruck) nominierten sich 32,4 % der Kinder selbst.

Beitrag der verschiedenen Methoden zur Identifikation der klassenintern leistungsfähigsten Kinder

Basierend auf den Formeln von Pegnato und Birch (1959, zitiert nach Wild, 1991) wurde für jede Methode die Effektivität und die Effizienz berechnet.

Die Tabellen zeigen die verschiedenen Methoden im Vergleich. In der zweiten Spalte ist aufgezeigt, wie viele Personen die ersten drei Ränge bei den Testwerten innerhalb der Klassen belegen. Ihre Summe wird als Grundquote bezeichnet. Die übrigen Spalten zeigen die Anzahl der durch die verschiedenen Methoden nominierten Personen (Selektionsquote), sowie die Trefferzahl. Letztere meint die Zahl der Personen, die durch die jeweilige Methode nominiert wurden und gleichzeitig die ersten drei Ränge bei den Testwerten innerhalb der Klassen belegen. Effektivitäts- und Effizienzwerte sind jeweils in den letzten zwei Zeilen der Tabelle abzulesen.

a) Mathematischer Bereich

Grundquote: Insgesamt fanden sich 35 Kinder auf den ersten drei Rängen des KFT 4-12+ R, klassenweise gereiht nach den Prozenträgen des mathematischen Testteils, wieder.

Die meisten Treffer wurden durch die LehrerInneneinschätzung erreicht (21), wie in Tabelle 10 in Zeile „Treffer“, Spalte LEM zu sehen ist. Dies führte zu einer maximalen Effektivität von 60 % durch die Einschätzungsmethode im mathematischen Bereich. Das bedeutet, dass 60 % der nach den Testwerten auf den ersten drei Rängen liegenden SchülerInnen durch die Einschätzung erkannt wurden. Durch eine verhältnismäßig hohe Selektionsquote (49), ist allerdings die Effizienz nicht wesentlich höher als bei der LehrerInnennomination (43 %).

Tab. 10: Verteilungen der Rangplätze, Effektivität und Effizienz der Methoden im mathematischen Bereich. N= 219

	QPR		LEM	LN-F1	SN F1	PN-F1 mit SN
Rang 1	14		12			12
Rang 2	10		28			10
Rang 3	11		9			15
Grundquote	35	Selektions quote	49	33	60	37
Stichprobe in %	16,0		22,4	15,1	27,4	16,9
		Treffer	21	13	16	12
		Effektivität	60,0 %	37,1 %	45,7 %	34,3 %
		Effizienz	42,9 %	39,4 %	26,7 %	32,4 %

Legende: QPR = Quantitativtestergebnisse, LEM = mathematische LehrerInneneinschätzung, LN = LehrerInnennomination, PN = Peernomination; SN = Selbstnomination; F1 = Frage 1.

Die LehrerInneneinschätzungsmethode eignete sich im mathematischen Bereich am besten, um die klassenintern leistungsfähigsten Kinder zu identifizieren. Es wurde über die Hälfte der im Test leistungsfähigsten Kinder erkannt. Mit den übrigen Methoden konnte nur weniger als die Hälfte erkannt werden.

b) Sprachlicher Bereich

Grundquote: Insgesamt befanden sich 36 Kinder auf den ersten drei Rängen, klassenweise gerangreicht nach den Prozenträgen des sprachlichen Testteils.

Die meisten Treffer (24) wurden durch die Selbstnomination – Frage 2, berechnet aus der Peernomination, erreicht. Tabelle 11 zeigt dies in Zeile „Treffer“, Spalte „SN-F2“. Es

ergab sich eine hohe Effektivität bei einer eher geringen Effizienz, wie man in den letzten zwei Zeilen der Tabelle sehen kann.

Die Peernomination –Frage 3 (Ausdruck) mit Einbezug der Selbstnomination erreichte dagegen bei einer Trefferzahl von 20 eine ausgeglichene Effektivität (56 %) und Effizienz (54 %) (siehe Tabelle 11 vorletzte und letzte Zeile, Spalte „PN-F3 mit SN“). Frage 3 zeigte bei der Peernomination einen Vorteil gegenüber der Frage 2 während bei der Selbstnomination Frage 2 im Vorteil war.

Tab. 11: Verteilungen der Rangplätze, Effektivität und Effizienz der Methoden im sprachlichen Bereich. N= 219

	VPR		LES	LN-F2	LN-F3	PN-F2 mit SN	PN-F3 mit SN	SN-F2	SN-F3
Rang 1	11		19			13	15		
Rang 2	16		8			15	13		
Rang 3	9		13			9	9		
Grundquote	36	Selektions quote	40	33	33	37	37	77	71
Stichprobe in %	16,4		18,3	15,1	15,1	16,9	16,9	35,5	32,4
		Treffer	15	13	16	19	20	24	13
		Effektivität	41,7 %	36,1 %	44,4 %	52,3 %	55,6 %	66,7 %	36,1 %
		Effizienz	37,5 %	39,4 %	48,5 %	51,1 %	54,1 %	31,2 %	18,3 %

Legende: VPR = Verbaltestergebnisse, LES = sprachliche LehrerInneneinschätzung, LN = LehrerInnennomination, PN = Peernomination, F 2 = Frage 2 (Erzählen), F 3 = Frage 3 (Ausdruck), SN = Selbstnomination.

Im sprachlichen Bereich stellte sich die Peernomination (Frage 3) als geeignetste Methode, um die klassenintern leistungsfähigsten Kinder zu identifizieren. Es wurde über die Hälfte der im Test leistungsfähigsten Kinder erkannt. Auch mit Hilfe von Frage 2 der Peernomination konnten über die Hälfte identifiziert werden. Die Effektivität fiel hier allerdings geringer aus als bei Frage 3. Die Selbstnomination (Frage 2) lieferte insgesamt die höchste Effektivitätsrate. Die Effizienzrate war in diesem Fall aber viel geringer, was die Anwendung dieses Verfahrens unattraktiver macht. Mit den übrigen Methoden konnten nur weniger als die Hälfte der klassenintern Leistungsfähigsten erkannt werden.

Kann durch Kombination der Methoden eine erfolgreichere Identifikation der klassenintern leistungsfähigsten Kinder erreicht werden?

Im Folgenden soll überprüft werden, ob eine Kombination verschiedener Methoden eine höhere Trefferanzahl der klassenbesten SchülerInnen mit sich bringt.

a) Mathematischer Bereich

Tab. 12: Effektivität und Effizienz der verschiedenen Methoden in Kombination, bezogen auf den mathematischen Bereich. N = 219

QPR			LEM & LN-F1	PN & LN-F1	PN-F1 & LEM	SN-F1 & LEM	LN & PN-F1 & LEM
Grundquote	35	Selektionsquote	54	50	63	83	66
Stichprobe in %	16,0		24,7	22,8	28,8	37,9	30,1
		Treffer	23	20	23	25	23
		Effektivität	65,7 %	57,1 %	65,7 %	71,4 %	65,7 %
		Effizienz	42,6 %	40,0 %	36,5 %	30,1 %	34,9 %

Legende: QPR = Quantitativtestergebnisse, LEM = mathematische LehrerInneneinschätzung, LN = LehrerInnennomination, PN = Peernomination, F 1 = Frage 1 (Mathe), SN = Selbstnomination.

Im mathematischen Bereich führte eine Kombination von LehrerInneneinschätzung mit Peernomination oder mit LehrerInnennomination zu einer Erhöhung der Effektivität. Es konnten zusätzlich zwei Kinder mehr korrekterweise auf die ersten drei Ränge zugeordnet werden. Dies wird deutlich im Vergleich der Trefferanzahl (21) in Tabelle 12, Zeile „Treffer“ Spalte „LEM“ und der Trefferanzahl (23) in Tabelle 12, Zeile „Treffer“, Spalte „PN-F1 & LEM“ oder „LN-F1 & LEM“. Außerdem erhöhte die Kombination von LehrerInneneinschätzung und Selbstnomination die Trefferzahl. Wie in Zeile „SN-F1 & LEM“ zu sehen ist, konnten vier Kinder mehr, insgesamt 25, korrekt den ersten drei Rängen zugeordnet werden.

In Anbetracht der Komplexität der Berechnungen wurde auf eine Kombination mit mehr als drei Methoden verzichtet.

b) Sprachlicher Bereich

Tab. 13: Effektivität und Effizienz der verschiedenen Methoden in Kombination, bezogen auf den sprachlichen Bereich, Frage 2. N = 219

	VPR		LES & LN-F2	PN & LN-F2	PN-F2 & LES	SN-F2 & LES	SN-F2 & LN F2	LN & PN-F2 & LES
Grundquote	36	Selektionsquote	47	54	56	93	87	63
Stichprobe in %	16,4		21,2	24,7	25,6	42,5	39,7	28,8
		Treffer	17	19	18	28	27	19
		Effektivität	47,2 %	52,8 %	50,0 %	77,8 %	75,0 %	52,8 %
		Effizienz	36,2 %	35,2 %	32,1 %	30,1 %	31,0 %	30,2 %

Legende: VPR = Verbaltestergebnisse, LES = sprachliche LehrerInneneinschätzung, LN = LehrerInnennomination, PN = Peernomination, F 2 = Frage 2 (Erzählen).

Tab. 14: Effektivität und Effizienz der verschiedenen Methoden in Kombination, bezogen auf den sprachlichen Bereich, Frage 3. N = 219

	VPR		LES & LN-F3	PN & LN-F3	PN-F3 & LES	SN-F2 & LN F3	LN & PN-F3 & LES
Grundquote	36	Selektionsquote	48	53	58	87	63
Stichprobe in %	16,4		21,9	24,2	26,5	39,7	28,8
		Treffer	17	22	23	29	23
		Effektivität	47,2 %	61,1 %	63,9 %	80,6 %	63,9 %
		Effizienz	35,4 %	41,5 %	39,7 %	33,3 %	36,5 %

Legende: VPR = Verbaltestergebnisse, LES = sprachliche LehrerInneneinschätzung, LN = LehrerInnennomination, PN = Peernomination, F 3 = Frage 3 (Ausdruck).

Aufgrund der geringen Trefferzahl durch die Selbstnomination – Frage 3 wurde auf eine Kombination mit dieser verzichtet.

Im sprachlichen Bereich führte eine Kombination der Selbstnomination – Frage 2 und der LehrerInnennomination – Frage 3 zur höchsten Effektivität (siehe Tabelle 14, Zeile „Effektivität“, Spalte „SN-F2 & LN-F3“). Es konnten zusätzlich fünf Kinder korrekt als besonders leistungsfähig erkannt werden (vergleiche Tab. 11. Zeile „Treffer“ Spalte „SN-F2“ und Tab. 14, Zeile „Treffer“, Spalte „SN-F2 & LN-F3“). Die Effizienz fiel allerdings in diesem Fall gering aus.

Die LehrerInneneinschätzung in Kombination mit der Peernomination - Frage 3 (Ausdruck) führte zu einer ausgeglicheneren Effektivität (64 %) und Effizienz (40 %). Die Werte der Kombination Peernomination – Frage 3 und LehrerInneneinschätzung sind in Tabelle 14 aufgeführt in der Spalte „PN-F3 & LES“. Bei dieser Methodenkombination konnten zusätzlich drei weitere Kinder korrekt auf die ersten drei Ränge zugeordnet

werden, wie sich im Vergleich mit der Trefferanzahl (20) in Tabelle 11, Zeile „Treffer“ Spalte „PN-F3 mit SN“ zeigt.

Eine Kombination der verschiedenen Methoden führte sowohl im mathematischen als auch im sprachlichen Bereich zu einer Erhöhung der Trefferzahl der klassenintern leistungsfähigsten SchülerInnen.

3.3.5 Ergebnisse zur Identifikation der besonders leistungsfähigen Kinder

Es wurde nun die Teilstichprobe der Kinder untersucht, die im Test einen sehr hohen Prozentrang erreichte und somit als besonders leistungsfähig bezeichnet werden kann. Ziel ist es, die Methode oder Methodenkombination zu eruieren, mit welcher diese Kinder am besten identifiziert werden können.

Die Tauglichkeit der Methoden zum Screening wird untersucht, indem die Effektivität auf 100 % gesetzt und die daraus resultierende Effizienz begutachtet wird. Somit ist gewährleistet, dass alle Kinder, die nach der festgelegten Grenze von $PR > 85$ im KFT 4-12+ R als besonders leistungsfähig bezeichnet werden, erkannt werden.

a) Mathematischer Bereich

Grundquote: Es erreichten 53 Kinder einen Prozentrang > 85 im mathematischen Teil des KFT 4-12+ R.

Tabelle 15 stellt die Effektivität und Effizienz verschiedener Herangehensweisen als Screeningmethoden dar. Hier sollen möglichst viele der Kinder identifiziert werden, die im KFT 4-12+ R einen $PR > 85$ erreicht haben.

Anspruch auf 100 % Effektivität

Die besonders leistungsfähigen Kinder im mathematischen Bereich wurden von den LehrerInnen sehr unterschiedlich eingeschätzt. Erst ab Rang 18 wurden alle besagten Kinder erfasst. Wie in Tabelle 15 in Spalte „LEM $PR \geq 18$ “ angeführt, schätzten die LehrerInnen insgesamt 217 (von 219) Kinder auf einen Prozentrang ab 18 ein. Die Effizienz war mit 24,4 % sehr gering (geringst mögliche bei einer Effektivität von 100 % in dieser Stichprobe).

Betrachtet man die gereihten Ergebnisse der Einschätzung, werden alle 53 Kinder innerhalb der ersten 24 Ränge erkannt (siehe Tabelle 15, Spalte „LEM 1. bis 24. Rang“). Das bedeutet, dass die Streuung der im Test besonders leistungsfähigen Kinder sich in der Einschätzung durch die LehrerInnen über alle Ränge hinweg ausdehnte. Diese Methode kann deshalb als sehr ineffizient bezeichnet werden.

Weder durch die LehrerInnennomination, noch durch die Peernomination konnten alle besonders leistungsfähigen Kinder der Stichprobe erkannt werden. Die LehrerInnen identifizierten nur 17, die Kinder 38 der 53 besonders begabten SchülerInnen. Das bedeutet, dass nicht alle besonders leistungsfähigen Kinder von ihren MitschülerInnen oder LehrerInnen nominiert wurden.

Da die Einschätzungen der LehrerInnen und SchülerInnen sehr stark streuen und dadurch die Effizienz der Methoden bei einer 100 %igen Effektivität sehr gering ausfällt, wurde in einem weiteren Schritt der Anspruch, alle besonders leistungsfähigen Kinder in der Stichprobe zu erkennen, vernachlässigt. Die Effektivität wurde stattdessen auf einen Mindestwert von 66,6 % festgelegt. Es wurde je nach Methode ermittelt, ab welchem Prozentrang oder Rang dies der Fall war und die resultierende Effizienz bestimmt.

Tab. 15: Effektivität und Effizienz der verschiedenen Methoden bezogen auf die Kinder mit PR > 85 im mathematischen Testteil. N= 219

QPR			LEM PR $\geq$ 18	LEM 1. bis 24. Rang	LEM PR $\geq$ 70	LEM 1. bis 10 Rang	PN-F1 1. bis 8. Rang	LN-F1
Grundquote (PR > 85)	53	Selektions quote	217	219	101	124	121	33
Stichprobe in %	24,2		99,1	100	46,1	56,6	55,3	15,1
		Treffer	53	53	38	42	39	17
		Effektivität	100 %	100 %	71,7 %	79,25 %	73,58 %	32,08 %
		Effizienz	24,42 %	24,20 %	37,62 %	33,87 %	32,23 %	51,52 %

Legende: QPR = Quantitativtestergebnisse, LEM = mathematische LehrerInneneinschätzung, PR = Prozentrang, LN = LehrerInnennomination, PN = Peernomination (mit SN), F 1 = Frage 1 (Mathe), SN = Selbstnomination.

Anspruch auf mindestens 66,6 % Effektivität (= Identifikation von mindestens 36 besonders leistungsfähigen Kindern)

Berücksichtigte man alle Kinder, die bei der LehrerInneneinschätzung einen Prozentrang von mindestens 70 zugeteilt bekommen haben, erreichte man die Mindestgrenze von 66,6 % Effektivität (d. h. es sollen mindestens zwei Drittel der besonders leistungsfähigen Kinder erkannt werden). Die Effizienz lag dann bei 37,6 %, wie in

Tabelle 15 in der Spalte „LEM PR $\geq$ 70“ zu sehen ist. Die Mindestgrenze wurde im gerangreichten Fall ab einer Grenze des 10. Ranges erreicht. Da es mehrere zehnte Ränge gab, betrug die Effektivität 79,3 % und die Effizienz 33,9 %.

Im Falle der Peernomination wurden ebenfalls über 70 % der besonders Leistungsfähigen erkannt, wenn man die Ränge eins bis acht berücksichtigte. Die Effizienz war dabei 32,23 % und somit nahezu gleich hoch wie bei der Einschätzungsmethode.

Um im mathematischen Bereich möglichst viele der besonders leistungsfähigen Kinder zu identifizieren, eignete sich die LehrerInneneinschätzung am besten. Bei den Nominationsmethoden wurde in keinem Fall eine Effektivität von 100 % erreicht. Die Peernomination eignete sich ungefähr gleich gut mit der LehrerInneneinschätzung, als die Identifikationsrate von 100 auf 66,6% gesenkt wurde.

b) Sprachlicher Bereich

Grundquote: Es erreichten 25 Kinder einen Prozentrang > 85 im verbalen Teil des KFT 4-12+ R.

Tabelle 16 gibt einen Überblick der Effektivität und Effizienz der verschiedenen Methoden zur Identifikation der besonders leistungsfähigen Kinder.

Anspruch auf 100 % Effektivität

Die im sprachlichen Bereich besonders leistungsfähigen Kinder der Stichprobe wurden von den LehrerInnen mit der LehrerInneneinschätzung auf einen Prozentrang ab 40 eingeschätzt. Insgesamt schätzten die LehrerInnen 185 Kinder auf einen Prozentrang ab 40 ein. Die Effizienz war dabei, wie in Tabelle 16 in Spalte „LES PR ≥ 40 %“ zu sehen, sehr gering.

Betrachtet man die gerangreichten Ergebnisse der Einschätzung, zeigte sich, dass die besonders leistungsfähigen Kinder von den LehrerInnen sehr unterschiedlich eingeschätzt wurden. Erst ab Rang 16 wurden alle besagten Kinder erfasst. Insgesamt wurden 179 Kinder von den LehrerInnen auf die Ränge 1 bis 16 gestellt, was zu einer

nur geringfügig höheren Effizienz als bei Betrachtung ohne Rangreihung führte. Dies ist in Tabelle 16 in Spalte „LEM 1. bis 16. Rang“ zu erkennen.

Die Nominationsmethoden erreichten auch im sprachlichen Bereich keine 100 %ige Effektivität. Jeweils nur neun der 25 Kinder mit einem PR > 85 im Verbaltestteil wurden von den LehrerInnen bei Frage zwei (Erzählen) und drei (Ausdruck) nominiert.

149 Kinder haben sich bei Frage 2 selbst nominiert oder sind von MitschülerInnen nominiert worden. Es sind 20 der besonders leistungsfähigen Kinder erkannt worden.

Bei Frage 3 sind 162 Kinder von sich selbst oder von MitschülerInnen nominiert worden. Dabei wurden 23 der 25 besonders leistungsfähigen Kinder identifiziert. Es wurden folglich mit keiner der Nominationsmethoden alle besonders leistungsfähigen Kinder identifiziert.

Tab. 16: Effektivität und Effizienz der verschiedenen Methoden bezogen auf die Kinder mit PR > 85 im sprachlichen Testteil. N= 219

VPR			LES PR ≥ 40 %	LES 1. bis 16. Rang	LES PR ≥ 65 %	LES 1. bis 6. Rang	LN- F2	LN-F3
Grundquote (PR > 85)	25	Selektions quote	185	179	103	74	33	33
Stichprobe in %	11,4		84,5	81,7	47,0	33,8	15,1	15,1
		Treffer	25	25	19	17	9	9
		Effektivität	100,0 %	100,0 %	76,0 %	68,0 %	36,0 %	36,0 %
		Effizienz	13,5 %	14,0 %	18,5 %	23,0 %	27,3 %	27,3 %

Legende: VPR = Verbaltestergebnisse, LES = sprachliche LehrerInneneinschätzung, PR = Prozentrang, LN = LehrerInnennomination, PN = Peernomination, F 2 = Frage 2 (Erzählen), SN = Selbstnomination.

Tab. 17: Effektivität und Effizienz der verschiedenen Methoden bezogen auf die Kinder mit PR > 85 im sprachlichen Testteil. N= 219

VPR			PN-F2 1. bis .8. Rang mit SN	PN-F3 1. bis 5. Rang mit SN	LES 1. bis 3. Rang
Grundquote (PR > 85)	25	Selektions quote	103	64	40
Stichprobe in %	11,4		47,0	29,2	18,3
		Treffer	17	18	9
		Effektivität	68,0 %	72,0 %	36,0 %
		Effizienz	16,5 %	28,1 %	22,5 %

Legende: VPR = Verbaltestergebnisse, LES = sprachliche LehrerInneneinschätzung, PR = Prozentrang, LN = LehrerInnennomination, PN = Peernomination, F 3 = Frage 3 (Ausdruck), SN = Selbstnomination.

Anspruch auf mindestens 66,6 % Effektivität (= Identifikation von mindestens 17 besonders leistungsfähigen Kindern)

Berücksichtigte man alle Kinder, die bei der LehrerInneneinschätzung einen Prozentrang von mindestens 65 zugeteilt bekommen haben, wurden mindestens zwei Drittel der besonders leistungsfähigen Kinder erkannt. Es wurde eine Effizienz von 18,5 % erreicht, wie in Tabelle 16 in der Spalte „LES PR $\geq$ 65“ zu sehen ist. Im gerangreichten Fall müssen die Ränge eins bis sechs berücksichtigt werden, um mindestens zwei Drittel der Gesuchten zu erkennen. Die Effizienz lag dann, wie in Tabelle 16 in Spalte „LES 1. bis 6. Rang“ zu sehen, bei 23,0 %.

Die Peernomination lieferte zwei Drittel Effektivität, wenn bei Frage 2 der erste bis achte Rang berücksichtigt wird. Wie in Tabelle 17 Spalte „PN-F2 1. bis 8. Rang“ zu sehen, war die Effizienz niedrig.

Bei Frage 3 wurden zwei Drittel erkannt, wenn die Ränge eins bis fünf berücksichtigt werden. Hier war die Effizienz mit 28,1 % in diesem Bereich am höchsten (siehe Tabelle 17, Spalte „PN-F3 1. bis 5. Rang“).

Um im sprachlichen Bereich möglichst viele der besonders leistungsfähigen Kinder zu identifizieren, eignete sich die Einschätzung der LehrerInnen am besten. Die Peernomination – Frage 3 eignete sich dagegen am besten, als die Identifikationsrate von 100 auf 66,6% gesenkt wurde.

Kann durch Kombination der Methoden eine erfolgreichere Identifikation der leistungsfähigsten Kinder erreicht werden?

In Anbetracht der Komplexität der Berechnungen und Darstellung wurde die Überprüfung der Gewinne durch eine Kombination der Verfahren selektiv mit den bei den vorangehenden Berechnungen erfolgreichsten Verfahren durchgeführt.

Wurden im mathematischen Bereich die zwei erfolgreichsten Vorgehensweisen bei einem Effektivitätsanspruch von mindestens 66,6 % zusammengelegt, ergab sich eine höhere Trefferanzahl. Durch die Kombination von „LEM 1. bis 10 Rang“ und „PN–F1 1. bis 8. Rang“ konnten drei besonders leistungsfähige Kinder mehr identifiziert werden.

Im sprachlichen Bereich wurde überprüft, ob eine Kombination der zwei erfolgreichsten Vorgehensweisen bei einem Effektivitätsanspruch von mindestens 66,6 % eine Erhöhung der Trefferanzahl bewirkte. Die Kombination von „LES PR $\geq$ 65 % und „PN–F3 1. bis 5. Rang“ ergab vier zusätzliche Treffer.

Es konnte durch eine Kombination der Verfahren eine erfolgreichere Identifikation der besonders leistungsfähigen Kinder erreicht werden.

3.3.6 Ergebnisse zum Einfluss verschiedener LehrerInnenmerkmale auf die Einschätzung der LehrerInnen

Zur näheren Beleuchtung der Heterogenität der Vergleichskomponenten der elf verschiedenen LehrerInnen, wurden mögliche Moderatorvariablen untersucht. Es wurde eine mehrfaktorielle ANOVA gerechnet. Dazu wurden als abhängige Variable einmal die Niveauebene, einmal die Differenzierungskomponente und einmal die Vergleichskomponente herangezogen. Als unabhängige Variablen dienten die dichotomisierten Variablen Unterrichtsdauer, Berufserfahrung und Wissen über Hochbegabung (Fortbildung besucht oder nicht).

Die Voraussetzungen für die Berechnung wurden überprüft. Es war eine Normalverteilung bei den verschiedenen Komponenten gegeben (K-S Test, Signifikanzniveau $p \geq .01$) (siehe Anhang 2). Die Unabhängigkeit der Faktorstufen war ebenfalls gegeben. Die Homoskedastizität wurde mit dem Levene-Test geprüft. Die Fehlervarianzen zwischen den Gruppen waren gleich (siehe Anhang 4).

Tab. 18: Mehrfaktorielle ANOVA zur Überprüfung der Moderatoreffekte. Niveauebene mathematischer Bereich. N= 11

	df	F	Sig.
Unterrichtsdauer	1	.435	.539
Berufserfahrung	1	.221	.658
FortbildungHB	1	.414	.548
Unterrichtsdauer * Berufserfahrung	1	.612	.470
Unterrichtsdauer * FortbildungHB	1	.002	.964
Berufserfahrung * FortbildungHB	0	-	-

Legende: df = Freiheitsgrade, Sig. = Signifikanz, FortbildungHB = Fortbildung zum Thema Hochbegabung besucht oder nicht/ Wissen über Hochbegabung.

Tab. 19: Mehrfaktorielle ANOVA zur Überprüfung der Moderatoreffekte. Niveauebene sprachlicher Bereich. N= 11

	df	F	Sig.
Unterrichtsdauer	1	1,036	.355
Berufserfahrung	1	.178	.691
FortbildungHB	1	.204	.670
Unterrichtsdauer *	1	.183	.687
Berufserfahrung			
Unterrichtsdauer *	1	.216	.662
FortbildungHB			
Berufserfahrung *	0	-	-
FortbildungHB			

Legende: df = Freiheitsgrade, Sig. = Signifikanz, FortbildungHB = Fortbildung zum Thema Hochbegabung besucht oder nicht/ Wissen über Hochbegabung.

Tab. 20: Mehrfaktorielle ANOVA zur Überprüfung der Moderatoreffekte. Differenzierungskomponente mathematischer Bereich. N= 11

	df	F	Sig.
Unterrichtsdauer	1	.119	.745
Berufserfahrung	1	.173	.694
FortbildungHB	1	2,322	.188
Unterrichtsdauer *	1	.322	.595
Berufserfahrung			
Unterrichtsdauer *	1	1,674	.252
FortbildungHB			
Berufserfahrung *	0	-	-
FortbildungHB			

Legende: df = Freiheitsgrade, Sig. = Signifikanz, FortbildungHB = Fortbildung zum Thema Hochbegabung besucht oder nicht/ Wissen über Hochbegabung.

Tab. 21: Mehrfaktorielle ANOVA zur Überprüfung der Moderatoreffekte. Differenzierungskomponente sprachlicher Bereich. N= 11

	df	F	Sig.
Unterrichtsdauer	1	0	.996
Berufserfahrung	1	.490	.515
FortbildungHB	1	2,091	.208
Unterrichtsdauer *	1	.429	.542
Berufserfahrung			
Unterrichtsdauer *	1	.236	.648
FortbildungHB			
Berufserfahrung *	0	-	-
FortbildungHB			

Legende: df = Freiheitsgrade, Sig. = Signifikanz, FortbildungHB = Fortbildung zum Thema Hochbegabung besucht oder nicht/ Wissen über Hochbegabung.

Tab. 22: Mehrfaktorielle ANOVA zur Überprüfung der Moderatoreffekte. Vergleichskomponente mathematischer Bereich. N= 11

	df	F	Sig.
Unterrichtsdauer	1	.107	.756
Berufserfahrung	1	.037	.855
FortbildungHB	1	.278	.620
Unterrichtsdauer *			
Berufserfahrung	1	.331	.590
Unterrichtsdauer *			
FortbildungHB	1	.130	.733
Berufserfahrung *			
FortbildungHB	0	-	-

Legende: df = Freiheitsgrade, Sig. = Signifikanz, FortbildungHB = Fortbildung zum Thema Hochbegabung besucht oder nicht/ Wissen über Hochbegabung.

Tab. 23: Mehrfaktorielle ANOVA zur Überprüfung der Moderatoreffekte. Vergleichskomponente sprachlicher Bereich. N= 11

	df	F	Sig.
Unterrichtsdauer	1	.160	.706
Berufserfahrung	1	.056	.822
FortbildungHB	1	.036	.856
Unterrichtsdauer *			
Berufserfahrung	1	.000	.989
Unterrichtsdauer *			
FortbildungHB	1	.163	.703
Berufserfahrung *			
FortbildungHB	0	-	-

Legende: df = Freiheitsgrade, Sig. = Signifikanz, FortbildungHB = Fortbildung zum Thema Hochbegabung besucht oder nicht/ Wissen über Hochbegabung.

Die Tabellen 18 bis 23 zeigen, dass keine signifikanten Unterschiede in der diagnostischen Kompetenz zwischen den Gruppen gefunden wurden. Die Stichprobe ist für die Berechnung einer ANOVA sehr klein, weswegen über die Wechselwirkungen zwischen Berufserfahrung und Wissen über Hochbegabung sowie über eine Wechselwirkung aller drei unabhängigen Variablen keine Aussagen getroffen werden konnten.

4 Zusammenfassung und Diskussion

Im folgenden Kapitel werden die Ergebnisse diskutiert und im Anschluss Fördermöglichkeiten diagnostischer Kompetenz aufgezeigt.

LehrerInnen wird zum Ende der Volksschulzeit eine verantwortungsvolle Aufgabe übergeben, indem sie eine wichtige Rolle bei der Schulübergangsentscheidung der Kinder spielen. Eine hohe Kompetenz hinsichtlich der Leistungsfähigkeitseinschätzung ist von großer Relevanz, um die Kinder auf einen zu ihren Leistungsfähigkeiten passenden Bildungsweg weisen zu können. Wie diese Kompetenz bei österreichischen VolksschullehrerInnen ausgeprägt ist, sollte in der vorliegenden Arbeit untersucht werden. Dazu wurde anhand einer Stichprobe aus elf österreichischen vierten Klassen der Volksschule und ihren KlassenlehrerInnen die diagnostische Kompetenz der LehrerInnen bewertet. Interessant war weiterhin, ob es Einflussgrößen gibt, die systematisch mit der Höhe der diagnostischen Kompetenz zusammen hängen. Mit der Identifikation solcher Merkmale könnte eine Förderung der Kompetenz wirkungsvoller geplant und gestaltet werden.

Es kann in vieler Hinsicht wichtig sein, besonders hohe Leistungsfähigkeit zu erkennen (z. B. um das Potenzial einer Gesellschaft zu nutzen, indem hochbegabte Kinder einem Hochbegabtenförderprogramm zugeführt werden). Es gibt verschiedene Vorgehensweisen, diese zu identifizieren. Die psychologische Testdiagnostik bietet hierbei die brauchbarsten Verfahren und wird deshalb als Königsweg bezeichnet. Die Anwendung ist allerdings aufwändig und kann infolge dessen nur schwer bei einer großen Population durchgeführt werden. Daher werden alternative Vorgehensweisen für eine erste überblicksartige Identifizierung Hochbegabter gesucht. Diese sollten eine möglichst genaue und treffsichere Vorauswahl potenziell besonders leistungsfähiger Kinder ermöglichen, die in weiterer Folge mit den differenzierten Verfahren begutachtet werden. In vorliegender Arbeit wurden verschiedene Methoden auf ihre Screeningtauglichkeit zur Identifikation besonders leistungsfähiger Kinder untersucht.

Eine Frage der Arbeit bezog sich auf die Güte der LehrerInneneinschätzungen. Die Ergebnisse der Kreuztabellen zeigen, dass die Lehrkräfte ungefähr gleich viele ihrer SchülerInnen in deren Leistungsfähigkeit überschätzten wie korrekt einschätzten. Die berechneten Niveauelementen verdeutlichen, dass das gesamte Klassenniveau von nahezu drei Viertel der LehrerInnen überschätzt wurde. Dies unterstreicht die schon in der Literatur berichteten Überschätzungstendenzen von LehrerInnen. Die Ergebnisse

stimmen mit denen von Hesse und Latzko (2009), Hosenfeld et al. (2002) und Lang (2008) überein. Österreichische Lehrkräfte scheinen, ähnlich wie deutsche Lehrkräfte, ein größeres kognitives Potenzial bei den Kindern zu sehen, als eigentlich vorhanden ist. Bei Lang (2008) war dieses Ergebnis noch mit Vorsicht zu interpretieren, da die untersuchte Stichprobe unterdurchschnittlich in der Intelligenztestung abschnitt, und es dadurch mit höherer Wahrscheinlichkeit zu einer Überschätzung kam. Da die Stichprobe der vorliegenden Arbeit keine Tendenz zu unterdurchschnittlichen Intelligenztestleistungen aufwies, bekräftigen die Ergebnisse die Überschätzung der Leistungsfähigkeit der SchülerInnen durch die Lehrkräfte.

Die Differenzierungskomponenten stellen die Einschätzung der Streuung der Leistungsfähigkeit innerhalb der Klassen dar. Die Verteilung wurde im mathematischen Bereich von fast allen LehrerInnen unterschätzt. Im sprachlichen Bereich hingegen wurde die Verteilung von fast drei Viertel der Lehrkräfte unter-, von zwei LehrerInnen über- und in einem Fall genau eingeschätzt. Diese Ergebnisse zur Differenzierungskomponente decken sich mit denen von Schrader und Helmke (1987). Sie fanden heraus, dass die Unterschätzung im Gegensatz zur Überschätzung stärker mit einem Leistungszuwachs leistungsschwacher SchülerInnen zusammenhängt. Sie nehmen an, dass das Verhalten der Lehrkraft teilweise auf ihrer Einschätzung basiert und dies wiederum von den Kindern wahrgenommen wird (Pygmalioneffekt, s. S. 9). Leistungsschwächere Kinder haben nach Meinung der Autoren durch eine homogenere Wahrnehmung der Leistungsunterschiede innerhalb der Klasse, die sich von der Lehrkraft auf die Schüler überträgt, eine günstigere Selbsteinschätzung und dadurch weniger „Demotivierungs- und Resignationserscheinungen“ (S. 35). Lang (2008) berichtete dagegen von einer Überschätzung der Streuung. Auch dieses Ergebnis steht mit hoher Wahrscheinlichkeit mit dem ungewöhnlich niedrigen Durchschnittsniveau der Leistungsfähigkeit der untersuchten oberösterreichischen Kinder in Zusammenhang und sollte deshalb nicht verallgemeinert werden. Weinert und Schrader (1986, zitiert nach Hesse & Latzko, 2009) bewerten eine leichte Überschätzung des Leistungsfähigkeitsniveaus und eine mäßige Unterschätzung der Streuung, wie in vorliegender Arbeit gefunden wurde, als positiv. Sie halten diese Perspektive günstig für das Verhalten der LehrerInnen, da so die „Lehrkräfte für sich immer wieder neue pädagogische Handlungsanreize generieren“ können (S. 30). Insgesamt haben sie die Einstellung, dass eine genaue Einschätzung durch die LehrerInnen in einer komplexen Umgebung wie Schulunterricht nicht möglich sei. Ob eine Verbindung zwischen

Urteilsgenauigkeit und Lernerfolg in der vorliegenden Studie vorhanden ist, konnte nicht untersucht werden, da der Lernerfolg nicht erfasst wurde.

Spinath (2005) argumentiert, dass den Lehrkräften zumindest bewusst sein sollte, dass ihre Einschätzungen ungenau sein können und sie bei wichtigen Entscheidungen darauf Rücksicht nehmen. Schrader und Helmke (1987) vermuten, dass eine positiv verzerrte Erwartungshaltung bezüglich der Leistungsfähigkeit der Klasse auf Seiten der LehrerInnen zu einer erhöhten Anstrengungsbereitschaft führen könnte. Die Forscher sehen als Folge eine „leistungsförderliche Atmosphäre“ (S. 34). Die Effektivität des Unterrichts könne allerdings nur gewährleistet sein, wenn die Überschätzung nicht zu groß sei. Ab einem bestimmten Grad an Überschätzung müsse mit Überforderung aufgrund zu geringer Lernvoraussetzungen der SchülerInnen für die Anforderungen des Unterrichts gerechnet werden. In ihrer Untersuchung von 32 LehrerInnen konnte diese Vermutung allerdings nicht bestätigt werden. Sowohl jene, die das Niveau unterschätzten als auch jene, die das Niveau stark überschätzten, erzielten in ihrer Klasse einen stärkeren Lernerfolg als jene, die leicht überschätzten. Eine Interpretation der Autoren liegt in einer nahezu gegenteiligen Begründung als im Vorfeld argumentiert. Sie äußern, dass der größere Lernerfolg das Resultat einer gesteigerten Motivation der LehrerInnen sein könne, die eher schwachen Kinder durch besonderen beruflichen Einsatz zu unterstützen. Hesse und Latzko (2009) betonen, dass eine regelmäßige Reflexion der meist implizit ablaufenden Beurteilungsprozesse wichtig sei, um mögliche Fehler einzudämmen. Sie fordern die Ausdehnung des Wissenserwerbs im Gebiet der Psychologie und der Methoden, damit die LehrerInnen in der Lage sind, die Reflexionen selbst durchführen zu können. Im folgenden Kapitel wird ausführlicher auf die Möglichkeiten eingegangen.

Die Vergleichskomponenten werden im Großteil der Studien zur diagnostischen Kompetenz als Maß für die Güte der LehrerInneneinschätzungen genommen. Sie befinden sich in vorliegender Arbeit mit $r = .6$ im mathematischen Bereich und $r = .59$ im sprachlichen Bereich in der Spannweite der in der Literatur (z. B. Hoge & Coladarci, 1989; Schrader & Helmke, 2001; Spinath, 2005) berichteten Mittelwerte. Nahezu drei Viertel der LehrerInnen schätzten die Rangreihung genau ein.

Rost, Sparfeldt und Schilling (2006) haben Gründe erarbeitet, die die Ungenauigkeit von LehrerInnenurteilen erklären könnten. Sie vermuten, dass das Verständnis der Lehrkräfte vom Begabungsbegriff ein anderes, breiter gefasstes ist, als das, was ein

Intelligenztest misst. Außerdem führen sie an, dass Kinder, die sich angepasster verhalten auch eher als begabter eingeschätzt werden. Eine hohe Leistungsmotivation, die nicht zwingend mit hoher Leistungsfähigkeit zusammenhängt, mag ebenfalls häufig dazu beitragen, ein Kind als besser begabt zu bewerten.

Die Ergebnisse zeigen, dass die Qualität der LehrerInneneinschätzung eher nicht zufrieden stellend ist. Da jedoch von ihrem Einschätzungsvermögen Bildungskarrieren von Kindern abhängen, sollten den Lehrkräften die Diskrepanzen zwischen ihrer Einschätzung und den Testwertergebnissen aufgezeigt und die Relevanz verdeutlicht werden. Wichtig ist, eine Verbesserung der diagnostischen Kompetenz und vorangehend, die Bereitschaft der Lehrkräfte zum Erarbeiten und Üben der Kompetenz.

Bei der Analyse der SchülerInnen, die die ersten drei Ränge beim Intelligenztest innerhalb der Klasse belegten, stellte sich für den mathematischen Bereich die LehrerInneneinschätzung als effektivste und effizienteste Methode heraus. Im sprachlichen Bereich konnte mit der Selbstnomination – Frage 2 (Erzählen) die höchste Effektivität erreicht werden, allerdings bei einer geringen Effizienz von einem Drittel. Ein ausgeglicheneres Bild ergab sich für die Peernomination – Frage 3 (Ausdruck). Hier lag die Effektivität zwar etwas tiefer, dafür aber die Effizienz höher. Im mathematischen Bereich konnte durch eine Kombination der LehrerInneneinschätzung und der Selbstnomination eine Erhöhung der Trefferzahl erreicht werden. Im sprachlichen Bereich ergab die Kombination von LehrerInneneinschätzung und Peernomination – Frage 3 eine höhere Trefferzahl.

Als Screeninginstrument für die Identifikation besonders leistungsfähiger Kinder der vierten Klasse Volksschule eigneten sich im mathematischen und im sprachlichen Bereich von den untersuchten Instrumenten die Einschätzungsmethoden am besten. Mit dem Anspruch, eine vollständige Identifikation zu erreichen, sollte folglich die LehrerInneneinschätzung die Methode der Wahl sein. Es konnte allerdings bei dieser Vorgehensweise keine akzeptable Effizienz erreicht werden. Sehr viele nicht besonders leistungsfähige Kinder wurden fälschlicherweise als besonders leistungsfähig eingeschätzt. Somit wäre die Durchführung einer LehrerInneneinschätzung besonders leistungsfähiger Kinder mit Anspruch auf hundertprozentige Identifikationsquote nicht zielführend. Wild (1991) kam in seiner Studie auf sehr ähnliche Ergebnisse. Geht man den Kompromiss ein, nicht alle hochleistungsfähigen Kinder zu erkennen, dafür aber eine höhere Effizienz zu erreichen, eignete sich im mathematischen Bereich die

Peernomination gleich gut wie die LehrerInneneinschätzung. Im sprachlichen Bereich zeigte sich unter diesen Bedingungen die höchste Effizienz bei Anwendung der Peernomination – Frage 3. Eine Kombination der erfolgreichsten Methoden, bei einem Anspruch auf mindestens 66,6 % Effektivität, brachte auch bei der Identifikation der insgesamt Begabtesten einen Zuwachs der Erkennensrate. Wild (1991) hingegen sah in keinem Fall Vorteile in einer Kombination mehrerer Methoden und auch keinen Vorteil der Peernomination. Vorliegende Arbeit zeigt jedoch, dass die Kombination von LehrerInneneinschätzung und der von Wild (1991) als sehr schlecht bewerteten Selbstnomination und auch die Kombination von LehrerInneneinschätzung und Peernomination, Informationsgewinn mit sich bringt. Im Vergleich zu den Ergebnissen, die Heller et al. (2005) bei Verwendung von Checklisten anführen, zeigte sich ein Vorteil sowohl der LehrerInneneinschätzung als auch der Nominationsmethoden gegenüber den Checklisten.

Generell ist zu sagen, dass keine der angewandten Methoden eine überzeugende Screeningtauglichkeit besitzt. Je nach Kontext könnte allerdings der Einsatz von Peernominationen argumentiert werden. In der Untersuchung zeigte sich, dass bei der Identifikation der klassenintern leistungsfähigsten Kinder durch die Peernomination (Frage 3) nicht nur über die Hälfte der im Test leistungsfähigsten Kinder erkannt wurde, sondern auch die Effizienzrate des Vorgehens relativ hoch ausfiel. Weniger als die Hälfte der Nominierten wurde von den SchülerInnen fälschlicherweise nominiert. Im Hinblick auf die Identifikation der allgemein leistungsfähigsten SchülerInnen zeigte sich zumindest eine vergleichbare Tauglichkeit der Peernomination mit der LehrerInnen-einschätzung im mathematischen Bereich. Hier war sie allerdings nur wirksam, als die Identifikationsrate von 100 auf 66,6% gesenkt wurde. Eine Kombination verschiedener Methoden schien wiederum einen Vorteil im Identifikationsvorgehen mit sich zu bringen. Zwischen den erhobenen Moderatorvariablen und den LehrerInneneinschätzungen konnte kein Zusammenhang festgestellt werden. Um anhand der berechneten statistischen Werte Rückschlüsse auf andere VolksschullehrerInnen zu ziehen, ist die Stichprobe zu klein.

4.1 Kritik und Ausblick

Kritisch ist zur vorliegenden Studie, und generell zu den meisten Studien, die sich mit dem Thema der diagnostischen Kompetenz beschäftigen, anzumerken, dass der Vergleich der Einschätzungen mit einer einmaligen Testleistung Mängel aufweist. Karing (2009) argumentierte, dass eine völlige Übereinstimmung von LehrerInnenurteilen und einer im Intelligenztest gemessenen Leistung sehr unwahrscheinlich sei, da das Urteil über eine längere Zeitspanne getroffen werde, wohingegen die Testung einen stichpunktartigen Ausschnitt der Kompetenz darstelle. Es sollte daher erwogen werden, die Intelligenztestung zu mehreren Zeitpunkten vorzunehmen oder die Anweisungen an die LehrerInnen zu spezifizieren.

Kritisch zu bewerten ist weiterhin, dass die in Tests erfasste Leistungsfähigkeit von einer Reihe von Faktoren abhängt, die in den meisten Studien unzureichend kontrolliert worden sind. Im Bericht der VERA 4 (Vergleichsarbeiten in vierten Grundschulklassen in sieben Bundesländern Deutschlands) wird eine Studie erwähnt, nach der das gesamte Leistungsniveau einer Klasse sinkt, je höher der Anteil der Kinder mit Deutsch als nicht dominanter Sprache ist. Ähnliche Auffälligkeiten sind im Vergleich der Leistungsfähigkeitsmessungen der niederösterreichischen (vorliegende Arbeit) und oberösterreichischen (Lang, 2008) Schulklassen zu beobachten (siehe Kapitel 3.1.2). Zukünftige Arbeiten müssten daher mehr Gewicht darauf legen, mögliche Störvariablen der Stichprobe zu kontrollieren. Anzustreben sind Stichproben, die in Bezug auf die Muttersprache repräsentativ für die jeweilige Bevölkerung sind. Das starke Abweichen der Intelligenztestergebnisse der Stichprobe von Lang (2008) von den deutschen Normen (im Gegensatz zur vorliegenden Arbeit) mag zusätzlich am Messzeitpunkt liegen. Die Erhebungen fanden früher im Schuljahr statt als die in Niederösterreich. Der österreichische Lehrplan der Volksschule (bmuk, 2003) sieht zum Beispiel die Bruchrechnung als Rechenoperation erst in der vierten Schulstufe vor. Bruchrechnung ist Inhalt einiger Items des KFT 4-12+ R. Es bleibt unklar, ob einige oberösterreichische Klassen, die zu einem früheren Zeitpunkt untersucht wurden, die Bruchrechnung im Unterricht noch nicht behandelt hatten, während einige niederösterreichischen Klassen dies eventuell aufgrund des späteren Messzeitpunktes schon gelernt hatten. Wichtig für die Vergleichbarkeit mehrerer Stichproben sind demzufolge die Berücksichtigung des Zeitpunkts der Erfassung und damit der Fortschritt des Lehrplans.

Weiterhin wäre eine nähere Betrachtung der diagnostischen Kompetenz möglich, wenn über die Intelligenzermessung hinaus die Schulnoten der Kinder mit erhoben werden. Dies brächte die Möglichkeit mit sich, zu beurteilen, ob das Erkennen von Underachievern ein besonderes Problem von LehrerInnen darstellt.

ForscherInnen im Bereich der diagnostischen Kompetenz und der Identifikation von besonderer kognitiver Begabung wird auf Grundlage der vorliegenden Arbeit nahe gelegt, in ihren Arbeiten zukünftig die Kombination verschiedener Methoden vorzunehmen. Außerdem sollte der Vollständigkeit halber beim Überprüfen möglicher Einflussstärken neben der Vergleichskomponente auch die Differenzierungs- und Niveauelemente mit in die Berechnungen einbezogen werden.

Das Konzept der diagnostischen Kompetenz ist schwer greifbar und nur ein Teil der von Schrader (1989) formulierten Definition ist messbar. Einstellungen und Motivation der LehrerInnen könnten ebenfalls Teilbereiche der diagnostischen Kompetenz darstellen. Die Definition sollte, um mehr Klarheit in den Forschungsbereich zu bringen, neu durchdacht und mit wissenschaftlichen Methoden überprüft werden.

4.2 Fördermöglichkeiten diagnostischer Kompetenz

Die berichteten Ergebnisse machen deutlich, dass bei den meisten VolksschullehrerInnen im Hinblick auf die kognitive Leistungsfähigkeit der Kinder eine mittlere diagnostische Kompetenz mit einer Überschätzung des Klassengesamtniveaus vorlag. Langfeldt sieht die „Diagnosekompetenz als Kernkompetenz für erfolgreiches Unterrichten und pädagogisches Handeln“ (2006, S. 199). Auch van Ophuysen (2006) zeigte in ihrer Untersuchung die Bedeutung einer genauen Einschätzung im Zusammenhang mit der Schullaufbahneempfehlung auf, die in der Regel nach dem vierten Schuljahr der Volksschule vorgenommen wird. Nach dem österreichischen Schulorganisationsgesetz § 40 ist die Aufnahme an einer allgemein bildenden höheren Schule (AHS) in erster Linie abhängig von der Benotung durch die LehrerInnen in den Fächern Deutsch, Lesen und Mathematik. Eine Fehleinschätzung kann nach van Ophuysen (2006) im Falle einer Unterschätzung zu „Unterforderung, suboptimalem Lernfortschritt und motivationalen Defiziten“, bei Überschätzung zu „Überforderung, schlechten Noten, Motivationsverlusten und Schulangst“ führen (S. 160). Karing (2009) bemängelt, dass die universitäre Ausbildung der Grundschullehrkräfte die diagnostische Kompetenz nicht fördere, obwohl die Kultusministerkonferenz dies fordere. In Österreich

sind die Lehrpläne der Bachelor-Studiengänge Lehramt Volksschule an den staatlichen Pädagogischen Hochschulen unterschiedlich gestaltet. Während Begabung in Vorarlberg und Kärnten nicht gesondert als Themenblock behandelt wird, gibt es in Salzburg ein ganzes Modul zur Begabungsförderung und auch in der Steiermark werden die Grundlagen der Begabungs- und Begabtenförderung gelehrt. Die übrigen Bundesländer haben das Thema Begabungsförderung zumindest als Wahlfach oder Wahlpflichtfach im Angebot (Pädagogische Hochschulen, 2010).

Auch Helmke, Hosenfeld und Schrader (2004) unterstreichen, u. a. vor dem Hintergrund von Ergebnissen von PISA 2000, die Wichtigkeit der Förderung diagnostischer Kompetenz bei Lehrkräften. Hohe diagnostische Kompetenz sei unabdingbar für die individuelle Förderung der einzelnen Schüler und für effektiven Unterricht. Vergleichsarbeiten, die, wie unten beschrieben, teilweise schon regelmäßig in deutschen Schulen durchgeführt werden, sollen nach Forderung der Autoren nicht nur als Erfassungsinstrument sondern ebenso als Interventionsinstrument (im Sinne von Verbesserung der diagnostischen Kompetenz) verwendet werden.

Zunächst wird eine Alternativmethode zur klassischen Erhebung der diagnostischen Kompetenz erläutert. Der „Simulierte Klassenraum“ nach Südkamp, Möller und Pohlmann (2008) scheint viel versprechend zu sein. Ein Computerprogramm simuliert SchülerInnen, welche nach vorprogrammierten Leistungsparametern Antworten auf Fragen geben, die in gleicher Weise auch in einer realen Unterrichtssituation gestellt werden könnten. Da sich gleiche Ergebnisse wie in den übrigen Studien zur diagnostischen Kompetenz ergaben, kann das Instrument als valide angesehen werden. Ein Vorteil gegenüber der klassischen Erhebung besteht im geringeren Zeitaufwand, und darin, dass Kinder nicht in die Studie mit einbezogen werden müssen. Eine weitere Erfassungsmethode stellt die Selbstdiagnose nach Helmke (2007, zitiert nach Hesse & Latzko, 2009) dar. In fünf Schritten wird ausführlich erklärt, wie Lehrkräfte ihre persönliche diagnostische Kompetenz mit Hilfe der eigenen Klasse einschätzen können. Im ersten Schritt wählen die LehrerInnen ein SchülerInnenmerkmal (z. B. die Leistungsfähigkeit) oder ein Aufgabenmerkmal, das sie einschätzen wollen. Danach treffen sie eine Prognose auf einer fünfstufigen Skala für jeden einzelnen Schüler und jede einzelne Schülerin ihrer Klasse für das gewählte SchülerInnenmerkmal. Im Falle der Einschätzung des Aufgabenmerkmals soll vorausgesagt werden, von wie vielen SchülerInnen die jeweilige Aufgabe gelöst werden wird. In einem dritten Schritt wird die tatsächliche Ausprägung der Variable erhoben, zum Beispiel anhand eines

standardisierten Tests. Die Ergebnisse des Tests müssen auf die fünfstufige Skala übertragen werden. Der nächste Schritt beinhaltet den Vergleich der Prognose mit den Ergebnissen aus der Testung. Dazu wird die Niveauekomponente berechnet. Die Lehrkräfte bekommen hierzu eine schriftliche Anleitung. Ein letzter Schritt gibt Hilfestellungen zur Analyse des Vergleichs, zeigt mögliche Fehlerquellen bei der Einschätzung auf und animiert zur aktiven Arbeit zur Verbesserung der diagnostischen Kompetenz.

Im Bildungsbericht Österreich 2009, herausgegeben von Specht, wird deutlich, dass Österreich als einziges Land LehrerInnen als alleinige Notengeber ohne weitere objektivierende Maßnahmen, wie zum Beispiel das Zentralabitur in Deutschland, zu Rate zieht. Ein Vorschlag für die Verbesserung der diagnostischen Kompetenz sind die in Deutschland angewendeten Vergleichsarbeiten in vierten Grundschulklassen in sieben Bundesländern (VERA; Berlin, Brandenburg, Bremen, Mecklenburg-Vorpommern, Nordrhein-Westfalen, Rheinland-Pfalz, Schleswig-Holstein). Alle ViertklässlerInnen werden anhand speziell dafür entwickelter mathematischer und sprachlicher Aufgaben getestet. Die Lehrkräfte werden aufgefordert, vorausszusagen, wie viele Kinder die Aufgaben lösen werden und können im Nachhinein sehen, wie gut sie eingeschätzt haben. Jeder Lehrer und jede Lehrerin kann über die Rückmeldung von VERA etwas über die eigenen Stärken und Schwächen bei der Einschätzung erfahren und sich mit anderen Lehrkräften vergleichen. Gründe, die zu einer ungenauen Einschätzung des Niveaus der Klasse führen, können zu hohe Ansprüche (z. B. wenn die Einschätzung im Vergleich zu KollegInnen hoch ausfällt) oder unerwartet schlechtes Abschneiden der SchülerInnen sein (z. B. weil die Lehrkraft nicht berücksichtigt hat, dass es sich um eine außergewöhnliche Testsituation gehandelt hat, sich zu sehr an den leistungsstärksten SchülerInnen orientiert hat oder die SchülerInnen unmotiviert oder ängstlich waren). Von Hesse und Latzko (2009) wird der Vorschlag geäußert, einen Fragenkatalog für teilnehmende Lehrkräfte zu entwickeln, der bei der Suche nach Gründen für ungenaue Einschätzungen helfen soll. Auf Basis dieser Liste sollte dann nach Methoden gesucht werden, anhand derer die Hypothesen überprüft werden könnten. Beispielsweise könnten ausgewählte Aufgaben erneut vorgegeben und gleichzeitig die Kinder beobachtet und befragt sowie ihnen Unterstützung geboten werden. Es sollte zudem nach Anhaltspunkten gesucht werden, die das Vorhandensein bestimmter Ursachen anzeigen. Eine weitere Übung besteht darin, ein Vorgehen für die Dokumentation der Leistungen auszuarbeiten. Außerdem sei ein Zusammentreffen und

ein Austausch verschiedener Lehrkräfte von Vorteil. So könnten unterschiedliche Sichtweisen zusammengeführt und voneinander gelernt werden. Das Endprodukt sollte eine Liste von Fehlerquellen verbunden mit Lösungsansätzen sein.

Zur Verbesserung der diagnostischen Kompetenz schlagen einige AutorInnen die Teilnahme der LehrerInnen an einschlägigen Seminaren vor (z. B. Ziegler & Stöger, 2003). Schrader und Helmke (2001) dagegen fordern, ähnlich wie im Bildungsbericht Österreich, die Verwendung von objektiven, normbezogenen Leistungstests. Die Lehrkräfte sollen einschätzen, wie die Kinder in der spezifischen Testung abschneiden und bekommen später Rückmeldung über die Ergebnisse. Die eigenen Einschätzungstendenzen könnten so reflektiert werden, und es komme zu einem Übungseffekt. Schrader und Helmke (2001) plädieren dafür, die Anwendung solcher Testungen in die Ausbildung von LehrerInnen stärker zu integrieren.

Wahl, Weinert und Huber (1997, zitiert nach Schrader, 2009) schlagen vor, dass LehrerInnen im alltäglichen Unterricht anhand absichtlich herbeigeführter Situationen die Leistungen der SchülerInnen vorhersagen und mit den tatsächlichen Leistungen in Form von Testdaten vergleichen. Die Lehrkräfte überführen implizite Urteile in explizite Urteile, welche mit Testergebnissen vergleichbar sind und erhalten somit automatisch Rückmeldung über die Genauigkeit ihrer Einschätzung. Setzt sich die Lehrkraft nun mit möglichen Gründen einer Unter- oder Überschätzung auseinander und nimmt sie weitere Anhaltspunkte für die Einschätzung hinzu, könnte dies nach Wahl, Weinert und Huber (1997, zitiert nach Schrader, 2009) zu einer verbesserten diagnostischen Kompetenz führen.

Um diese Fördermöglichkeiten zu realisieren, müssen den Lehrkräften Testmaterial und Informationsmaterial zur Durchführung der genannten Maßnahmen zur Verfügung gestellt werden. Im Idealfall sollte in die universitäre bzw. fachhochschulische Ausbildung ein solcher Bestandteil integriert werden, in dem die zukünftigen Lehrkräfte Diagnostizieren professionell erlernen.

Ein weiterer Vorschlag besteht in der Erstellung einer Broschüre für Lehrkräfte, die ein durchschnittlich kognitiv begabtes Kind anschaulich darstellt, um den LehrerInnen Orientierung für eine möglichst faire Einschätzung zu geben. Es sollte der Schwerpunkt darauf gelegt werden, die Vorgehensweise der psychologischen Diagnostik zu erklären. Ohne Informationen, was ein Kind in leistungsbezogenen Aufgaben können muss, um

als durchschnittlich begabt zu gelten, ist es für eine Lehrkraft äußerst schwer, die kognitive Leistungsfähigkeit einzuschätzen.

5 Resümee

Die vorliegende Studie beschäftigt sich mit der diagnostischen Kompetenz von österreichischen VolksschullehrerInnen sowie der Eignung von LehrerInnen-einschätzung, LehrerInnennomination und Peernomination (mit der Möglichkeit der Selbstnomination) zur Identifikation besonders leistungsfähiger Kinder im mathematischen und sprachlichen Bereich. Zu diesem Zwecke wurden 219 SchülerInnen aus elf Klassen in Niederösterreich mit einem standardisierten Intelligenztestverfahren untersucht. Die jeweiligen KlassenlehrerInnen schätzten die Leistungsfähigkeit aller Kinder anhand des Einschätzungsverfahrens ein. Zusätzlich nahmen die Lehrkräfte und die SchülerInnen eine Nomination vor, in welcher sie die jeweils im mathematischen und sprachlichen Bereich besten SchülerInnen nennen sollten. Der Vergleich der Intelligenztestwerte mit den LehrerInneneinschätzungen zeigte einen mittleren Zusammenhang. Außerdem konnte eine Überschätzung des Klassenniveaus und eine leichte Unterschätzung der Streuung der Leistungsfähigkeit innerhalb der Klassen festgestellt werden. Als geeignetstes Instrument zur Identifikation der besonders leistungsfähigen Kinder im mathematischen Bereich stellte sich von den angewendeten Verfahren die LehrerInneneinschätzung heraus. Im sprachlichen Bereich zeigte sich ein Vorteil der Peernomination. Durch eine Kombination verschiedener Methoden konnte eine Erhöhung der korrekt identifizierten besonders Leistungsfähigen erreicht werden. Es wird aufgrund der Ergebnisse dazu geraten, das Thema des Diagnostizierens mehr in die LehrerInnenausbildung zu integrieren.

„Hiermit bestätige ich, dass die vorliegende Arbeit in allen relevanten Teilen selbstständig durchgeführt wurde.“

Wien, den 20. April 2011

(Katharina Klockgether)

Literaturverzeichnis

- Amelang, M., Bartussek, D., Stemmler, G. & Hagemann, D. (2006). *Differentielle Psychologie und Persönlichkeitsforschung*. Stuttgart: Kohlhammer.
- Bortz, J. & Döring, N. (1995). *Forschungsmethoden und Evaluation für Sozialwissenschaftler*. Heidelberg: Springer.
- Brown, S. W., Renzulli, J. S., Chen, C.-H., Gubbins, E. J., Siegle, D. & Zhang, W. (2005). Assumptions underlying the identification of gifted and talented students. *Gifted Child Quarterly*, 49 (1), 68-79.
- Bundesministerium für Unterricht, Kunst und Kultur. (2003). *Lehrplan der Volksschule, Siebenter Teil, Bildungs- und Lehraufgaben sowie Lehrstoff und didaktische Grundsätze der Pflichtgegenstände der Grundschule und der Volksschuloberstufe, Grundschule – Mathematik*. Retrieved from http://www.bmukk.gv.at/medienpool/3996/VS7T_Mathematik.pdf
- Fels, C. (1999). *Identifizierung und Förderung Hochbegabter in den Schulen der Bundesrepublik Deutschland*. Bern: Haupt.
- Gölitz, D., Roick, T. & Hasselhorn, M. (2006). *DEMAT 4. Deutscher Mathematiktest für vierte Klassen*. Göttingen: Hogrefe.
- Hagen, E. (1989). *Die Identifizierung Hochbegabter. Grundlagen der Diagnose außergewöhnlicher Begabungen*. Heidelberg: Asanger.
- Hanses, P. & Rost, D. (1998). Das Drama der hochbegabten Underachiever. *Zeitschrift für Pädagogische Psychologie*, 12 (1), 53-71.
- Heller, K. (Hrsg.) (2001). *Hochbegabung im Kindes- und Jugendalter*. Göttingen: Hogrefe.
- Heller, K. & Perleth, C. (2007). *Münchner Hochbegabungstestbatterie für die Primarstufe (MHBT-P)*. Göttingen: Hogrefe.
- Heller, K. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Weinheim: Beltz.
- Heller, K., Reimann, R. & Senfter, A. (2005). *Hochbegabung im Grundschulalter*. Münster: Lit.
- Helmke, A., Hosenfeld, I. & Schrader, F.-W. (2004). Vergleichsarbeiten als Instrument zur Verbesserung der Diagnosekompetenz von Lehrkräften. In R. A. C. Griesse (Hrsg.), *Schulmanagement und Schulentwicklung* (S. 119-144). Hohengehren: Schneider.
- Hesse, I. & Latzko, B. (2009). *Diagnostik für Lehrkräfte*. Opladen: Budrich.

- Hinnant, B., O'Brien, M. & Ghazarian, S. (2009). The longitudinal relations of teacher expectations to achievement in the early school years. *Journal of Educational Psychology*, 101 (3), 662-670.
- Hoge, R. & Coladarci, T. (1989). Teacher-based judgements of academic achievement: A review of literature. *Review of Educational Research*, 59 (3), 297-313.
- Holling, H. & Kanning, U. P. (1999). *Hochbegabung - Forschungsergebnisse und Fördermöglichkeiten*. Göttingen: Hogrefe.
- Holocher-Ertl, S., Kubinger, K. & Hohensinn, C. (2008). Identifying children who may be cognitively gifted: The gap between practical demands and scientific supply. *Psychology Science Quarterly*, 50 (2), 97-111.
- Hosenfeld, I., Helmke, A. & Schrader, F.-W. (2002). Diagnostische Kompetenz: Unterrichts- und lernrelevante Schülermerkmale und deren Einschätzung durch Lehrkräfte in der Unterrichtsstudie SALVE. In M. P. & J. Doll (Hrsg.), *Bildungsqualität von Schule: Schulische und außerschulische Bedingungen mathematischer, naturwissenschaftlicher und überfachlicher Kompetenzen* (S. 65-82). Weinheim: Beltz.
- Hurwitz, J., Elliott, S. & Braden, J. (2007). The influence of test familiarity and student disability status upon teachers' judgments of students' test performance. *School Psychology Quarterly*, 22 (2), 115-144.
- Huser, J. (1999). *Lichtblick für helle Köpfe. Ein Wegweiser zur Erkennung und Förderung von hohen Fähigkeiten bei Kindern und Jugendlichen aller Schulstufen*. Zürich: Lehrmittelverlag.
- Jäger, A. O. (1984). Intelligenzstrukturforschung: Konkurrierende Modelle, neue Entwicklungen, Perspektiven. *Psychologische Rundschau*, 35 (1), 21-35.
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23 (3-4), 197-209.
- Klauer, K. & Leutner, D. Intelligenz und Begabung (2010). In D. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 304-311). Weinheim: Beltz.
- Kolb, K. & Jussim, L. (1993). Teacher expectations and underachieving gifted children. *Roeper Review*, 17 (1), 26-30.
- Kubinger, K. D. (2006). *Psychologische Diagnostik. Theorie und Praxis psychologischen Diagnostizierens*. Wien: Hogrefe.

- Lang, E. (2008). *LehrerInneneinschätzung, LehrerInnen- und Peernomination: Sind diese für die Identifikation von mathematischer bzw. sprachlicher Begabung hilfreich?* Universität Wien: Diplomarbeit.
- Langfeldt, H.–P. (2006). *Psychologie für die Schule*. Weinheim: Beltz.
- Lenhard, W. & Schneider, W. (2006). *ELFE 1-6. Leseverständnistest für Erst- bis Sechstklässler*. Göttingen: Hogrefe.
- Lorenz, C. & Artelt, C. (2009). Fachspezifität und Stabilität diagnostischer Kompetenz von Grundschullehrkräften in den Fächern Deutsch und Mathematik. *Zeitschrift für Pädagogische Psychologie*, 23 (3-4), 211-222.
- Mühle, G. (1968). Definitions- und Methodenprobleme der Begabtenforschung. In H. Roth (Hrsg.), *Begabung und Lernen* (S. 69-97). Stuttgart: Klett.
- Neber, H. (2004). Teacher identification of students for gifted programs: Nominations to a summer school for highly-gifted students. *Psychology Science Quarterly*, 46 (3), 348–362.
- Nicholls, J. (1978). The development of the concept of effort and ability, perception of academic attainment, and the understanding that difficult tasks require more ability. *Child Development*, 49 (4), 800-814.
- Pädagogische Hochschule Kärnten. (2010). *Curriculum der pädagogischen Hochschule Kärnten*. Retrieved from <http://www.ph-kaernten.ac.at>
- Pädagogische Hochschule Niederösterreich. (2010). *Curriculum der pädagogischen Hochschule Niederösterreich*. Retrieved from www.ph-noe.ac.at
- Pädagogische Hochschule Oberösterreich. (2010). *Curriculum der pädagogischen Hochschule Oberösterreich*. Retrieved from <http://www.ph-ooe.at>
- Pädagogische Hochschule Salzburg. (2010). *Curriculum der pädagogischen Hochschule Salzburg*. Retrieved from <http://www.phsalzburg.at>
- Pädagogische Hochschule Steiermark. (2010). *Curriculum der pädagogischen Hochschule Steiermark*. Retrieved from <http://www.phst.at>
- Pädagogische Hochschule Tirol. (2010). *Curriculum der pädagogischen Hochschule Tirol*. Retrieved from <http://www.ph-tirol.ac.at>
- Pädagogische Hochschule Vorarlberg. (2010). *Curriculum der pädagogischen Hochschule Vorarlberg*. Retrieved from <http://www.ph-vorarlberg.ac.at>
- Pädagogische Hochschule Wien. (2010). *Curriculum der pädagogischen Hochschule Wien*. Retrieved from <http://www.phwien.ac.at>

- Pekrun, R., Frenzel, A. & Götz, T. (2010). Bildungspsychologie des Sekundärbereichs. In C. Spiel, B. Schober, P. Wagner & R. Reimann (Hrsg.). *Bildungspsychologie* (S. 111-130). Göttingen: Hogrefe.
- Raven, J. C., Raven, J. & Court, J. H. (1999). *SPM Manual (Deutsche Bearbeitung und Normierung von S. Bulheller und H. Häcker)*. Frankfurt: Swets & Zeitlinger B.V.
- Rost, D. (1991). Identifizierung von Hochbegabung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 23 (3), 197-231.
- Rost, D., Sparfeldt, J. & Schilling, S. (2006). Hochbegabung. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 187-222). Heidelberg: Springer.
- Schrader, F.-W. (2009). Anmerkungen zum Themenschwerpunkt Diagnostische Kompetenz von Lehrkräften. The diagnostic competence of teachers. *Zeitschrift für Pädagogische Psychologie*, 23 (3-4), 237-245.
- Schrader, F.-W. (2006). *Diagnostische Kompetenz von Eltern und Lehrern*. In D. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (S. 95-100). Weinheim: Beltz.
- Schrader, F.-W. & Helmke, A. (2001). Alltägliche Leistungsbeurteilung durch Lehrer. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 45-58). Weinheim: Beltz.
- Schrader, F.-W. (1989). *Diagnostische Kompetenzen von Lehrern und ihre Bedeutung für die Gestaltung und Effektivität des Unterrichts*. Frankfurt am Main: Peter Lang.
- Schrader, F.-W. & Helmke, A. (1987). Diagnostische Kompetenz von Lehrern: Komponenten und Wirkungen. *Empirische Pädagogik*, 1 (1), 27-52.
- Schweizer, K. (Hrsg.) (2006). *Leistung und Leistungsdiagnostik*. Frankfurt am Main: Springer.
- Seidl, C. (2010, Juni). Umfrage: Wir lernen zu wenig und oft das Falsche. *Der Standard*. Retrieved from <http://derstandard.at/1277336768703/STANDARD-Umfrage-79-Prozent-sagen-Talente-bleiben-in-den-Schulen-ungenutzt>
- Sommer, U., Fink, A. & Neubauer, A. C. (2008). Detection of high ability children by teachers and parents: Psychometric quality of new rating checklists for the assessment of intellectual, creative and social ability. *Psychology Science Quarterly*, 50 (2), 189-205.
- Specht, W. (Hrsg.) (2009). *Nationaler Bildungsbericht Österreich 2009. Band 1: Das Schulsystem im Spiegel von Daten und Indikatoren*. Graz: Leykam.
- Spinath, B. (2010). Bildungspsychologie des Primärbereichs. In C. Spiel, B. Schober, P. Wagner & R. Reimann (Hrsg.). *Bildungspsychologie* (S. 81-100). Göttingen: Hogrefe.

- Spinath, B. (2005). Akkuratheit der Einschätzung von Schülermerkmalen durch Lehrer und das Konstrukt der diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 19 (1/2), 85-95.
- Stern, W. (1916). Psychologische Begabung und Begabungsdiagnose. In P. Petersen (Hrsg.), *Der Aufstieg der Begabten* (S. 105-120). Leipzig: G. Teubner.
- Stöger, H. (2003). Wie gut stimmen Begabungseinschätzungen von Eltern, Lehrkräften und Schülern mit Intelligenztestmessungen überein? *LVH aktuell*, 10, 5-9.
- Stryck, T. (2005). *VERA 2004 – VERgleichsArbeiten in Jahrgangsstufe 4*. Berlin: Senatsverwaltung für Bildung, Jugend und Sport.
- Südkamp, A., Pohlmann, B. & Möller, J. (2008). Der Simulierte Klassenraum. Eine experimentelle Untersuchung zur diagnostischen Kompetenz. *Zeitschrift für Pädagogische Psychologie*, 22 (3-4), 261-276.
- Thorndike, R. L. & Hagen, E. P. (1971). *Cognitive Abilities Test*. Boston: Houghton-Mifflin.
- Urban, K. K. (1992). Begabungsförderung im Vorschulalter. In A. H. H. Nickel (Hrsg.), *Begabung und Hochbegabung* (S. 159-169). Bern: Hans Huber.
- van Ophuysen, S. (2006). Vergleich diagnostischer Entscheidungen von Novizen und Experten am Beispiel der Schullaufbahnenempfehlung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 38 (4), 154-161.
- Wahl, D., Weinert, F. E. & Huber, G. (1997). *Psychologie für die Schulpraxis*. München: Kösel.
- Weiß, R. H. (2006). *Grundintelligenztest Skala 2 – Revision*. Göttingen: Hogrefe.
- Wild, K.-P. (1991). *Identifikation hochbegabter Schüler - Lehrer und Schüler als Datenquellen*. Heidelberg: Asanger.
- Wild, K.-P. & Rost, D. (1995). Klassengröße und Genauigkeit von Schülerbeurteilungen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 27 (1), 78-90.
- Ziegler, A. & Stöger, H. (2003). Identification of Underachievement: An empirical study on the agreement among various diagnostic sources. *Gifted and Talented International*, 18 (2), 87-94.
- Zöfel, P. (2003). *Statistik für Psychologen*. München: Pearson Studium.

Abbildungsverzeichnis

Abbildung 1: Verteilung der Quantitativtestwerte und Verbaltestwerte, Stichprobe NÖ (Niederösterreich) . N = 219 und Stichprobe OÖ (Oberösterreich) aus Lang (2008). N = 233.....	31
Abbildung 2: Ausschnitt des Antwortbogens des KFT 4-12+ R.....	34

Tabellenverzeichnis

Tabelle 1: Statistische Kennwerte der Stichprobe. N = 219.....	42
Tabelle 2: Kreuztabelle der LehrerInneneinschätzung im mathematischen Bereich und der Testergebnisse im mathematischen Bereich in Prozenträngen, gestaffelt in fünf Intervalle. N= 219.....	43
Tabelle 3: Kreuztabelle der LehrerInneneinschätzung im sprachlichen Bereich und der Testergebnisse im sprachlichen Bereich in Prozenträngen, gestaffelt in fünf Intervalle. N= 219.....	44
Tabelle 4: Mittelwerte der LehrerInneneinschätzung (Mathe), Mittelwerte der Testergebnisse (Mathe) und Niveauelemente pro Lehrkraft bzw. Klasse (N bis X). N= 219.....	45
Tabelle 5: Mittelwerte der LehrerInneneinschätzung (Sprache), Mittelwerte der Testergebnisse (Sprache) und Niveauelemente pro Lehrkraft bzw. Klasse (N bis X). N= 219.....	46
Tabelle 6: Standardabweichungen der LehrerInneneinschätzung (Mathe), Standardabweichungen der Testergebnisse (Mathe) und Differenzierungskomponente pro Lehrkraft bzw. Klasse (N bis X). N= 219.....	46
Tabelle 7: Standardabweichungen der LehrerInneneinschätzung (Sprache), Standardabweichungen der Testergebnisse (Sprache) und Differenzierungskomponente pro Lehrkraft bzw. Klasse (N bis X). N= 219.....	47
Tabelle 8: Vergleichskomponenten des mathematischen Bereichs pro Lehrkraft. N = 219.....	48

Tabelle 9: Vergleichskomponenten des sprachlichen Bereichs pro Lehrkraft. N = 219.....	48
Tabelle 10: Verteilungen der Rangplätze, Effektivität und Effizienz der Methoden im mathematischen Bereich. N= 219.....	51
Tabelle 11: Verteilungen der Rangplätze, Effektivität und Effizienz der Methoden im sprachlichen Bereich. N= 219.....	52
Tabelle 12: Effektivität und Effizienz der verschiedenen Methoden in Kombination, bezogen auf den mathematischen Bereich. N = 219.....	53
Tabelle 13: Effektivität und Effizienz der verschiedenen Methoden in Kombination, bezogen auf den sprachlichen Bereich, Frage 2. N = 219.....	54
Tabelle 14: Effektivität und Effizienz der verschiedenen Methoden in Kombination, bezogen auf den sprachlichen Bereich, Frage 3. N = 219.....	54
Tabelle 15: Effektivität und Effizienz der verschiedenen Methoden bezogen auf die Kinder mit PR > 85 im mathematischen Testteil. N= 219.....	56
Tabelle 16: Effektivität und Effizienz der verschiedenen Methoden bezogen auf die Kinder mit PR > 85 im sprachlichen Testteil. N= 219.....	58
Tabelle 17: Effektivität und Effizienz der verschiedenen Methoden bezogen auf die Kinder mit PR > 85 im sprachlichen Testteil. N= 219.....	58
Tabelle 18: Mehrfaktorielle ANOVA zur Überprüfung der Moderatoreffekte. Niveauekomponente mathematischer Bereich. N= 11.....	60
Tabelle 19: Mehrfaktorielle ANOVA zur Überprüfung der Moderatoreffekte. Niveauekomponente sprachlicher Bereich. N= 11.....	61
Tabelle 20: Mehrfaktorielle ANOVA zur Überprüfung der Moderatoreffekte. Differenzierungskomponente mathematischer Bereich. N= 11.....	61
Tabelle 21: Mehrfaktorielle ANOVA zur Überprüfung der Moderatoreffekte. Differenzierungskomponente sprachlicher Bereich. N= 11.....	61
Tabelle 22: Mehrfaktorielle ANOVA zur Überprüfung der Moderatoreffekte. Vergleichskomponente mathematischer Bereich. N= 11.....	62
Tabelle 23: Mehrfaktorielle ANOVA zur Überprüfung der Moderatoreffekte. Vergleichskomponente sprachlicher Bereich. N= 11.....	62

Tabelle 24: Prüfung der Normalverteilung der VPR, LEM, LES und QPR mit dem Kolmogorov-Smirnov-Anpassungstest. N = 219.....	86
Tabelle 25: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse N mit dem Kolmogorov-Smirnov-Anpassungstest. N = 21.....	86
Tabelle 26: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse O mit dem Kolmogorov-Smirnov-Anpassungstest. N = 20.....	86
Tabelle 27: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse P mit dem Kolmogorov-Smirnov-Anpassungstest. N = 16.....	86
Tabelle 28: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse Q mit dem Kolmogorov-Smirnov-Anpassungstest. N = 20.....	87
Tabelle 29: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse R mit dem Kolmogorov-Smirnov-Anpassungstest. N = 23.....	87
Tabelle 30: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse S mit dem Kolmogorov-Smirnov-Anpassungstest. N = 14.....	87
Tabelle 31: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse T mit dem Kolmogorov-Smirnov-Anpassungstest. N = 22.....	87
Tabelle 32: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse U mit dem Kolmogorov-Smirnov-Anpassungstest. N = 24.....	87
Tabelle 33: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse V mit dem Kolmogorov-Smirnov-Anpassungstest. N = 21.....	88
Tabelle 34: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse W mit dem Kolmogorov-Smirnov-Anpassungstest. N = 18.....	88
Tabelle 35: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse X mit dem Kolmogorov-Smirnov-Anpassungstest. N = 18.....	88
Tabelle 36: Prüfung der Normalverteilung der Niveauelemente im sprachlichen und mathematischen Bereich mit dem Kolmogorov-Smirnov-Anpassungstest. N = 11.....	88
Tabelle 37: Prüfung der Normalverteilung der Differenzierungselemente im sprachlichen und mathematischen Bereich mit dem Kolmogorov-Smirnov-Anpassungstest. N = 11.....	88

Tabelle 38: Prüfung der Normalverteilung der Vergleichskomponenten im sprachlichen und mathematischen Bereich mit dem Kolmogorov-Smirnov-Anpassungstest. N = 11.....	89
Tabelle 39: Deskriptive Statistik der Verbaltestergebnisse der einzelnen Klassen (kodiert mit N bis X). N= 219.....	90
Tabelle 40: Deskriptive Statistik der Quantitativtestergebnisse der einzelnen Klassen (kodiert mit N bis X). N= 219.....	90
Tabelle 41: Deskriptive Statistik der mathematischen LehrerInneneinschätzung der einzelnen Klassen (kodiert mit N bis X). N= 219.....	90
Tabelle 42: Deskriptive Statistik der sprachlichen LehrerInneneinschätzung der einzelnen Klassen (kodiert mit N bis X). N= 219.....	90
Tabelle 43: Prüfung der Homogenität der Varianzen der Niveauelementen mit Levene-Test. N = 11.....	91
Tabelle 44: Prüfung der Homogenität der Varianzen der Differenzierungskomponenten mit Levene-Test. N = 11.....	91
Tabelle 45: Prüfung der Homogenität der Varianzen der Vergleichskomponenten mit Levene-Test. N = 11.....	91

Anhang

Anhang 1: Verwendete Abkürzungen

Anhang 2: Prüfung der Normalverteilungen

Anhang 3: Deskriptive Ergebnisse der Klassen

Anhang 4: Prüfung der Voraussetzungen der ANOVA

Anhang 5: Schreiben an den Landesschulrat Niederösterreich

Anhang 6: Elternbrief

Anhang 7: LehrerInnenfragebogen

Anhang 8: Anleitung zur Durchführung der Peernomination

Anhang 9: SchülerInnenfragebogen

Anhang 1: Verwendete Abkürzungen

df =	Freiheitsgrade
DK =	Differenzierungskomponente
HB =	Hochbegabung
IQ =	Intelligenzquotient
KFT 4-12+ R =	Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision
LEM =	LehrerInneneinschätzung der mathematischen Leistungsfähigkeit der Kinder in PR
LES =	LehrerInneneinschätzung der sprachlichen Leistungsfähigkeit der Kinder in PR
LK =	Lehrkraft
M =	Mittelwert
N/n =	Stichprobenumfang
NK =	Niveauelemente
QPR =	Ergebnisse der Kinder im quantitativen/mathematischen Teil des KFT 4-12+ R in Prozentangaben
r =	Korrelationskoeffizient
SD =	Standardabweichung
U =	Unterschätzung
Ü =	Überschätzung
VK =	Vergleichskomponente
VPR =	Ergebnisse der Kinder im verbalen/sprachlichen Teil des KFT 4-12+ R in Prozentangaben

Anhang 2: Prüfung der Normalverteilungen

Tab. 24: Prüfung der Normalverteilung der VPR, LEM, LES und QPR mit dem Kolmogorov-Smirnov-Anpassungstest. N = 219

	VPR	LEM	LES	QPR
M	50,34	64,39	60,81	54,71
SD	28,17	20,62	20,60	30,81
K-S-Z	1,363	1,186	1,231	1,745
Sig. 2-seitig	.049	.120	.097	.005

Legende: VPR = Verbaltestergebnisse, LEM = mathematische LehrerInneneinschätzung, LES = sprachliche LehrerInneneinschätzung, QPR = Quantitativtestergebnisse, M = Mittelwert, SD = Standardabweichung, K-S-Z = Kolmogorov-Smirnov Prüfgröße Z, Sig. = Signifikanz.

Prüfung der Normalverteilung der VPR, LEM, LES und QPR der einzelnen Klassen

Tab. 25: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse N mit dem Kolmogorov-Smirnov-Anpassungstest. N = 21

	VPR	LEM	LES	QPR
M	49,57	70,71	56,90	65,88
SD	27,64	18,66	20,34	26,83
K-S-Z	.619	.962	.600	.592
Sig. 2-seitig	.838	.313	.865	.874

Legende: VPR = Verbaltestergebnisse, LEM = mathematische LehrerInneneinschätzung, LES = sprachliche LehrerInneneinschätzung, QPR = Quantitativtestergebnisse, M = Mittelwert, SD = Standardabweichung, K-S-Z = Kolmogorov-Smirnov Prüfgröße Z, Sig. = Signifikanz.

Tab. 26: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse O mit dem Kolmogorov-Smirnov-Anpassungstest. N = 20

	VPR	LEM	LES	QPR
M	63,50	60,75	55,15	63,43
SD	24,36	13,60	12,61	29,90
K-S-Z	.717	.903	.893	.782
Sig. 2-seitig	.683	.389	.402	.574

Legende: VPR = Verbaltestergebnisse, LEM = mathematische LehrerInneneinschätzung, LES = sprachliche LehrerInneneinschätzung, QPR = Quantitativtestergebnisse, M = Mittelwert, SD = Standardabweichung, K-S-Z = Kolmogorov-Smirnov Prüfgröße Z, Sig. = Signifikanz.

Tab. 27: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse P mit dem Kolmogorov-Smirnov-Anpassungstest. N = 16

	VPR	LEM	LES	QPR
M	69,72	64,25	58,19	72,10
SD	18,81	16,85	15,92	26,87
K-S-Z	.757	.774	.609	.641
Sig. 2-seitig	.616	.587	.852	.806

Legende: VPR = Verbaltestergebnisse, LEM = mathematische LehrerInneneinschätzung, LES = sprachliche LehrerInneneinschätzung, QPR = Quantitativtestergebnisse, M = Mittelwert, SD = Standardabweichung, K-S-Z = Kolmogorov-Smirnov Prüfgröße Z, Sig. = Signifikanz.

Tab. 28: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse Q mit dem Kolmogorov-Smirnov-Anpassungstest. N = 20

	VPR	LEM	LES	QPR
M	47,76	59,30	55,35	52,15
SD	24,44	18,43	16,35	34,02
K-S-Z	.543	.808	.617	.708
Sig. 2-seitig	.930	.531	.840	.697

Legende: VPR = Verbaltestergebnisse, LEM = mathematische LehrerInneneinschätzung, LES = sprachliche LehrerInneneinschätzung, QPR = Quantitativtestergebnisse, M = Mittelwert, SD = Standardabweichung, K-S-Z = Kolmogorov-Smirnov Prüfgröße Z, Sig. = Signifikanz.

Tab. 29: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse R mit dem Kolmogorov-Smirnov-Anpassungstest. N = 23

	VPR	LEM	LES	QPR
M	34,26	68,61	67,26	47,73
SD	26,40	22,61	22,88	30,24
K-S-Z	1,004	.730	.541	.695
Sig. 2-seitig	.266	.660	.931	.720

Legende: VPR = Verbaltestergebnisse, LEM = mathematische LehrerInneneinschätzung, LES = sprachliche LehrerInneneinschätzung, QPR = Quantitativtestergebnisse, M = Mittelwert, SD = Standardabweichung, K-S-Z = Kolmogorov-Smirnov Prüfgröße Z, Sig. = Signifikanz.

Tab. 30: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse S mit dem Kolmogorov-Smirnov-Anpassungstest. N = 14

	VPR	LEM	LES	QPR
M	31,93	54,29	55,00	41,64
SD	23,35	30,06	26,75	31,32
K-S-Z	.674	.630	.812	.564
Sig. 2-seitig	.754	.822	.525	.908

Legende: VPR = Verbaltestergebnisse, LEM = mathematische LehrerInneneinschätzung, LES = sprachliche LehrerInneneinschätzung, QPR = Quantitativtestergebnisse, M = Mittelwert, SD = Standardabweichung, K-S-Z = Kolmogorov-Smirnov Prüfgröße Z, Sig. = Signifikanz.

Tab. 31: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse T mit dem Kolmogorov-Smirnov-Anpassungstest. N = 22

	VPR	LEM	LES	QPR
M	38,10	60,86	63,45	44,91
SD	31,10	19,60	19,53	31,50
K-S-Z	1,098	.722	.646	.621
Sig. 2-seitig	.179	.674	.797	.836

Legende: VPR = Verbaltestergebnisse, LEM = mathematische LehrerInneneinschätzung, LES = sprachliche LehrerInneneinschätzung, QPR = Quantitativtestergebnisse, M = Mittelwert, SD = Standardabweichung, K-S-Z = Kolmogorov-Smirnov Prüfgröße Z, Sig. = Signifikanz.

Tab. 32: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse U mit dem Kolmogorov-Smirnov-Anpassungstest. N = 24

	VPR	LEM	LES	QPR
M	76,69	62,25	57,17	73,88
SD	15,42	17,92	19,67	19,02
K-S-Z	.935	.718	.844	.866
Sig. 2-seitig	.346	.682	.474	.442

Legende: VPR = Verbaltestergebnisse, LEM = mathematische LehrerInneneinschätzung, LES = sprachliche LehrerInneneinschätzung, QPR = Quantitativtestergebnisse, M = Mittelwert, SD = Standardabweichung, K-S-Z = Kolmogorov-Smirnov Prüfgröße Z, Sig. = Signifikanz.

Tab. 33: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse V mit dem Kolmogorov-Smirnov-Anpassungstest. N = 21

	VPR	LEM	LES	QPR
M	65,67	74,76	72,76	60,86
SD	21,45	22,61	20,44	30,91
K-S-Z	.958	1,402	1,182	.683
Sig. 2-seitig	.318	.039	.122	.739

Legende: VPR = Verbaltestergebnisse, LEM = mathematische LehrerInneneinschätzung, LES = sprachliche LehrerInneneinschätzung, QPR = Quantitativtestergebnisse, M = Mittelwert, SD = Standardabweichung, K-S-Z = Kolmogorov-Smirnov Prüfgröße Z, Sig. = Signifikanz.

Tab. 34: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse W mit dem Kolmogorov-Smirnov-Anpassungstest. N = 18

	VPR	LEM	LES	QPR
M	34,25	74,50	74,11	42,56
SD	21,09	18,58	17,95	22,51
K-S-Z	.684	.988	.794	.576
Sig. 2-seitig	.737	.283	.554	.894

Legende: VPR = Verbaltestergebnisse, LEM = mathematische LehrerInneneinschätzung, LES = sprachliche LehrerInneneinschätzung, QPR = Quantitativtestergebnisse, M = Mittelwert, SD = Standardabweichung, K-S-Z = Kolmogorov-Smirnov Prüfgröße Z, Sig. = Signifikanz.

Tab. 35: Prüfung der Normalverteilung der VPR, LEM, LES und QPR innerhalb der Klasse X mit dem Kolmogorov-Smirnov-Anpassungstest. N = 18

	VPR	LEM	LES	QPR
M	36,68	55,25	51,75	32,30
SD	24,27	19,70	22,20	29,00
K-S-Z	.797	.625	.484	.951
Sig. 2-seitig	.549	.830	.973	.326

Legende: VPR = Verbaltestergebnisse, LEM = mathematische LehrerInneneinschätzung, LES = sprachliche LehrerInneneinschätzung, QPR = Quantitativtestergebnisse, M = Mittelwert, SD = Standardabweichung, K-S-Z = Kolmogorov-Smirnov Prüfgröße Z, Sig. = Signifikanz.

Prüfung der Normalverteilung der Komponenten

Tab. 36: Prüfung der Normalverteilung der Niveauelemente im sprachlichen und mathematischen Bereich mit dem Kolmogorov-Smirnov-Anpassungstest. N = 11

	NK Sprache	NK Mathematik
M	10,81	9,02
SD	18,81	14,81
K-S-Z	.494	.472
Sig. 2-seitig	.968	.979

Legende: NK = Niveauelemente, M = Mittelwert, SD = Standardabweichung, K-S-Z = Kolmogorov-Smirnov Prüfgröße Z, Sig. = Signifikanz.

Tab. 37: Prüfung der Normalverteilung der Differenzierungskomponenten im sprachlichen und mathematischen Bereich mit dem Kolmogorov-Smirnov-Anpassungstest. N = 11

	DK Sprache	DK Mathematik
M	0,86	0,71
SD	0,22	0,16
K-S-Z	.512	.440
Sig. 2-seitig	.956	.990

Legende: VK = Differenzierungskomponente, M = Mittelwert, SD = Standardabweichung, K-S-Z = Kolmogorov-Smirnov Prüfgröße Z, Sig. = Signifikanz.

Tab. 38: Prüfung der Normalverteilung der Vergleichskomponenten im sprachlichen und mathematischen Bereich mit dem Kolmogorov-Smirnov-Anpassungstest. N = 11

	VK Sprache	VK Mathematik
M	0,59	0,60
SD	0,21	0,15
K-S-Z	.922	.445
Sig. 2-seitig	.363	.989

Legende: VK = Vergleichskomponente, M = Mittelwert, SD = Standardabweichung, K-S-Z = Kolmogorov-Smirnov Prüfgröße Z, Sig. = Signifikanz.

Anhang 3: Deskriptive Ergebnisse der Klassen

Tab. 39: Deskriptive Statistik der Verbaltestergebnisse der einzelnen Klassen (kodiert mit N bis X). N= 219

Klasse	N	O	P	Q	R	S	T	U	V	W	X
n	21	20	16	20	23	14	22	24	21	18	20
M	49,57	63,50	69,72	47,76	34,26	31,93	38,10	76,69	65,67	34,25	36,68
SD	27,64	24,36	18,81	24,44	26,40	23,35	31,10	15,42	21,45	21,10	24,27
Min	2	21	34	0,1	3	2	0,1	50	21	1,5	1,5
Max	99	98	96,5	93	80	81	100	100	93	69	88

Legende: n = Stichprobenumfang, M = Mittelwert, SD = Standardabweichung, Min = minimale Ausprägung, Max = maximale Ausprägung.

Tab. 40: Deskriptive Statistik der Quantitativtestergebnisse der einzelnen Klassen (kodiert mit N bis X). N= 219

Klasse	N	O	P	Q	R	S	T	U	V	W	X
n	21	20	16	20	23	14	22	24	21	18	20
M	65,88	63,43	72,10	52,15	47,74	41,64	44,91	73,88	60,86	42,56	32,30
SD	26,83	29,90	26,88	34,02	30,24	31,32	31,50	19,02	30,91	22,51	29,00
Min	16	12	13	2	3	2	0,6	34	10	4	2
Max	99	99	100	98	99	86	98	100	99	88	90

Legende: n = Stichprobenumfang, M = Mittelwert, SD = Standardabweichung, Min = minimale Ausprägung, Max = maximale Ausprägung.

Tab. 41: Deskriptive Statistik der mathematischen LehrerInneneinschätzung der einzelnen Klassen (kodiert mit N bis X). N= 219

Klasse	N	O	P	Q	R	S	T	U	V	W	X
N	21	20	16	20	23	14	22	24	21	18	20
M	70,71	60,75	64,25	59,30	68,61	54,29	60,86	62,25	74,76	74,50	55,25
SD	18,66	13,60	16,85	18,43	22,61	30,06	19,60	17,92	22,61	18,58	19,70
Min	30	40	44	28	30	5	30	18	30	25	20
Max	95	90	90	80	100	95	98	90	100	98	90

Legende: n = Stichprobenumfang, M = Mittelwert, SD = Standardabweichung, Min = minimale Ausprägung, Max = maximale Ausprägung.

Tab. 42: Deskriptive Statistik der sprachlichen LehrerInneneinschätzung der einzelnen Klassen (kodiert mit N bis X). N= 219

Klasse	N	O	P	Q	R	S	T	U	V	W	X
N	21	20	16	20	23	14	22	24	21	18	20
M	56,90	55,15	58,19	55,35	67,26	55,00	63,45	57,17	72,76	74,11	51,75
SD	20,34	12,61	15,92	16,35	22,61	26,75	19,53	19,67	20,44	17,95	22,20
Min	25	38	37	30	30	10	30	12	30	25	10
Max	95	70	80	85	100	95	95	90	95	95	90

Legende: n = Stichprobenumfang, M = Mittelwert, SD = Standardabweichung, Min = minimale Ausprägung, Max = maximale Ausprägung.

Anhang 4: Prüfung der Voraussetzungen der ANOVA

Tab. 43: Prüfung der Homogenität der Varianzen der Differenzierungskomponenten mit Levene-Test. N = 11

	F	df1	df2	Sig.
DK Mathematik	1,742	5	5	.279
DK Sprache	1,305	5	5	.389

Signifikanzniveau $p < .05$

Legende: DK = Differenzierungskomponente, df = Freiheitsgrade, Sig. = Signifikanz.

Tab. 44: Prüfung der Homogenität der Varianzen der Niveauelementen mit Levene-Test. N = 11

	F	df1	df2	Sig.
NK Mathematik	1,442	5	5	.349
NK Sprache	1,366	5	5	.370

Signifikanzniveau $p < .05$

Legende: NK = Niveauelement, df = Freiheitsgrade, Sig. = Signifikanz.

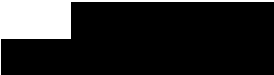
Tab. 45: Prüfung der Homogenität der Varianzen der Vergleichskomponenten mit Levene-Test. N = 11

	F	df1	df2	Sig.
VK Mathematik	1,506	5	5	.332
VK Sprache	2,451	5	5	.174

Signifikanzniveau $p < .05$

Legende: VK = Vergleichskomponente, df = Freiheitsgrade, Sig. = Signifikanz.

Katharina Klockgether
Lustgasse 10/5
1030 Wien



Mag. Christina Unterberger
Landesschulrat für Niederösterreich
Schulpsychologie-Bildungsberatung
Landesreferat

Rennbahnstraße 29
3109 St. Pölten

Wien, den 23. März 2010

Betrifft:
Ansuchen um Genehmigung einer empirischen Untersuchung

Sehr geehrte Frau Mag. Unterberger,

im Rahmen meiner Diplomarbeit an der Fakultät für Psychologie der Universität Wien zum Thema Begabungseinschätzung von SchülerInnen durch LehrerInnen und MitschülerInnen, möchte ich gerne Erhebungen an niederösterreichischen Volksschulen vornehmen.

Die Studie soll im Wesentlichen als eine Replikation bzw. eine Stichprobenerweiterung einer im November und Dezember 2007 an oberösterreichischen Volksschulen durchgeführten Diplomstudie fungieren. Die Erhebung im Jahre 2007, welche vom LSR Oberösterreich genehmigt wurde, ist damals ohne Komplikationen durchgeführt worden und soll nun möglichst ähnlich in Niederösterreich wiederholt werden, um generalisierbarere Schlüsse aus den Ergebnissen ziehen zu können.

Die Arbeit soll zur Klärung der Frage beitragen, wie gut SchülerInnen die hohen Leistungsfähigkeiten ihrer MitschülerInnen im Vergleich zu den LehrerInnen einschätzen können und inwieweit diese Einschätzungen mit den Ergebnissen eines standardisierten Leistungstest übereinstimmen. Dies ist z. B. in Hinblick auf individuelle Förderungen besonders wichtig. Es ist außerdem beabsichtigt, die Studie zu veröffentlichen.

Im Umfang von ca. 1,5 Schulstunden soll mit SchülerInnen der 4. Klassen der KFT 4-12+ R in einer Kurzform (ca. 50 min), die sich aus einem verbalen Teil (Wortschatz-Test und Wortanalogien-Test) sowie einem quantitativen Teil (Mengenvergleiche und Zahlenreihen) zusammensetzt, als standardisierter Leistungstest durchgeführt werden. Außerdem wird den Kindern ein Fragebogen vorgegeben, auf dem sie die jeweils

sprachlich und mathematisch begabtesten Kinder der Klasse aufschreiben sollen. Selbstverständlich stehen dabei die Bedürfnisse der Kinder im Vordergrund!

Oben genannter Fragebogen soll ebenso von den LehrerInnen ausgefüllt werden. Zusätzlich sollen diese alle anwesenden Kinder der Klasse auf einer Skala von 1 bis 100 in ihrer jeweiligen sprachlichen und mathematischen Leistungsfähigkeit einschätzen. Weiterhin werden an die Lehrkraft einige wenige Fragen zu ihrer Ausbildung gestellt. Ich stehe mit den DirektorInnen und LehrerInnen der Volksschulen [REDACTED]

[REDACTED] in Kontakt und habe von ihren Seiten eine Zusage zur Genehmigung der Durchführung der Studie an ihren Schulen.

Ich ersuche hiermit um die Genehmigung der Durchführung dieser Studie an den Volksschulen [REDACTED]
[REDACTED] durch den LSR Niederösterreich.

Mit freundlichen Grüßen und herzlichem Dank,
Katharina Klockgether

Link zur Beschreibung des KFT 4-12+ R:
<http://www.testzentrale.de/?id=692&mod=detail>

Anbei:
-Schreiben an die Eltern
-Fragebogen für die Kinder (peerfb)
-Fragebögen für die LehrerInnen (lehrerfb)

Wien, den 12. Mai 2010

Liebe Eltern,

unterstützt durch die Klassenlehrerin Ihres Kindes würde ich sehr gerne im Juni einen Teil meiner Diplomarbeitsstudie (Universität Wien) in der Klasse Ihres Kindes durchführen. In dieser Studie geht es darum herauszufinden, wie gut Lehrkräfte, Schüler und Schülerinnen selbst die Begabung von Schülern und Schülerinnen einschätzen können. Dieses Thema ist wichtig, um z.B. in Zukunft Lehrer/innen bei der Einschätzung von Begabung besser zu unterstützen und damit Kinder individueller fördern zu können. Konkret werden die Kinder einige mathematische und sprachliche Aufgaben bearbeiten (die schon häufig bei Kindern in dieser Altersstufe eingesetzt wurden). Außerdem werden sie gebeten einen kurzen Fragebogen auszufüllen. Beides zusammen wird etwas mehr als eine Schulstunde dauern und wird in enger Kooperation mit der Lehrerin durchgeführt.

Die Ergebnisse werden streng vertraulich behandelt, nur anonymisiert ausgewertet und auch nicht an die Lehrerin oder die Schule weitergegeben. Die Namen der Kinder müssen zwecks Zuordnung der Daten zunächst mit erfasst werden, werden allerdings ausschließlich von der Testleiterin eingesehen.

Für diese Studie ist es sehr wichtig, dass möglichst alle Kinder teilnehmen, da die Klasse sozusagen als eine Durchschnittsklasse für vierte Klassen in Österreich stehen soll. (Es werden auch noch Kinder anderer Schulen untersucht).

Falls irgendwelche Unklarheiten oder Bedenken bestehen, zögern Sie nicht, mir ein e-mail zu schreiben oder mich anzurufen.

Mit freundlichen Grüßen und herzlichem Dank für Ihre Unterstützung!

Katharina Klockgether



Den Abschnitt bitte bis Donnerstag, den 27.05.2010 ausgefüllt Ihrem Kind in die Schule mitgeben.

Ich erkläre mich

☐ einverstanden

☐ nicht einverstanden

dass mein Sohn/ meine Tochter

(Nachname, Vorname)

an der Studie teilnimmt.

(Unterschrift des Erziehungsberechtigten)

Sonstige Bemerkungen:

Sehr geehrte Kollegin,

Vielen Dank, dass Sie bereit sind, mich bei meiner Diplomarbeit zu unterstützen. Im Rahmen der Studie soll geklärt werden, (1) wie gut SchülerInnen die hohen Leistungsfähigkeiten ihrer MitschülerInnen einschätzen können im Vergleich zu Lehrkräften und (2) wie weit diese Beurteilungen mit den Ergebnissen eines standardisierten Leistungstest übereinstimmen. Die gewonnenen Erkenntnisse sind v.a. in Hinblick auf individuelle Förderungen besonders hilfreich. Eine Genehmigung des Landesschulrats für Niederösterreich liegt vor.

Im Folgenden bitte ich Sie zunächst die genaue Bezeichnung der Volksschule und einige Daten zur Person einzutragen und anschließend einen kurzen Fragebogen sowie einen Einschätzungsbogen auszufüllen. Ihre SchülerInnen werden einen Test mit mathematischen und sprachlichen Aufgaben bearbeiten. Danach werde ich den Kindern ebenfalls einen Fragebogen vorgeben, der dem ähnelt, den Sie bekommen.

Selbstverständlich werden alle erhobenen Daten und Ergebnisse streng vertraulich unter Wahrung des Datenschutzes behandelt. Bitte beachten Sie, dass keine individuellen Rückmeldungen erfolgen, sondern die Ergebnisse ausschließlich Forschungszwecken dienen.

Mit freundlichen Grüßen und herzlichem Dank,

Katharina Klockgether

Name der Volksschule: _____ Klasse: _____

Geschlecht: ☐ weiblich ☐ männlich

Seit wann unterrichten Sie diese Klasse? Seit der ☐ 1. Klasse ☐ 2. Kl. ☐ 3. Kl. ☐ 4. Kl.

Seit wie vielen Jahren üben Sie den Beruf der VS-Lehrerin aus? Seit _____ Jahr/en

Haben Sie Fortbildungen im Bereich der Hochbegabung besucht?

☐ Nein ☐ Ja, im Ausmaß von ca. _____ Fortbildungstagen.

Haben Sie im Rahmen Ihrer Ausbildung zum Thema Hochbegabung etwas gehört?

gar nichts ☐ ☐ ☐ ☐ ☐ sehr viel

Haben Sie sich privat näher mit dem Thema Hochbegabung beschäftigt? ☐ Nein ☐ Ja

Wenn ja (in Ausbildung oder privat), geben Sie bitte kurz an, mit was genau und in welchem Rahmen Sie sich mit dem Thema beschäftigt haben:

Bitte beantworten Sie die folgenden Fragen:

Schreiben Sie dazu **jeweils drei** SchülerInnen mit dem **Vornamen** auf. Sie können dieselben Schüler bei mehreren Fragen nennen.

Welche drei Kinder aus Ihrer Klasse, **die an dieser Untersuchung teilnehmen**, können am besten **rechnen**?

1. _____
2. _____
3. _____

Welche drei Kinder aus Ihrer Klasse, die an dieser Untersuchung teilnehmen, können am besten **interessante Geschichten erzählen/schreiben**?

1. _____
2. _____
3. _____

Welche drei Kinder aus Ihrer Klasse, die an dieser Untersuchung teilnehmen, können sich am besten **ausdrücken** (Wortschatz)?

1. _____
2. _____
3. _____

Falls Sie der Meinung sind, dass angeführte SchülerInnen in jenem Bereich **herausragend** sind, kennzeichnen Sie diese bitte mit einem **Stern (*)** hinter dem Namen.

Im Folgenden werden Sie gebeten, alle anwesenden SchülerInnen hinsichtlich ihrer/seiner sprachlichen Fähigkeit und anschließend hinsichtlich ihrer/seiner mathematischen Fähigkeit einzuschätzen.

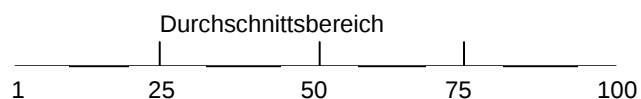
Was wird unter sprachlicher und mathematischer Fähigkeit verstanden?

Bei der Einschätzung der „sprachlichen Fähigkeit“ berücksichtigen Sie bitte den Wortschatz, die Ausdrucksgewandtheit, die richtige Anwendung der Grammatik, die Fremdspracheneignung etc.

Bei der Einschätzung der „mathematischen Fähigkeit“ denken Sie bitte an Fähigkeiten der Mengenerfassung und –differenzierung, den Umgang mit den Grundrechenarten, das schlussfolgernde Denkvermögen etc..

Nach welcher Methode wird die Fähigkeit eingeschätzt?

Stellen Sie sich bitte eine Menge von **100 SchülerInnen** vor, die **die 4. Schulstufe** besuchen und **das gesamte Leistungsspektrum abdecken (Skala von 1 bis 100)**.



Diese Schülerinnen sind nach ihrer Fähigkeit der Reihe nach aufgestellt (1 = SchülerIn mit der **geringsten** Fähigkeit; 100 = SchülerIn mit der **höchst möglichen** Fähigkeit). Überlegen Sie nun für **jede/jeden Ihrer SchülerInnen** mit welcher/welchem der 100 SchülerInnen, diese/dieser hinsichtlich der Fähigkeit vergleichbar wäre.

Beispiel:

Scheint Ihnen einer Ihrer Schüler hinsichtlich seiner Fähigkeit mit einem typischen Schüler auf Position 55 vergleichbar, dann schreiben Sie die Positionsnummer 55 auf. Dies bedeutet, dass 54 typische SchülerInnen der 4. Schulstufe eine schlechtere und 54 SchülerInnen eine bessere Fähigkeit mit sich bringen als Ihr Schüler.

Wichtig: Die Einschätzung der Fähigkeit der SchülerInnen soll an der **Menge typischer SchülerInnen der 4. Klasse** gemessen werden. Dies bedeutet **nicht**, dass Sie die SchülerInnen Ihrer derzeitigen 4. Klasse **in eine interne Rangreihe** bringen sollen. Sie können durchaus an **mehrere SchülerInnen die gleiche Positionsnummer** vergeben.

Bitte schätzen Sie nun jede/n anwesende/n SchülerIn betreffend der **sprachlichen Fähigkeit** ein.

[illegible]

Bitte schätzen Sie nun jede/n anwesende/n SchülerIn betreffend der **mathematischen Fähigkeit** ein.

[illegible]

Anhang 8: Anleitung zur Durchführung der Peernomination

Verteilen der Zettel

TL (Testleiterin): „Jeder von euch hat besondere Fähigkeiten, das heißt, jeder kann etwas besonders gut. Wahrscheinlich habt ihr das schon beobachtet. Ihr kennt euch schon länger und wisst deshalb auch, welche Mitschülerinnen und Mitschüler etwas besonders gut können.

Manche könne vielleicht besonders gut zeichnen, turnen usw.

Ich kenne euch erst seit heute und möchte gerne etwas über eure Fähigkeiten erfahren. Alle Fähigkeiten kann ich aber nicht abfragen. Ihr bekommt zwei Blätter mit drei Fragen und am Ende sollt ihr mir noch eure Muttersprache aufschreiben. Die Zettel werde ich nachher einsammeln und ich verspreche euch, niemand außer mir bekommt sie zu lesen.

Bitte schreibt zuerst euren Vornamen, den Namen der Schule und eure Klasse in die oberste Zeile. Ich lese den Text jetzt laut vor und ihr lest bitte still mit:

Denke an ein Kind, das jetzt in deiner Klasse sitzt und am besten rechnen kann. Welches Mädchen oder welcher Bub ist das?

Schreibe bis zu drei (also einen, zwei oder drei) sehr gute Rechnerinnen und Rechner der Klasse mit Vornamen auf. Bitte nur Personen aufschreiben, die heute mitmachen.

Du kannst dich auch selbst aufschreiben, wenn du zu den 3 Besten gehörst.

Bitte überlegt genau und schreibt nicht einfach eure Freunde auf, weil sie so nett sind, sondern nur, wenn sie zu den drei Besten gehören.

Schaut bitte auch nicht zu euren Sitznachbarn, denn ich möchte eure Meinung, und nicht die von eurem Nachbarn. Kennt sich jeder aus?

Wer mit der Beantwortung der ersten Frage fertig ist, legt bitte den Stift aus der Hand.“

Nachdem alle Kinder mit der Beantwortung der ersten Frage fertig sind:

TL: „Ich lese nun die zweite Frage vor:

Denke an ein Kind, das jetzt in deiner Klasse sitzt, und das am besten interessante Geschichten erzählen oder schreiben kann. Welches Mädchen oder welcher Bub ist das?

Schreibe bis zu drei Mitschülerinnen oder Mitschüler mit Vornamen auf. Du kannst auch wieder dieselben Personen, wie bei der ersten Frage aufschreiben. ***Du kannst dich auch selbst aufschreiben, wenn du zu den drei Besten gehörst.***

Überlegt nun wieder und schreibt die Namen auf!

Wer mit der Beantwortung der zweiten Frage fertig ist, dreht bitte das Blatt um.“

Nachdem alle Kinder mit der Beantwortung der zweiten Frage fertig sind:

TL: „Ich lese euch nun die dritte Frage vor:

Denke an ein Kind, das jetzt in deiner Klasse sitzt, und das sich am besten ausdrücken kann.

Damit meine ich, dass dieses Kind viele Wörter kennt, diesem Kind viele Wörter mit derselben oder ähnlichen Bedeutung einfallen, zum Beispiel: gehen, laufen, spazieren usw. und dass das Kind etwas genau beschreiben kann.

Welches Mädchen oder welcher Bub ist das?

Schreibe bis zu drei Mitschülerinnen oder Mitschüler mit Vornamen auf. Du kannst dich auch selbst aufschreiben, wenn du zu den drei Besten gehörst. Wer mit der Beantwortung der dritten Frage fertig ist, legt bitte den Stift aus der Hand.“

Die Zettel werden eingesammelt.

„Danke für eure Hilfe! Ihr ward sehr fleißig.“

Vorname: _____

Schule: _____ Klasse: _____

Bitte beantworte die folgenden Fragen:

Denke an ein Kind, das jetzt in deiner Klasse sitzt und am besten rechnen kann. Welches Mädchen oder welcher Bub ist das?

Schreibe bis zu drei sehr gute Rechnerinnen und Rechner der Klasse mit Vornamen auf. Du kannst dich auch selbst aufschreiben, wenn du zu den drei Besten gehörst.

1. _____

2. _____

3. _____

Denke an ein Kind, das jetzt in deiner Klasse sitzt, und das am besten interessante Geschichten erzählen oder schreiben kann.

Welches Mädchen oder welcher Bub ist das?

Schreibe bis zu drei Mitschülerinnen oder Mitschüler mit Vornamen auf. Du kannst dich auch selbst aufschreiben, wenn du zu den drei Besten gehörst.

1. _____

2. _____

3. _____

Denke an ein Kind, das jetzt in deiner Klasse sitzt, und das sich am besten ausdrücken kann.

Welches Mädchen oder welcher Bub ist das?

Schreibe bis zu drei Mitschülerinnen oder Mitschüler mit Vornamen auf. Du kannst dich auch selbst aufschreiben, wenn du zu den drei Besten gehörst.

1. _____

2. _____

3. _____

Deine **Muttersprache** ist _____



Danke für deine Mitarbeit!

Lebenslauf

Persönliche Daten:

Katharina Klockgether
geboren am 21.11.1985 in Göttingen

Bildungsweg:

2005	Abitur am Beethovengymnasium Bonn
seit 2005	Studium der Psychologie an der Universität Wien

Praktische Erfahrungen:

2001 – 2007	Tätigkeit als Kindergruppenleiterin in Bonn
2007	6-wöchiges Forschungspraktikum in der Medizinischen Psychologie des Universitätsklinikums Bonn
2009/10	Sprachstanderfassung der Vorschulkinder am Zwi-Perez-Chajes Kindergarten Wien
2009/10	Kinderbetreuung im Mutter-Kind-Heim Müllnergasse (Wien) der Caritas Socialis
2011	8-wöchiges Praktikum an der Psychiatrie des Universitätsklinikums Aachen (Forschung und Klinik)
2009 – 2011	Wissenschaftliche Mitarbeit am Charlotte Bühler Institut e. V.: Erstellen eines Qualitätsprogramms zur Identifikation und Förderung von Hochbegabten im Vorschulalter