

Einschätzung kindlicher Entwicklung
Zur Aussagekraft von informellen und formellen Diagnoseleistungen

Inauguraldissertation
zur
Erlangung der Würde
einer Doktorin der Philosophie
vorgelegt der
Fakultät für Psychologie
der Universität Basel

von

Marie-Claire Frischknecht Brunner

aus Heiden/AR

Wetzikon, 2014



Genehmigt von der Fakultät für Psychologie

auf Antrag von

Prof. Dr. Alexander Grob
Ass.-Prof. Dr. Sakari Lemola

Basel, den _____

Prof. Dr. Roselind Lieb



Danksagung

In erster Linie möchte ich Prof. Dr. Alexander Grob meinen herzlichsten Dank aussprechen. Er hat mich mit seiner fachlich hochkompetenten kritisch-konstruktiven und wertschätzenden Begleitung in allen Phasen des langen Weges vom Anfangsgedanken bis zur Fertigstellung dieser Dissertation wohlwollend unterstützt. Ohne seinen geduldigen, konstruktiven Beistand, wäre ich diesen Weg nicht zu Ende gegangen. Seinem Vertrauen habe ich es auch zu verdanken, dass ich am Projekt der Entwicklung der *Intelligence and Development Scales – Preschool* mitwirken konnte. Beide Projekte ermöglichten mir neue Lernprozesse und führten zu einer Bereicherung auf vielen Ebenen.

Ein herzlicher Dank gilt auch Ass.-Prof. Dr. Sakari Lemola dessen kollegiale unterstützende Art ich im Laufe dieser Jahre ebenfalls kennen und schätzen gelernt habe und der sich nun zu meiner grossen Freude als Zweitgutachter für die vorliegende Dissertation zur Verfügung gestellt hat. Schliesslich möchte ich Prof. Dr. Rainer Greifeneder danken, welcher sich als Kommissionsvorsitzender zur Verfügung gestellt hat.

Ein besonderer Dank gilt meiner Projektpartnerin und Co-Autorin Giselle Reimann. Mit ihrer geduldigen Art und optimistischen Haltung hat sie wesentlich dazu beigetragen, dass ich mich von der Praxis her kommend in die wissenschaftliche Denkweise einarbeiten konnte. Giselle Reimann hat mich in diesem langjährigen Prozess der Dissertationserstellung mit ihren hervorragenden Fachkenntnissen und auch persönlichem Beistand begleitet und unterstützt. Dafür möchte ich mich ganz herzlich bedanken.

Und schliesslich möchte ich meinem Ehemann und Partner Ueli Frischknecht herzlichst danken. Mit seiner persönlichen Unterstützung und stets positiven und ermunternden Einstellung zu meinem Vorhaben war es für mich möglich diesen Weg zu gehen und mit meinen familiären Aufgaben und beruflichen Verpflichtungen zu vereinbaren. Ein herzlicher Dank gilt auch meinen beiden Töchtern Estelle und Gloria, die stets für einen gesunden Ausgleich zwischen Arbeit und Familie sorgten. Auch meiner Freundin Susi Mäder möchte ich für ihre persönliche Unterstützung meinen herzlichen Dank aussprechen.

Inhaltsverzeichnis

Danksagung	II
Zusammenfassung	VI
Einleitung	1
Theoretischer Hintergrund	5
Informelle Diagnostik: Elterneinschätzung.....	5
Vergleich von Elterneinschätzung und Testleistung.....	10
Formelle Diagnoseleistung: Interpretation von Leistungsprofilen.....	12
Forschungsfragen	14
Methode	16
Studie 1 (Frischknecht, Reimann & Grob, 2014).....	16
Studie 2 (Frischknecht, Reimann, Gut, Ledermann & Grob, 2013)	17
Studie 3 (Reimann, Stoeklin, Lavalley, Gut, Frischknecht & Grob, 2012)	18
Zusammenfassung der Resultate	21
1. Informelle Diagnoseleistung: Elterneinschätzung im Vorschulalter	21
2. Informelle Diagnoseleistung: Elterneinschätzung zu relevanten Schulfächern im Primarschulalter	21
3. Diagnoseleistung im formellen Kontext: Zur Validität von intraindividuellen Profilanalysen.....	22
Gesamt-Diskussion	23
Stärken und Schwächen	26

Praktische Implikationen.....	28
Konklusion und Ausblick.....	29
Literatur.....	31
Anhang	39

Zusammenfassung

Ziel der vorliegenden Dissertation, welche 3 Studien beinhaltet, ist es, den aktuellen Forschungsstand bezüglich Validität informeller und formeller Diagnoseleistungen zu erweitern. Dabei geht es um die Akkuratheit von Elterneinschätzung kindlicher Entwicklung und Schulleistungen in Mathematik und Deutsch und darüber hinaus um die Validität von diagnostischen Schlussfolgerungen aufgrund der Interpretation von intraindividuellen Leistungsprofilen.

Zusammengefasst zeigen die Resultate der Studien, dass Elterneinschätzungen bezüglich relevanten Entwicklungsbereichen und dem Gesamtentwicklungsstand bei 3-5 jährigen Kindern sowie bei Leistungseinschätzungen in den Schulfächern Mathematik und Deutsch bei 5-10 Jährigen positiv mit entsprechenden Testleistungen des Kindes korrelieren. Dennoch geben Elterneinschätzungen infolge ihrer geringen Akkuratheit kaum Hinweise über das Ausmass einer Entwicklungsabweichung. Die vorliegende Dissertation zeigt auf, dass die Elternwahrnehmung zwar Hinweise auf Entwicklungsauffälligkeiten sowie mögliche Schwächen in den Leistungsfächern Deutsch und Mathematik geben kann aber nicht ausreicht, um im Sinne eines Screenings frühzeitig Entwicklungsdefizite zu erkennen und unterstreicht somit die Notwendigkeit des Einsatzes von validierten Entwicklungsscreenings ebenso wie die Anwendung von Entwicklungstests durch Experten. Des Weiteren konnte bezüglich professioneller Diagnostik aufgezeigt werden, dass durch das intraindividuelle Profil zusätzliche Varianz in den mathematischen Kompetenzen erklärt werden kann und damit ein Beitrag zu diagnostischen Schlussfolgerungen geleistet werden kann. Gleichzeitig erscheint es sinnvoll Elterneinschätzungen im diagnostischen Prozess als wertvolle Ergänzung zu integrieren.

Einleitung

Gemäss dem bio-ökologischen Modell von Bronfenbrenner (1979/1986) ist die kindliche Entwicklung in einem multiplen Kontext, bestehend aus einem Mikrosystem, Mesosystem, Exosystem und Makrosystem zu verstehen. Zu den proximalen Faktoren und damit zum Mikrosystem des Kindes gehören einerseits die kognitiven Fähigkeiten und die sozio-emotionale Entwicklung des Kindes selbst und andererseits externe Faktoren wie die häusliche Umgebung und die Schule. Damit wird der unmittelbaren Lebens- und Lernumgebung des Kindes eine zentrale Rolle in der kindlichen Entwicklung attestiert. Während die kognitive Leistungsfähigkeit einer der besten Prädiktoren für die schulische Leistung darstellt (Deary, Strand, Smith & Fernandez, 2007; Kahana, Youngstrom & Glutting, 2002), zeigen Studien zur elterlichen Erwartungshaltung bezüglich schulischer Karriere, dass diese ebenfalls einen relevanten Einfluss auf die Leistungen des Kindes haben (Byrnes & Wasik, 2009; Halle, Kurtz-Costes & Mahoney, 1997; Phillipson & Phillipson, 2007; Pomerantz & Dong, 2006). Dabei spielt die Passung (Hunt & Paraskevopoulos, 1980; Schrader, 2006) zwischen kindlicher Leistungsfähigkeit und elterlicher Erwartung, bzw. elterlicher Einschätzung des kindlichen Potentials, eine wichtige Rolle für eine positiv verlaufende Entwicklung des Kindes.

Die Elterneinschätzung kindlicher Entwicklung und Leistungsfähigkeit wird in der Literatur als informelle Diagnose bezeichnet, welche meist beiläufig und unsystematisch im Rahmen des alltäglichen erzieherischen Handelns entsteht und sich aus impliziten, subjektiven Urteilen, Einschätzungen und Erwartungen zusammensetzt (Hascher, 2005; Schrader, 2006). Im Gegensatz dazu werden bei einer formellen Diagnostik Instrumente eingesetzt die es ermöglichen, sowohl den

Gesamtentwicklungsstand als auch einzelne Funktionsbereiche in einer differenzierten entwicklungsdiagnostischen Abklärung zu testen. Sowohl in der formellen als auch in der informellen Diagnostik werden nebst einer Gesamteinschätzung des allgemeinen Entwicklungsstandes Stärken und Schwächen in einzelnen Entwicklungsbereichen beobachtet und thematisiert. Bezogen auf die Elterneinschätzung im diagnostischen Prozess hat die Auskunft der Eltern bezüglich einzelner Funktionsbereiche bereits im ersten Anamnesegespräch grosses Gewicht, indem Eltern über den Entwicklungsstand in einzelnen Funktionsbereichen ihres Kindes und über dessen individuellen Entwicklungsverlauf berichten (Priller, 1996). Die elterliche Einschätzung einer Stärke oder Schwäche beim eigenen Kind erfolgt dabei nicht nur im Vergleich zu Gleichaltrigen, sondern u.a. auch in Bezug auf andere Entwicklungsbereiche bei ihrem eigenen Kind. In diesem Zusammenhang sind Eltern zuhause implizit mit einem intraindividuellen Kompetenzprofil ihres Kindes konfrontiert. In Bezug auf die formelle Diagnostik zeigt sich der intraindividuelle Ansatz auch in der Praxis der Schulpsychologie. Bei der Gesamteinschätzung der kognitiven Fähigkeiten erfolgt diese durch die Ermittlung der allgemeinen Profilhöhe, beispielsweise in Form des IQ, welcher auf dem interindividuellen Vergleich mit einer Referenzgruppe basiert. Gleichzeitig liegt auch ein intraindividuelles Leistungsprofil vor. Das intraindividuelle Leistungsprofil bezieht sich auf das Profilmuster des einzelnen Kindes, welches aufgrund der Subtestresultate in Form von Stärken und Schwächen eines Kindes sichtbar wird (Davison & Kuang, 2000; Ding, 2001). Zwei Profile mit der exakt gleichen allgemeinen Profilhöhe können ein unterschiedliches Profilmuster bezüglich hohen und tiefen Wertpunkten aufweisen (Watkins & Glutting, 2000). Diagnostiker stützen sich deshalb nicht nur auf die allgemeine Profilhöhe, sondern auch auf das intraindividuelle Profil des Kindes selbst, indem sie aufgrund einer Subtestanalyse

Schlussfolgerungen bezüglich Stärken und Schwächen des Kindes ziehen (Pfeiffer, Reddy, Kletzel, Schmelzer & Boyer, 2000). Die Eruierung von individuellen Stärken und Schwächen eines Kindes soll dazu dienen, massgeschneiderte Fördermassnahmen einzuleiten (Grob, Hagmann-von Arx & Bodmer, 2009), wobei der Einbezug von Eltern bei der Planung von schulischen Massnahmen (Schrader, 2006) oder bei Frühfördermassnahmen (Rennen-Allhoff, Allhoff, Bowi & Laaser, 1993) deren Akzeptanz fördert und bei den Eltern zu einer höheren Identifikation mit den getroffenen Massnahmen führt. Gemäss dem Zürcher Fit-Konzept (Largo & Jenni, 2005) kann die kindliche Entwicklung besser verlaufen und das Wohlbefinden des Kindes gesteigert werden, je besser es Eltern und Fachleuten gelingt sich auf das individuelle Entwicklungsprofil des Kindes einzustellen (Jenni, Benz & Latal, 2011). Die erfolgreiche Zusammenführung von informeller und formeller Diagnostik ist deshalb ein wichtiger Bestandteil einer entwicklungspsychologischen oder – pädiatrischen Beratungssituation. Erkenntnisse zur Akkuratheit elterlicher Einschätzung und zur Validität von diagnostischen Schlussfolgerungen können zur Erhöhung der Beratungsqualität beitragen.

Die vorliegende Arbeit hat erstens zum Ziel die Übereinstimmung zwischen dem aktuellen kindlichen Entwicklungs- und Leistungsstand und der elterlichen Einschätzung zu überprüfen. Davon ausgehend, dass die Elterneinschätzung als informelle Diagnose meist beiläufig und unsystematisch im Rahmen des alltäglichen erzieherischen Handelns entsteht (Hascher, 2005; Schrader, 2006) ist zu erwarten, dass sich Eltern für die Generierung ihrer Einschätzung nicht an einer Struktur eines validen Screeningverfahrens orientieren. Daher soll der Frage nachgegangen werden, ob es den Eltern dennoch möglich ist, durch ihre informelle Diagnoseleistung den Entwicklungsstand ihres Kindes im Vorschulalter so adäquat zu beurteilen, dass sie im

Sinne eines Screenings einen Beitrag zur Früherkennung von Entwicklungsdefiziten leisten könnten, und ob Eltern in der Lage sind bei ihrem Kind im Schulalter das Leistungsniveau in zentralen Schulfächern wie Mathematik und Deutsch im Sinne einer guten Passung richtig einzuschätzen.

Zweitens soll die Validität der formellen Diagnoseleistung überprüft werden. Die Validität von interindividuellen Unterschieden, welche aufgrund von standardisierten entwicklungsdiagnostischen Instrumenten ermittelt wurden, ist gut dokumentiert (Alloway, 2009; Bull, Espy & Wiebe, 2008; Floyd, Evans & McGrew, 2003; Kahana et al., 2002; Watkins & Glutting, 2000). In der schulpsychologischen Praxis wird allerdings häufig auf das intraindividuelle Profil Bezug genommen. Die Interpretation anhand der eruierten Stärken und Schwächen innerhalb eines Leistungsprofils wird jedoch bezüglich zusätzlicher prädiktiver Validität (Glutting, Watkins, Konold & McDermott, 2006) in der Literatur infrage gestellt.

Theoretischer Hintergrund

Informelle Diagnostik: Elterneinschätzung

Die Bedeutung der Früherkennung entwicklungsgefährdeter Kinder wird in der Fachliteratur als zentral erachtet, hat sich doch gezeigt, dass Investitionen in die Frühförderung für einen Staat bedeutende Einsparungen nach ziehen (Schweinhart et al., 2005). Die Hälfte der entwicklungsauffälligen Kinder wird erst mit Eintritt ins Schulsystem erkannt (Glascoe, 1999). Um möglichst früh auf Entwicklungsdefizite reagieren zu können, müssten diese jedoch schon im vorschulischen Kontext erkannt werden. Eltern gelten als wichtige Informationsquellen bezüglich der Entwicklung ihres Kindes. Walker und Wiske (1981) gehen davon aus, dass Eltern über die meisten Informationen betreffend dem aktuellen Verhalten und der Entwicklungsgeschichte ihres Kindes verfügen und deshalb die besten Lieferanten von Informationen sind. Bestimmte Verhaltensweisen die mit Entwicklungsverzögerungen aufgrund von Wahrnehmungsdefiziten, Verhaltensauffälligkeiten, sozialen oder motorischen Beeinträchtigungen einhergehen, fallen zum Teil erst durch Langzeitbeobachtung auf und werden bei Kurzzeitbeobachtungen in der kinderärztlichen Praxis nicht unbedingt manifest (Rennen-Allhoff, Allhoff, Bowi & Laaser, 1993). Wie zuverlässig retrospektive Elternangaben sind, scheint unter anderem vom Inhalt der Frage anhängig. Besondere Meilensteine, wie beispielsweise erste Schritte oder das Geburtsgewicht können Eltern recht genau erinnern, während Angaben bezüglich Meilensteinen, die sich graduell entwickeln, wie die ersten Worte oder der Beginn des Lächelns weniger zuverlässig erinnert werden (Rennen-Allhoff, 1991). Elternangaben zu aktuellen Beobachtungen des Kindes und seinem Entwicklungsstand weisen hingegen darauf hin, dass die elterliche Beobachtung in Bezug auf körperlich bedingte Behinderungen effektiv ist (Rennen-Allhoff et al., 1993). Es stellt sich jedoch die Frage, ob Eltern

ebenfalls abnorme Entwicklungsverläufe sehen können, wenn es um die Einschätzung von weiteren relevanten Entwicklungsbereichen wie die intellektuelle Leistungsfähigkeit, beziehungsweise deren Korrelate im Sinne von einzelnen Wahrnehmungsbereichen oder sozial-emotionale Kompetenzen, Psychomotorik, mathematische Kompetenzen oder Sprachentwicklung geht. So wurden bei einer Schuleingangsuntersuchung am häufigsten Sprachstörungen (15.4%), graphomotorische Störungen (6.6%) und allgemeine Entwicklungsstörungen (6.4%) diagnostiziert (Tröster, Flender & Reineke, 2005). Die Tatsache, dass Entwicklungsauffälligkeiten in diesen Bereichen zum Teil erst bei Schul- oder Kindergarteneintritt festgestellt werden (Glascoe, 1999; Tröster et al. 2005) lässt vermuten, dass Eltern Entwicklungsabweichungen zu wenig zuverlässig erkennen können.

Eltern beobachten ihr Kind im natürlichen Alltagssetting und vergleichen sein Verhalten mit jenem von Altersgenossen (Chen, Lin, Wen & Wu, 2007; Glascoe, 1999; Glascoe & Dworkin, 1995; Glascoe & Sandler, 1995), wobei verschiedene Alltagsbeschäftigungen den Entwicklungsstand in unterschiedlichen Funktions- und Entwicklungsbereichen aufzeigen können und so den Eltern die Möglichkeit bieten die Fähigkeiten des eigenen Kindes zu erleben. Ihre Kenntnisse über die Kompetenzen des Kindes in schulischen Fächern generieren Eltern einerseits anhand eigener Beobachtungen, das Aufgabenverhalten des Kindes, aber auch über schriftliche Rückmeldungen aus der Schule.

Aus der Definition der informellen Diagnose (Hascher, 2005; Schrader, 2006) lässt sich ableiten, dass die Elterneinschätzung im Sinne einer elterlichen Wahrnehmung der Leistungsfähigkeit oder des Entwicklungsstandes des eigenen Kindes verschiedenen Einflüssen unterworfen ist, welche in der professionellen Diagnostik weitgehend nicht zum Tragen kommen. Diverse Untersuchungen weisen darauf hin, dass Eltern in der

Regel dazu neigen ihre Kinder bezüglich aktuellem Entwicklungsstand oder schulischem Leistungsvermögen zu überschätzen (Chen et al., 2007; Deimann, Kastner-Koller, Benka, Kainz & Schmidt, 2005; Malhi, Kashyap & Dua, 2005; Petermann, Stein & Macha, 2006; Buch, Sparfeldt & Rost, 2006; Willinger, Schaunig, Jantscher, Schmoeger et al., 2011; Willinger & Eisenwort, 2005). Nachfolgend sollen daher zwei Modelle zu Wirkungsmechanismen vorgestellt werden, welche bei der elterlichen Einschätzung der kindlichen Leistungsfähigkeit eine Rolle spielen könnten. Dabei wird die von (Bronfenbrenner 1979/1986) definierte Einbettung der kindlichen Entwicklung in einen multiplen Kontext berücksichtigt.

Semin und Papadopoulous (1989) entwarfen ein Modell *der erweiterten Identität der Mutter*, welches am Beispiel der sozial-emotionalen Kompetenz des Kindes und dem daraus folgenden kindlichen und mütterlichen Verhalten illustriert wird. Wenn dem Kind ein „faux pas“ passiert und es nicht in der Lage ist, diesen selbst wieder gut zu machen, übernimmt die Mutter einen Teil des Unvermögens des Kindes und führt auf diese Weise die Handlung zu Ende, indem sie zum Beispiel Betroffenheit und Bedauern ausdrückt, sich für das Kind entschuldigt und sich auch verantwortlich fühlt für das Geschehen. Indem sich die Mutter in diese Situation hinein gibt wird sie selbst Teil davon und erlebt das Gelingen oder Scheitern der Handlung als Teil ihrer selbst. Semin und Papadopoulous (1989) gehen davon aus, dass diese Dynamik in der Mutter-Kind-Dyade mit zunehmendem Alter und der damit einhergehenden Zunahme der Kompetenzen und Eigenständigkeit des Kindes abnimmt. In Bezug auf die elterliche Einschätzung von kindlichen Kompetenzen lässt sich aus diesem Modell ableiten, dass sowohl die Mutter, als auch der Vater, als ebenfalls massgeblich an der Erziehung und Betreuung des Kindes beteiligte Person, bei einer Bewertung der Fähigkeiten ihres Kindes möglicherweise im Sinne der erweiterten Identität auch eine Bewertung der

eigenen Kompetenzen vornehmen, wobei mit der Abnahme der erweiterten Identität auch ein Rückgang der Bewertung der eigenen Kompetenzen zugunsten einer Bewertung der Kompetenzen des Kindes einhergehen sollte. Aus diesem Modell lässt sich ableiten, dass bei einer Konstellation, in der das Kind Entwicklungsverzögerungen in einem oder mehreren relevanten Bereichen aufweist, dessen Eltern die erweiterte Identität länger aufrechterhalten müssten, als Eltern von entwicklungsunauffälligen Kindern, was zur Folge haben könnte, dass bei einer Beurteilung die Bewertung der eigenen Kompetenzen länger eine Rolle spielt, und damit verzögerte Entwicklungsverläufe beim Kind nicht rechtzeitig wahrgenommen werden können.

Das zweite Modell bezieht sich auf die Leistungsfähigkeit und ihren Stellenwert in unserer Gesellschaft, und damit auf die Kultur und ihre zugrundeliegenden ideologischen Gegebenheiten, welche nach Bronfenbrenner (1979/1986) dem Makrosystem zugeordnet werden können. Festinger (1954) zeigt in seinem *Modell der kognitiven Dissonanz* zwei mögliche Faktoren auf, welche die Einschätzung der Leistungsfähigkeit wesentlich beeinflussen können. Zum einen besteht in der westlichen Kultur eine Werthaltung welche impliziert, dass eine hohe Leistung mehr Wert hat, als eine geringe Leistung. Daraus folgt die Bestrebung sich in jeder Hinsicht verbessern zu müssen oder besser als Andere zu sein, was zwischen Individuen zu einem stetigen kompetitiven Prozess führt. Je besser die Bewältigung einer Aufgabe gelingt, desto attraktiver und wünschenswerter gilt dies in unserer Gesellschaft. In Bezug auf die Bewältigung von Entwicklungsaufgaben und schulischen Erfolg kann das bedeuten, dass es umso attraktiver und wünschenswerter ist, je besser dies gelingt. Gleichzeitig sind Individuen nach Festingers Theorie der kognitiven Dissonanz bestrebt sich Vergleichsreferenzen zu suchen, welche in etwa dem eigenen Leistungsniveau entsprechen. Eine zu hohe Diskrepanz führt Individuen dazu die eigene Wahrnehmung

bezüglich einer Fähigkeit dahingehend anzupassen, die Dissonanz und die damit entstehende Stresssituation zu vermindern, was wiederum zu einer Verzerrung der eigenen Wahrnehmung führen kann. In Bezug auf eine bestehende Entwicklungsverzögerung beim Kind könnte aus Festingers Theorie abgeleitet werden, dass Eltern durch den Vergleich mit anderen Kindern unter Umständen eine verzögerte Entwicklung bei ihrem Kind wahrnehmen, aber die eigene Wahrnehmung im Sinne einer Angleichung an die Referenzgrösse so weit verändern, dass die Diskrepanz zwischen einem unauffälligen und auffälligen Entwicklungsstand nicht mehr erkannt werden kann. Zugleich sind sie der Werthaltung westlicher Kultur unterworfen, welche eine stetige Verbesserung der Fähigkeiten „diktiert“ und die erfolgreiche Bewältigung einer Herausforderung honoriert.

Aus den Modellen von Semin und Papadopoulos (1989) und Festinger (1954) lässt sich ableiten, dass die elterliche Wahrnehmung dahingehend beeinflusst sein kann, dass Eltern die Fähigkeiten ihres Kindes höher einschätzen, als diese tatsächlich sind, und aufgrund des in der westlichen Kultur allgegenwärtigen stetigen kompetitiven Prozesses, die Kompetenzen ihres Kindes auch leicht höher einschätzen als der Durchschnitt der Vergleichspopulation. Es stellt sich daher die Frage wie aussagekräftig Elterneinschätzungen sein können.

Um zu überprüfen wie valide Elterneinschätzungen zum aktuellen Entwicklungs- oder Leistungsstand des Kindes sind, können sie mit Akten, Zeugnissen, Testergebnissen und Beobachtungen von anderen Beurteilern verglichen werden. Dabei werden Diagnoseleistungen aus verschiedenen Quellen miteinander verglichen, nämlich eine Diagnoseleistung, die aus der elterlichen Wahrnehmung resultiert und eine Diagnoseleistung, welche auf wissenschaftlich erprobten Methoden mit vorgegebenen Qualitätsstandards beruht (Schrader, 2006).

Vergleich von Elterneinschätzung und Testleistung

Studien, zur Akkuratheit oder Validität von Elterneinschätzungen fokussieren auf die korrelativen Zusammenhänge zwischen Elterneinschätzung und objektiv eingeschätzter kindlicher Leistung sowie die Sensitivität (Prozentsatz der Kinder mit tatsächlich vorliegender Entwicklungsverzögerung und korrekter elterlicher Identifikation), die Spezifität (Prozentsatz der Kinder ohne Entwicklungsverzögerung und korrekter elterlicher Identifikation) und den Positiv-Prädiktiven Vorhersagewert (PPV) von Elterneinschätzungen.

Korrelative Zusammenhänge. Bei Kindern im Alter von 3 – 6 Jahren werden betreffend relevanten Entwicklungsbereichen unterschiedliche Korrelationswerte berichtet. In Studien mit nicht-klinischen Stichproben wurden bisher je nach Funktionsbereich lediglich moderate oder gar keine substantiellen Zusammenhänge zwischen Elterneinschätzung und Testleistung des Kindes gefunden (Koch, Kastner-Koller, Deimann, Kossmeier, Koitz & Steiner, 2011; Storck, Webel, Steinmacher, Kupferschmidt & Bode, 1998). Für den Gesamtentwicklungsstand oder die Globalentwicklung konnten substantiell positive Korrelationen zwischen Einschätzung und Testergebnissen nachgewiesen werden (Glascoe & Sandler, 1995; Koch et al., 2011). Die Elterneinschätzung von schulischen Leistungen im Bereich Mathematik und Sprachkompetenzen korreliert substantiell positiv mit objektiven Testergebnissen (Phillipson & Phillipson, 2007), wobei bezüglich Mathematik zum Teil stärkere Zusammenhänge für Jungen gefunden wurden (Furnham, 2000; Lummis & Stevenson, 1990; Rätty, Vänskä, Kasanen & Kärkkäinen, 2002; Waschbusch, Daleiden & Drabman, 2000). Zudem scheint das Geschlecht des Kindes bei der Erwartung und bei der Beurteilung von Leistungserfolgen des Kindes eine Rolle zu spielen (Furnham, 2000, Lummis & Stevenson 1990; Rätty et al., 2002; Tiedemann, 2000). Darüber hinaus

scheint der Bildungsstand der Mutter einen Einfluss auf die Leistungseinschätzung und die Leistung selbst zu haben (Davis-Kean, 2005; Halle, Kurtz-Costes & Mahoney, 1997; Lee, 2010).

Sensitivität, Spezifität, Positiv-Prädiktiver Wert. Bei Kindern im Alter von 3-6 Jahren kommen Untersuchungen mit nicht-klinischen Stichproben (Koch et al., 2011; Miller, 1986; Storck et al., 1998) bezüglich Sensitivität und Spezifität oder auch Positiv-Prädiktivem Vorhersagewert (PPV) von Elterneinschätzungen zu weniger befriedigenden Werten als Studien mit klinischen Stichproben (z.B. Chen et al., 2007; Chung, Chang, Chen, Tang & Wong, 2010; Glascoe & Dworkin, 1995; Malhi et al., 2005; Pulsifer, Hoon, Plamer, Copalan & Capute, 1994; Willinger & Eisenwort, 2005). So konnten Storck et al. (1998) in ihrer Studie aufzeigen, dass die Eltern im Rahmen einer Vorsorgeuntersuchung bei Kindern im Alter von 5 Jahren unabhängig von der spezifischen Teilfunktion nur ca. 30% der entwicklungsauffälligen Kinder korrekt identifizierten. Nicht nur bei der Einschätzung einzelner Entwicklungsbereiche, sondern auch bei der Einschätzung des Gesamtentwicklungsstandes werden bei nicht-klinischen Stichproben geringe Sensitivitätsmasse und geringe Positiv-Prädiktive Werte erreicht (Glascoe, 1994).

In Bezug auf Schulleistungen konnten Studien zur Elterneinschätzung aufzeigen, dass Eltern positive Einschätzungskategorien bevorzugen, wenn sie die Schulleistungen ihres Kindes beurteilen sollen, was dazu führt, dass eher Kategorien gewählt werden, welche für eine überdurchschnittliche Leistung stehen, während Einschätzungskategorien, welche auf unterdurchschnittliche Leistung hinweisen, seltener gewählt werden (Räty et al., 2002).

Da keine standardisierten Instrumente zur Erfassung der Elterneinschätzung vorliegen unterscheiden sich die Studien vor allem dahingehend, wie

Elterneinschätzungen operationalisiert wurden, was eine Vergleichbarkeit der Ergebnisse erschwert. Auch die Erfassung des Leistungspotentials beim Kind ist unterschiedlich und erfolgt nicht immer über standardisierte Verfahren (Kovas, Haworth, Petrill & Plomin, 2007; Lummis & Stevenson, 1990; Phillipson & Phillipson, 2007; Tiedemann, 2000). Bei standardisierten Testverfahren ist der Cut-off klar definiert: Als auffällig gilt eine Leistung, die mehr als eine Standardabweichung vom Mittelwert abweicht. Bei den Elterneinschätzungen wird dies je nach Studie unterschiedlich definiert.

Formelle Diagnoseleistung: Interpretation von Leistungsprofilen

In der formellen Diagnostik werden validierte Verfahren in Form von Entwicklungs- und Leistungstest eingesetzt. Die kognitive Leistungsfähigkeit (IQ) stellt einer der besten Prädiktoren für die schulische Leistung dar (Deary et al. 2007; Kahana et al. 2002). So zeigte sich am Beispiel von rechenschwachen Kindern, dass diese im Vergleich zu Kontrollkindern häufig eine tiefere Gesamtintelligenz aufweisen (Andersson, 2010; Geary, Hoard, Nugent & Baily, 2012). Zusätzlich zur Gesamteinschätzung der intellektuellen Fähigkeiten wird das Vorliegen eines Leistungsprofils, welches relevante Wahrnehmungsbereiche beinhaltet, im schulpsychologischen Alltag genutzt für die Empfehlung von Fördermassnahmen. So zeigt eine Studie aus dem Jahre 2000, dass im amerikanischen Raum 89% der Schulpsychologen klinischen Nutzen aus der Subtestanalyse ziehen (Pfeiffer et al., 2000), wobei 70% davon eine Profilanalyse, also die Interpretation von Stärken und Schwächen anhand des intraindividuellen Profilmusters des Kindes unabhängig von der Gesamtprofilhöhe, als valables Instrument erachteten. Davon abgeleitete massgeschneiderte Fördermassnahmen sollten beispielsweise in Mathematik einer Verfestigung und Kumulation von Defiziten entgegenwirken (Fritz & Ricken, 2005;

Grob et al., 2009, Lorenz, 2009; von Aster, Schweiter, Weinold Zulauf, 2007). Dieses in der schulpsychologischen Praxis verbreitete Vorgehen (Pfeiffer et al., 2000; Jenni, Benz & Latal, 2011) wird einerseits unterstützt durch Studienbefunde, die darauf hinweisen, dass das intraindividuelle Profil der kognitiven Variablen über die Gesamtprofilhöhe - das heisst über den IQ - hinaus einen Beitrag bei der Vorhersage von Schulleistungen wie beispielsweise Mathematik leisten kann (Naglieri, 2000; Pritchard, Livingston, Reynolds & Moses, 2000) andererseits wird die zusätzliche prädiktive Validität eines intraindividuellen Profils in der Literatur infrage gestellt (Glutting et al., 2006).

Forschungsfragen

Wie adäquat Eltern die kindliche Entwicklung und das Leistungspotential des Kindes wahrnehmen, spielt in zweierlei Hinsicht eine Rolle. Zum einen können Eltern zur Früherfassung von Entwicklungsdefiziten beitragen, wenn sie anhand ihrer Beobachtungen abnorme Entwicklungsverläufe beobachten, zum anderen kann eine Passung (Hunt & Paraskevopoulos, 1980) zwischen den an das Kind gestellten Anforderungen und seinen Möglichkeiten und Fähigkeiten das Erziehungsverhalten und die kindliche Entwicklung positiv beeinflussen, während eine fehlende Passung zu einer gestörten Entwicklung führen kann (Grob & Jaschinski, 2003; Schrader, 2006). Bei Kindern, die auf Fördermassnahmen angewiesen sind, kann eine adäquate Einschätzung der Eltern bezüglich der zu unterstützenden Leistungsdefizite zu einer höheren Identifikation mit der getroffenen Massnahme führen. Empfehlungen für unterstützende Massnahmen erfolgen in der Praxis wiederum aufgrund der Gesamtprofilhöhe des Entwicklungsstandes oder des IQ, sowie der Interpretation des intraindividuellen Leistungsprofils des Kindes.

Angesichts der hohen Bedeutung der elterlichen Wahrnehmung wie auch der Findung von adäquaten Unterstützungsmassnahmen für einen positiven Entwicklungsverlauf stellt sich die Frage, inwieweit die informelle Diagnoseleistung der Eltern die tatsächlichen kindlichen Kompetenzen erfassen kann und inwiefern die in der Praxis angewandte Interpretation von intraindividuellen Leistungsprofilen, einen Beitrag zur Empfehlung von spezifischen Fördermassnahmen leisten kann.

Ziel der vorliegenden Arbeit ist es die Akkuratheit von Elterneinschätzungen als informelle Diagnoseleistung bei Kindern im Vorschul- und im frühen Schulalter zu überprüfen und die prädiktive Validität von intraindividuellen Leistungsprofilen als Bestandteil der formellen Diagnoseleistung zu untersuchen.

Aus den bisher dargelegten Überlegungen lassen sich drei Hauptfragestellungen ableiten:

1. Zur Aussagekraft von informeller Diagnoseleistung im Vorschulalter:
Sind Eltern aufgrund ihrer Beobachtung in der Lage, bei ihrem Kind im Vorschulalter allfällige Entwicklungsverzögerungen so zu erkennen, dass sie die Funktion eines Screenings in diesem Stadium wahrnehmen könnten?
2. Zur Aussagekraft von informeller Diagnoseleistung im Primarschulalter:
Wie gut entsprechen Leistungseinschätzungen der Mütter den tatsächlichen Kompetenzen der Kinder in den relevanten Schulfächern Mathematik und Deutsch?
In Bezug auf mögliche Einflussfaktoren, stellt sich zudem die Frage ob sich das Geschlecht des Kindes oder der Bildungsstand der Mutter auf die Einschätzung und/oder auf die Leistung des Kindes selbst auswirken.
3. Zur Aussagekraft von formeller Diagnoseleistung im Primarschulalter:
Inwiefern kann nebst der Interpretation der allgemeinen Profilhöhe die in der Praxis angewandte intraindividuelle Analyse von Untertestwerten einen Beitrag zur Vorhersage von Leistungsschwächen leisten?

Methoden

Die Daten zur Beantwortung der Forschungsfrage 1 stammen aus der Normierungsstichprobe der Intelligence and Development Scales – Preschool (IDS-P, Grob, Reimann, Gut & Frischknecht, 2013) die für Kinder im Alter von 3;0 bis 5;11 Jahren konzipiert sind und die Daten zur Beantwortung der Forschungsfragen 2 und 3 stammen aus der Normierungsstichprobe der Intelligence and Development Scales (IDS, Grob, Meyer & Hagmann von Arx, 2009), die den Altersbereich von 5;0 bis 10;11 Jahren abdecken. Mit beiden Versionen liegen umfangreiche Testbatterien vor, die zuverlässig zwischen normal entwickelten Kindern und verschiedenen klinischen Gruppen unterscheiden. Die repräsentativen Normierungsstichproben garantieren eine professionelle Erhebung des kognitiven Leistungspotentials und des Entwicklungsstandes in den Bereichen der allgemeinen Entwicklung. Parallel zur testpsychologischen Untersuchung der Kinder wurden die Eltern mittels eines Fragebogens zu ihrer Leistungseinschätzung befragt. Die Elterneinschätzung wird in der vorliegenden Arbeit der „informellen Diagnostik“, welche einem nicht normierten Elternfragebogen entspricht, zugeordnet.

Studie 1 (Frischknecht, Reimann & Grob, 2014)

Ziel der Studie 1 war, zu untersuchen ob Eltern bei ihren Kindern im Vorschulalter in der Lage sind ein Entwicklungsdefizit in relevanten Entwicklungsbereichen oder in der Gesamtentwicklung zu erkennen. Diesbezüglich wurde nebst den korrelativen Zusammenhängen zwischen Elterneinschätzung und Testleistung des Kindes insbesondere überprüft, ob Elterneinschätzungen die Kriterien für ein akkurates Screening erfüllen in Bezug auf hinreichende Sensitivität, Spezifität, PPV (positiv prädiktiver Vorhersagewert), RATZ-Index (Relativer Anstieg der Trefferquote gegenüber der Zufallstrefferquote) und AUC-Werte (Angabewert zur

Akkuratheit von Elterneinschätzungen). Dazu wurden in einer nicht-klinischen Stichprobe die Elterneinschätzungen zum Entwicklungsstand von 539 Kindern ($M = 4;6$ Jahre, $SD = 0;10$ Jahre; 48 % Mädchen) mit den Testleistungen derselben Kinder in den Intelligence and Development Scales – Preschool (IDS-P, Grob et al., 2013), verglichen. Die Stichprobe bestand aus einer Teilmenge der Normierungsstichprobe ($N = 713$) der IDS-P, welche im Einzelsetting durchgeführt wurde. Die Kinder wurden in der Schweiz (53%), in Deutschland (21%) und Österreich (26%) über Spielgruppen, Tagesstätten, Kindergarten und Elternforen rekrutiert. In die vorliegende Studie wurden die Kinder einbezogen, für die der von den Eltern beantwortete Fragebogen vorlag. Die Testleistungen umfassten die relevanten Entwicklungsbereiche in Form der Kognitiven Entwicklung (7 Untertests) und der Allgemeinen Entwicklung (8 Untertests). Im Elternfragebogen wurden die Eltern dazu aufgefordert anhand von konkreten, alltagsnahen Beobachtungsbeispielen eine altersbezogene Einschätzung zum Entwicklungsstand ihres Kindes abzugeben. Es wurde darauf geachtet, dass die Beispielitems möglichst genau die geforderten Leistungen im entsprechenden Untertest abbilden. Zusätzlich mussten die Eltern eine Einschätzung zum Gesamtentwicklungsstand vornehmen.

Studie 2 (Frischknecht, Reimann, Gut, Ledermann & Grob, 2013)

Das Ziel der Studie 2 war, zu überprüfen, wie genau Mütter die Mathematik- und Sprachleistungen ihrer 6- bis 10-jährigen Kinder einzuschätzen vermögen. Die untersuchte Stichprobe bestand aus einer Teilmenge der Normierungsstichprobe ($N = 1330$) der IDS und umfasste 382 Kindern (190 Jungen und 192 Mädchen, $M = 7.98$ Jahre, $SD = 1.24$ Jahre) welche über die Grundschule (1. - 4. Klasse) in der Schweiz, Deutschland und Österreich rekrutiert worden waren. Die Untersuchung verfolgte drei Ziele: Erstens sollte die Übereinstimmung der mütterlichen Einschätzung mit der

kindlichen Testleistung untersucht und geprüft werden, ob Überschätzungen häufiger vorkommen als Unterschätzungen und ob allfällige geschlechtsspezifische Effekte bei der Genauigkeitseinschätzung vorhanden sind. Zweitens wurde der Zusammenhang zwischen mütterlicher Leistungseinschätzung in den Schulfächern Mathematik und Deutsch und der tatsächlichen Testleistung erfasst. Des Weiteren wurde getestet, ob der Zusammenhang zwischen der Testleistung und der Leistungseinschätzung durch das Geschlecht des Kindes moderiert wird, wobei der Bildungsstand der Mutter als Kovariate einbezogen wurde. Drittens wurde untersucht, ob die mütterlichen Einschätzungen der Leistungen ihrer Kinder je nach Geschlecht unterschiedlich ausfallen und ob sich in den Testresultaten geschlechtsspezifische Unterschiede finden lassen. Die mütterliche Einschätzung wurde über einen Fragebogen erhoben und anschliessend verglichen mit der Testleistung der Kinder in den Bereichen Sprachentwicklung und Mathematisches Denken der IDS (Intelligence and Development Scales; Grob et al., 2009).

Studie 3 (Reimann, Stoeklin, Lavalée, Gut, Frischknecht & Grob, 2012)

In Studie 3 lag der Fokus auf der Ebene des Kindes und seinen kognitiven und motivationalen Voraussetzungen, die insbesondere eine Rolle spielen für seine Kompetenzen in Mathematik. Für die vorliegende Arbeit interessiert insbesondere die Fragestellung ob das intraindividuelle Profil über die allgemeine Profilhöhe hinaus genutzt werden kann um die mathematischen Kompetenzen besser vorauszusagen. Dazu wurden 567 Kinder, ($M = 8.03$ Jahre, $SD = 1.63$ Jahre, 53% Mädchen) untersucht, die eine bedeutsame Variabilität in ihrem Profil aufwiesen. Zunächst wurde untersucht, welche kognitiven Variablen prädiktiv waren für mathematische Kompetenzen und in einem zweiten Schritt wurde das häufigste kognitive Profil ermittelt um allfällige Muster zu finden, welche zusätzliche Information für die Vorhersage der

mathematischen Kompetenz liefern könnten. Schliesslich wurde untersucht, ob dieses Profil über die Profilhöhe hinaus einen zusätzlichen Beitrag zur Vorhersage der mathematischen Kompetenz leisten kann. Da sieben von neun Untertests im Profil kognitive Variablen darstellten und die Korrelation zwischen Profilhöhe und der allgemeinen Intelligenz $r = .96$, ($p < .001$) betrug, wurde nahezu die allgemeine Intelligenz abgebildet, weshalb die Profilhöhe als Repräsentant für die allgemeine Intelligenz verwendet werden konnte.

Tabelle 1

Übersicht über die Studien

	Relevante Variablen für die jeweiligen Studien	Stichprobe	Analysemethode
Studie 1	<p>IDS-P - Testleistung</p> <ul style="list-style-type: none"> • Kognition • Allgemeine Entwicklung (Insgesamt 15 Untertests) <p>Elternfragebogen</p> <ul style="list-style-type: none"> • Elterneinschätzung anhand von Alltagsbeispielen • Elterneinschätzung zum Gesamtentwicklungsstand 	<p>Kinder im Alter von 3;0 bis 5;11 Jahren. ($N = 539$; $M = 4;6$ Jahre, $SD = 0;10$ Jahre)</p>	<p>Korrelationsanalysen</p> <p>Ermittlung von Sensitivität, Spezifität, PPV- RAZ-Index, ROC</p> <p>Kurven-Analyse</p>
Studie 2	<p>IDS - Testleistung</p> <ul style="list-style-type: none"> • Mathematik • Sprache <p>Elternfragebogen</p> <ul style="list-style-type: none"> • Einschätzung der Mutter für die schulischen Leistungen des Kindes im Fach Mathematik • Einschätzung der Mutter für die schulischen Leistungen des Kindes im Fach Deutsch 	<p>Schulkinder ($N = 382$; $M = 7.98$ Jahre, $SD = 1.24$ Jahre)</p>	<p>Chi-Quadrat-Tests</p> <p>Korrelationsanalysen</p> <p>Fischer's Z-Transformation</p> <p>Hierarchische Regressionsanalyse</p> <p>t-Tests</p>
Studie 3	<p>IDS - Testleistung</p> <ul style="list-style-type: none"> • Mathematik • Kognition • Motivation 	<p>Schulkinder mit grosser intraindividuellem Varianz ($N = 567$)</p>	<p>Multiple sowie Hierarchische Regressionsanalysen</p> <p>Profilanalyse via Multidimensionale Skalierung</p>

Zusammenfassung der Resultate

Nachfolgend werden die Resultate der Studien, welche Bestandteil dieser kumulativen Dissertation sind, anhand der Fragestellungen präsentiert:

1. Informelle Diagnoseleistung: Elterneinschätzung im Vorschulalter

Die Elterneinschätzung bei Kindern im Vorschulalter korrelierte in dieser nicht-klinischen Stichprobe fast durchweg positiv mit der Testleistung des Kindes. Insgesamt müssen die Korrelationen jedoch als gering bezeichnet werden. Bei der elterlichen Einschätzung der Gesamtentwicklung wurde hingegen ein Korrelationswert im mittleren Bereich erzielt. Die Analysen in Bezug auf Sensitivität, Positiv-prädiktiver Wert PPV, RAZ-Index, ROC-Kurve (AUC-Wert) zeigten auf, dass die Elterneinschätzung in Hinblick auf einzelne Entwicklungsbereiche die Bedingungen für ein Screening nicht erfüllen kann. Lediglich bei den Spezifitätswerten wurde bei der Einschätzung von Entwicklungsbereichen in der Mehrheit der Bereiche sowie bei der Einschätzung der Gesamtentwicklung das geforderte Mass von mindestens 70% (American Psychological Association, 1985; Aylward, 1997) erreicht. Die Einschätzung der Gesamtentwicklung zeigte, dass bezüglich Sensitivität und Positiv Prädiktivem Wert die geforderten Masse zwar ebenfalls nicht erreicht wurden, hingegen der Spezifitätswert, der RAZ-Index sowie die Resultate der ROC-Analyse auf eine moderate Verbesserung gegenüber einer Zufallseinschätzung hinweisen.

2. Informelle Diagnoseleistung: Elterneinschätzung zu relevanten Schulfächern im Primarschulalter

Bei den Kindern im Alter von 5-10 Jahren zeigte sich betreffend der schulischen Fächer Deutsch und Mathematik, dass bezüglich Übereinstimmung von mütterlicher Einschätzung und kindlicher Leistung, bei beiden Geschlechtern Überschätzungen

häufiger vorkamen als Unterschätzungen. Im Vergleich zu den Testleistungen, die in etwa gleiche Anteile an über – und unterdurchschnittlichen Leistungen auswiesen, war bei der mütterlichen Einschätzung die Kategorie „überdurchschnittlich“ deutlich häufiger vertreten. Die korrelativen Zusammenhänge zwischen elterlicher Einschätzung und tatsächlichen Testleistungen lagen im schwachen bis mittleren Bereich, wobei bei den Korrelationskoeffizienten im Bereich Mathematik kein Unterschied zwischen Mädchen und Jungen hervorging, hingegen im Bereich Sprache ein höherer Korrelationskoeffizient zugunsten der Jungen ausgewiesen wurde. Die Regressionsanalyse zeigte, dass sowohl der Bildungsstand, als auch die Einschätzung der Mutter signifikante Prädiktoren darstellten in Bezug auf die Testleistungen in Sprache und Mathematik. Im Bereich Sprache zeigte sich zusätzlich, dass der Zusammenhang zwischen Einschätzung und Testleistung abhängig ist vom Geschlecht des Kindes. Des Weiteren zeigten sich geschlechtsspezifische Unterschiede sowohl in der Testleistung als auch in der Mütterereinschätzung indem im Bereich Mathematik Testleistung und Mütterereinschätzung bei den Jungen höher waren als bei den Mädchen. Im Bereich Sprache wurde hingegen leistungsmässig kein Geschlechtsunterschied festgestellt, aber die Mütterereinschätzung fiel bei den Mädchen signifikant höher aus, als bei den Jungen.

3. Diagnoseleistung im formellen Kontext: Zur Validität von intraindividuellen Profilanalysen

In Bezug auf die Aussagekraft von intraindividuellen Profilen bei mathematischen Schwierigkeiten zeigte sich, dass intraindividuelle Stärken und Schwächen über die Profilhöhe hinaus 8.8% der Varianz erklären konnten, was einem mittleren Effekt entspricht (Cohen, 1988) während sich die Gesamt-Profilhöhe als stärkster Prädiktor für die mathematischen Kompetenzen erwies.

Gesamt-Diskussion

Ziel der vorliegenden kumulativen Dissertation war es, den aktuellen Forschungsstand zur Akkuratheit von Elterneinschätzung bezüglich des Entwicklungsstandes ihres Kindes in relevanten Entwicklungsbereichen und der schulischen Kompetenzen Mathematik und Deutsch zu erweitern, sowie die Aussagekraft von professionell generierten Leistungsprofilen bezüglich möglicher Teilleistungsschwächen in Mathematik zu ermitteln. Dabei wurde der Fokus auf die Aussagekraft von informellen Diagnoseleistungen und auf die Aussagekraft von spezifischen formellen Diagnoseleistungen gelegt. Informelle und formelle Diagnoseleistungen sind Bestandteile von entwicklungspsychologischen oder – pädiatrischen Beratungssituationen. Ein erweitertes Wissen über die elterliche Wahrnehmung erlaubt es, Elternaussagen im Kontext von institutionellen Beratungsstellen besser zu verstehen und deren Gewicht bei Anamnesegesprächen so einzuordnen, dass sie in der Beratung optimal genutzt werden können. Gleichzeitig bildet das erweiterte Wissen über die Aussagekraft von Stärken und Schwächen in Bezug auf das individuelle Leistungsprofil des Kindes einen wichtigen Hintergrund bei der Planung von massgeschneiderten Fördermassnahmen, welche die Eltern verstehen und mittragen sollten.

In Bezug auf die in der vorliegenden Dissertation formulierten Fragestellungen weisen die Resultate der Studien 1 - 3 folgende Ergebnisse auf:

- 1) Die elterliche Einschätzung bei 3-5jährigen Kindern vermochte in der Mehrheit der Entwicklungs- und Funktionsbereiche einen Hinweis auf den jeweiligen Entwicklungsstand zu geben, wobei die Einschätzung des Gesamtentwicklungsstandes den Eltern am besten gelang. Die Elterneinschätzung erreichte jedoch nicht die Qualität eines Screeninginstruments.

2) Die mütterliche Einschätzungen bezüglich Leistungen in Mathematik und Deutsch bei Kindern im Alter von 5-10 Jahren korrelierten substantiell positiv mit den Testleistungen der Kinder. Überschätzungen kamen deutlich häufiger vor, als Unterschätzungen, wobei sich dieses Verhältnis im Hinblick auf das Geschlecht des Kindes nicht voneinander unterschied. Bei der mütterlichen Einschätzung war die Kategorie „überdurchschnittlich“ deutlich häufiger vertreten, als dies bei den tatsächlichen Testleistungen, die einer Normalverteilung entsprachen, der Fall war. Der prädiktive Wert der Mütterereinschätzungen unter Berücksichtigung des Bildungsstandes leistete einen signifikanten, wenn auch kleinen Beitrag bei der Vorhersage der Testleistungen. Ein signifikanter Geschlechtsunterschied konnte im Bereich Mathematik sowohl in der mütterlichen Einschätzung, als auch in der Testleistung zugunsten der Jungen nachgewiesen werden.

3) Sowohl die Interpretation von interindividuellen Profilhöhen, als auch die Interpretation des intraindividuellen Profils ist für Empfehlungen in der Praxis hilfreich. Nebst der allgemeinen Profilhöhe, welche sich als signifikantester Prädiktor für Mathematikleistungen auch bei hoher intraindividuellem Profilvarianz erwies, vermochte das intraindividuelle Profil, wenn auch nur in geringer Ausprägung, zusätzliche Information zu geben.

Für die in den Studien 1 und 2 nachgewiesene inakkurate Elterneinschätzung können sowohl das Modell von Semin und Papadopoulos (1989) als auch das Modell von Festinger (1954) einen Ansatz für mögliche Wirkungsmechanismen darstellen: Die Schwierigkeit das Ausmass eines Entwicklungs- oder Leistungsdefizits zu erkennen, könnte darauf beruhen, dass Eltern bei der Einschätzung in Folge der Mutter/Vater-Kind-Diade die Kompetenzen des Kindes nicht isoliert von sich selbst erleben und deshalb auch zu wenig genau einschätzen können, was das Kind tatsächlich kann. Je

grösser bei einem Kind der Entwicklungsrückstand im kognitiven und/oder im allgemeinen Bereich ist, umso grösser ist die Diskrepanz zur Referenzgrösse, wenn Eltern ihr Kind mit Kindern vergleichen, die keine Entwicklungsauffälligkeiten zeigen. Um die dadurch entstehende Stresssituation zu mildern könnte die Wahrnehmung dahingehend angepasst werden, dass die Diskrepanz als weniger gross eingeschätzt wird als sie unter Umständen ist. Zudem wird ein Kind mit Teilleistungsschwächen Defizite aufweisen, die den schulischen Erfolg hemmen können (Studie 3) was in einem Gesellschaftsbild, das sich kompetitiv orientiert und schulischen Erfolg honoriert wenig Platz hat. Zusammengenommen könnten diese Effekte die elterliche Diagnoseleistung dahingehend beeinflussen, dass eine Überschätzung der Kompetenzen des Kindes und eine Verkennung von tatsächlich vorhandenen Defiziten resultieren.

Weiter ist zu beachten, dass bei Kindern im Alter von 3-5 Jahren die Einschätzung der Gesamtentwicklung den Eltern etwas besser gelang als die Einschätzung einzelner Entwicklungs- oder Funktionsbereiche (Studie 1). Entsprechend zeigte sich bei der formellen Diagnose bei Kindern im Alter von 5-10 Jahren, dass der Gesamt-IQ in Bezug auf die mathematischen Kompetenzen am aussagekräftigsten war, während die intraindividuelle Profilanalyse deutlich weniger Varianz aufklären konnte (Studie 3). Möglicherweise können Kinder Schwächen auch mit Stärken ausgleichen, so dass einzelne Schwächen nicht zwangsläufig zu schlechten Schulleistungen in gewissen Fächern führen müssen. Ebenso könnten Kinder im Vorschulalter Schwächen bereits durch Stärken ausgleichen, weshalb Eltern vielleicht nicht erkennen, dass in einem isolierten Bereich eine Schwäche vorliegt. Gleichzeitig ist jedoch zu beachten, dass die Reliabilität eines Gesamtwertes immer höher ist, als diejenige von einzelnen Untertests, was sowohl bei den Elterneinschätzungen, wie auch beim standardisierten Test zum Tragen kommt.

Insgesamt zeigen die Resultate auf, dass die elterliche Wahrnehmung einzelner Kompetenzen und des Entwicklungsstandes des Kindes nur bedingt mit den Resultaten welche durch formelle Diagnostik erhoben werden, übereinstimmt. Zusätzlich zeigte sich am Beispiel der mathematischen Kompetenzen, dass die allgemeine Profilhöhe ein besserer Prädiktor für die Mathematikleistungen darstellt, als die Interpretation von intraindividuellen Profilanalysen. Im diagnostischen Beratungsalltag gilt es daher darauf achten, dass zum einen die intraindividuelle Profilanalyse einen Hinweis auf mögliche Schwächen geben kann, aber nicht überbewertet werden sollte und zum anderen, dass für die Eltern Schwächen in einzelnen Entwicklungs- oder Leistungsbereichen oder eine Verzögerung der Gesamtentwicklung eventuell nicht offensichtlich waren.

Stärken und Schwächen

Eine Stärke der vorliegenden Dissertation liegt darin, dass sowohl informelle, als auch formelle Diagnoseleistungen kritisch beleuchtet wurden, indem nicht nur die Aussagekraft der Elterneinschätzungen untersucht wurde, sondern auch die Aussagekraft von Interpretationen, welche im professionellen Kontext vorgenommen werden. Die Elterneinschätzung wurde gezielt im Hinblick auf spezifische Entwicklungs- und Funktionsbereiche, sowie auf schulisch relevante Leistungsbereiche überprüft. Zur Testung des Kindes wurden mit der IDS und IDS-P zwei umfangreiche gut validierte und reliable Testbatterien eingesetzt, welche sowohl Daten zum kognitiven, als auch zum allgemeinen Entwicklungsstand lieferten und damit eine gut abgestützte Aussage in beiden Bereichen und zur Gesamtentwicklung erlaubten. Zur Erfassung des Potentials des Kindes in den Studien 2 und 3 wurden auf diese Weise die Fähigkeiten in Mathematik und Sprache nicht mittels klassischer Schultests, sondern relativ curriculumsfrei mit einem standardisierten Verfahren erfasst. Dies liess eine generelle Bewertung der Fähigkeiten in diesen Leistungsbereichen zu. Die

angewendeten Tests differenzieren zuverlässig zwischen auffälligem und unauffälligem Entwicklungsstand (Manual IDS (2009) und IDS-P (2013)). Die verwendeten Stichproben stammen bei allen drei Studien aus Normierungsstudien, welche einem nicht-klinischen Setting entsprechen. Damit handelt es sich um umfangreiche Norm-Stichproben aus dem Grundschul- und Vorschulalter. Dies ist insofern von Bedeutung, dass viele Studien zur Elterneinschätzung mit Kindern im Vorschulalter auf einer Stichprobe basieren, welche in einem klinischen Setting erhoben wurde, wo im Gegensatz zu den in dieser Arbeit untersuchten Stichproben bereits ein Verdacht auf eine Entwicklungsverzögerung in einem oder mehreren Teilbereichen vorgelegen hatte, oder die Eltern evtl. durch Resultate von bereits erfolgten Leistungstests oder Entwicklungsscreenings beeinflusst waren. Eine weitere Stärke der vorliegenden Arbeit liegt in der praktischen Bedeutsamkeit der Fragestellung sowie der vorliegenden Befunde für den diagnostischen Alltag im Bereich der Entwicklungspsychologie und der Pädagogik.

Hinsichtlich Einschränkungen in der vorliegenden Arbeit sei darauf hingewiesen, dass für die Elterneinschätzungen nicht auf standardisierte Fragebögen zurückgegriffen werden konnte, was eine Herausforderung sowohl bezüglich der Item-Formulierungen, als auch bezüglich der Setzung eines Cut-offs darstellte. Des Weiteren stand den Eltern pro Bereich (Entwicklungsbereich oder Schulfach) nur ein Item zur Verfügung, weshalb kein Kennwert für die interne Konsistenz berechnet werden konnte. Für zukünftige Studien empfiehlt es sich, die Elterneinschätzung mittels mehrerer Items zu erfassen um die Reliabilität zu erhöhen. Worauf sich Eltern bei ihrer Einschätzung vorrangig stützen, konnte zwar nicht eruiert werden, die Resultate der Studie 2 konnten jedoch aufzeigen, dass Mütter bei Ihrer Einschätzung zum Teil vom Geschlecht des Kindes beeinflusst werden, und dass der Bildungsstand der Mutter teilweise bei der Einschätzung und auch

bei der kindlichen Leistung eine Rolle spielen kann. Des Weiteren muss beachtet werden, dass zur Definition des Entwicklungs- und Leistungsstandes des Kindes dessen Testleistungen in den IDS-P und IDS diene, wobei zu bemerken ist, dass bei Anwendung eines Testverfahrens immer auch Messfehler in Kauf genommen werden müssen. Bei allen drei Studien handelt es sich um einen Querschnittsvergleich. In Bezug auf die Elterneinschätzung (Studien 1 und 2) können damit Veränderungen in der elterlichen Wahrnehmung im Sinne einer zunehmenden Passung bei zunehmender Kompetenz des Kindes (Hunt & Paraskevopoulos, 1980) oder der Einfluss einer möglichen Verringerung der Mutter-Kind-Dyade (Semin und Papadopoulous, 1989) nicht erfasst werden und in Bezug auf Studie 3 konnte die Auswirkung von Profilveränderungen auf Lernschwierigkeiten nicht überprüft werden. Des Weiteren sei darauf hingewiesen, dass die Studien im deutschsprachigen europäischen Raum durchgeführt wurden und der kulturbedingte Hintergrund eine Rolle spielen kann.

Praktische Implikationen

Die positiven Korrelationen zwischen Elterneinschätzung und kindlicher Testleistung weisen darauf hin, dass Eltern zwar einen Hinweis geben können, auf den Entwicklungs- oder Leistungsstand des Kindes, jedoch aufgrund ungenügender Akkuratheit das Ausmass eines Leistungsdefizits zum grossen Teil nicht erkennen. Es muss davon ausgegangen werden, dass ein bedeutender Teil der Eltern nicht erkennen kann, ob ihr Kind in einem spezifischen Bereich oder in seiner Gesamtentwicklung eine Verzögerung aufweist, welche weiterführende professionelle Abklärungen erfordern würde. Es ist jedoch zu beachten, dass eine leichte Überschätzung seitens der Eltern auch entwicklungsförderlich wirken kann, indem sie ihrem Kind günstige Entwicklungsanreize anbieten, weil sie davon ausgehen, dass das Kind diese auch bewältigen kann (Deimann et al., 2005). Des Weiteren kann eine inakkurate

Einschätzung der Eltern, insbesondere der Mütter, auch einen protektiven Einfluss auf das Familiensystem haben (Willinger et al. 2011). Auf der anderen Seite birgt eine hohe Überschätzung seitens der Eltern die Gefahr, dass bedeutende Entwicklungsdefizite nicht rechtzeitig erkannt werden, das Kind mit überhöhten Anforderungen konfrontiert wird oder aus Mangel an Akzeptanz stützende Massnahmen von den Eltern nicht mitgetragen werden. Die Elterneinschätzung kann eine testpsychologische Einschätzung ergänzen, aber nicht ersetzen. In Bezug auf Rechenschwächen erweist sich eine möglichst frühe und insbesondere umfassende Diagnostik mittels allgemeiner Entwicklungstests wie beispielsweise mittels den IDS (Grob, et al., 2009) oder IDS-P (Grob et al. 2013) als sinnvoll, da Schwierigkeiten im Lernprozess mit unterschiedlichen Schwächen im kognitiven Bereich einhergehen können und die Erkennung derselben Hinweise zur einer möglichst angepassten Förderplanung geben kann. Auch wenn elterliche Beobachtungen in Bezug auf die Ausprägung eines Defizits wenig aussagekräftig sind, können sie dennoch wertvolle Hinweise geben auf Stärken und Schwächen. Im Sinne von sich ergänzenden Informationsquellen sollten in einer Beratungssituation übereinstimmende, aber auch divergente Beobachtungen möglichst genutzt werden damit einerseits eine adäquates Unterstützungssetting empfohlen werden kann und andererseits die elterliche Wahrnehmung dahingehend geschärft werden kann, dass Eltern zuhause ihr Kind im Sinne einer möglichst guten Passung seinen Stärken und Schwächen entsprechend begleiten, fördern und fordern können.

Konklusion und Ausblick

Eine zentrale Schlussfolgerung dieser Dissertation ist, dass Elterneinschätzungen kein valides Screening zur Erkennung von Defiziten in relevanten Entwicklungsdomänen darstellen, was eine Erklärung dafür sein kann, dass bei vielen Kindern Entwicklungsdefizite erst beim Übertritt ins Schulsystem erkannt werden

(Glascoe, 1999). In Bezug auf die schulischen Fächer Deutsch und Mathematik muss die Müttereneinschätzung ebenfalls als nur beschränkt valide angesehen werden. Die Anwendung von Entwicklungstests im Kontext einer formellen Diagnose zur Einschätzung des Entwicklungsstandes im Vorschulalter aber auch zur Einschätzung von schulischen Fähigkeiten ist bei Fragstellungen in der Praxis unerlässlich, zeigte sich doch, dass die Profilhöhe ein bedeutender Prädiktor von schulischen Leistungen ist. Darüber hinaus kann die von den Diagnostikern angewandte Interpretation von intraindividuellen Profilen dazu beitragen, die individuellen Förderbedürfnisse eines Kindes zu identifizieren, wobei die Aussagekraft des intraindividuellen Profils nicht überinterpretiert werden sollte. Zusammenfassend verdeutlichen die vorliegenden Ergebnisse, dass die Einschätzung kindlicher Entwicklung und Kompetenzen sowohl im informellen, als auch im formellen Kontext eine Herausforderung darstellt, sobald die Einschätzung abhängig ist von Interpretationen durch den Beurteiler/die Beurteilerin. Da Auswirkungen von Stärken und Schwächen abhängig davon sind, wie die Umwelt darauf reagiert (Jenni et al., 2011), sollten in entwicklungspsychologischen und – pädiatrischen Beratungen sowohl Beobachtungen und Wahrnehmungen der Eltern, wie auch die Interpretation des intraindividuellen Entwicklungsprofils angemessen integriert werden.

Mit einem längsschnittlichen Design könnte zukünftige Forschung die Veränderung elterlicher Wahrnehmung in Bezug auf den kindlichen Entwicklungsstand berücksichtigen, aber auch Veränderungen im Mikrosystem Kind, Familie, Schule (Bronfenbrenner 1979/1986) welche sich durch eine professionelle Diagnostik und daraus folgenden Fördermassnahmen einstellen kann.

Literatur

- Alloway, T. P. (2009). Working memory, but not IQ, predicts subsequent learning in children with learning difficulties. *European Journal of Psychological Assessment*, 25, 92-98. doi: 10.1027/1015-5759.25.2.92
- American Psychological Association. Standards for Educational and Psychological Tests. Washington, DC: American Psychological Association; 1985.
- Andersson, U. (2010). Skill development in different components of arithmetic and basic cognitive functions: Findings from a 3-year longitudinal study of children with different types of learning difficulties. *Journal of Educational Psychology*, 102, 115-134. doi: 10.1037/a0016838
- Aylward, G. P. (1997). Conceptual issues in developmental screening and assessment. *Journal of Developmental and Behavioral Pediatrics*, 18, 340-349.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.
- Bronfenbrenner, U. (1986). Ecology of the family as a context for human development: Research perspectives. *Developmental Psychology*, 22, 723-742. Gefunden auf <http://rinaldipsych.synthasite.com/resources/Bronfenbrenner.pdf>
- Buch, S. R., Sparfeldt, J. R. & Rost, D. H. (2006). Eltern beurteilen die Entwicklung ihrer hochbegabten Kinder. Sonderdruck aus: *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 38, 53-61.
- Bull, R., Espy, K. A. & Wiebe, S. A. (2008). Short-term memory, working memory and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, 33, 205-228.
- Byrnes, J. P. & Wasik, B. A. (2009). Factors predictive of mathematics achievement in kindergarten, first and third grades: An opportunity-propensity analysis.

- Contemporary Educational Psychology*, 34, 167-183.
doi:10.1016/j.cedpsych.2009.01.002
- Chen, Ch., Lin, Ch.-H., Wen, S.-H. & Wu, Ch.-H. (2007). How effectively do parents discern their children's cognitive deficits at a preschool age? *Journal of the Chinese Medical Association* 70 (10), 445-450.
- Chung, C.-Y., Liu, W.-Y., Chang, Ch.-J., Chen, C.-L., Tang, S. F.-T. & Wong, M.-K. (2010). The relationship between parental concerns and final diagnosis in children with developmental delay. *Journal of Child Neurology*, 26, 413-419.
- Cohen, J. (1988). *Statistical power for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Davis-Kean, P. E. (2005). The influence of parent education and family income of child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, 19, 294-304.
- Davison, M. L. & Kuang, H. (2000). Profiles patterns: Research and professional interpretation. *School Psychology Quarterly*, 15, 457-464.
- Deary, I. J., Strand, S., Smith, P. & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13-21.
- Deimann, P., Kastner-Koller, U., Benka, M., Kainz, S. & Schmidt, H. (2005). Mütter als Entwicklungsdiagnostikerinnen. Der Entwicklungsstand von Kindergartenkindern im Urteil ihrer Mütter. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37, 122-134.
- Ding, C. S. (2001). Profile analysis: Multidimensional scaling approach. *Practical Assessment, Research & Evaluation*, 7. Retrieved from <http://PAREonline.net/getvn.asp?v=7&n=16>

- Festinger, L. (1954). A Theory of Social Comparison Processes. *Human Relations*, 7: 117. doi: 10.1177/001872675400700202
- Floyd, R. G., Evans, J. J. & McGrew, K. S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools*, 40, 155-171.
- Fritz, A. & Ricken, G. (2005). Früherkennung von Kindern mit Schwierigkeiten im Erwerb von Rechenfertigkeiten. In M. Hasselhorn, H. Marx & W. Schneider (Hrsg.), *Diagnostik von Mathematikleistungen (S.5-27)*. Göttingen: Hogrefe.
- Furnham, A., (2000). Parents' estimates of their own and their children's multiple intelligences. *British Journal of Development Psychology*, 18, 583-594.
- Geary, D. C., Hoard, M. K., Nugent, L. & Bailey, D. H. (2012) Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology*, 104, 206-223. doi: 10.1037/a0025398
- Glascoe, F. P. (1994). It's not what it seems. The relationship between parents' concerns and children with global delays. *Clinical Pediatrics*, 33, 292-296.
- Glascoe, F. P. (1999). Using parents' concerns to detect and address developmental and behavioural problems. *Journal of the Society of Pediatric Nurses*, 4 (1), 24-35.
- Glascoe, F. P. & Dworkin, P. H. (1995). The role of parents in the detection of developmental and behavioural problems. *Pediatrics*, 95, 829-836.
- Glascoe, F. P. & Sandler, H. (1995). Value of parents' estimates of children's developmental ages. *The Journal of Pediatrics*, 127 (5), 831-835.
- Glutting, J. J., Watkins, M. W., Konold, T. R. & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-

- IV in estimating reading and math achievement on the WIAT-II. *The Journal of Special Education*, 40, 103-114. doi:10.1177/00224669060400020101
- Grob, A. & Jaschinski, U. (2003). *Erwachsen werden. Entwicklungspsychologie des Jugendalters*. Weinheim: Beltz.
- Grob, A., Hagmann-von Arx, P. & Bodmer, N. M. (2009). Entwicklungsdiagnostik. In S. Schneider & J. Margraf (Hrsg.), *Lehrbuch der Verhaltenstherapie, Band 3: Störungen im Kindes und Jugendalter* (S. 145-158). Heidelberg: Springer.
- Grob, A., Meyer, C. & Hagmann-von Arx, P. (2009). *Intelligence and Development Scales (IDS)*. Bern: Hans Huber.
- Grob, A., Reimann, A., Gut, J. & Frischknecht M.-C. (2013). *Intelligence and Development Scales – Preschool (IDS-P)*. Bern: Hans Huber.
- Halle, T. G., Kurtz-Costes, B., Mahoney, J. L., (1997). Family influences of school achievement in low-income african american children. *Journal of Educational Psychology*, 89, 527-537.
- Hascher, T. (2005). Diagnostizieren in der Schule. In A. Barth, C. Kloeft, J. Fabian, S. Huber, H. Rosenbusch & H. Sassenscheidt (Hrsg.), *PraxisWissen SchulLeitung* (S.1-8). Bonn: WoltersKluwer.
- Hunt, J. McV. & Paraskevopoulos, J. (1980). Children's psychological development as a function of inaccuracy of their mothers' knowledge of their abilities. *The Journal of Genetic Psychology*, 136, 285-298.
- Jenni, O., Benz, C. & Latal, B. (2011). Das Entwicklungsprofil: Wenn die kindliche Entwicklung nicht im Gleichschritt verläuft. *Pädiatrie up2date*, 2, 199-228. doi: 10.1055/s-0030-1256448
- Kahana, S. Y., Youngstrom, E. A. & Glutting, J. J. (2002). Factor and subtest discrepancies of the Differential Ability Scales: Examining prevalence and validity

- in predicting academic achievement. *Assessment*, 9, 82-93.
doi:10.1177.10731911102009001010
- Köbberling, J., Richter, K., Trampisch, HJ., & Windeler, J. (1991). *Methodologie der medizinischen Diagnostik. Entwicklung, Beurteilung und Anwendung von Diagnoseverfahren in der Medizin*. Berlin Heidelberg New-York: Springer-Verlag.
- Koch, H., Kastner-Koller, U., Deimann, P., Kossmeier, Ch., Koitz, C. & Steiner, M. (2011). The development of kindergarten children as evaluated by their kindergarten teachers and mothers. *Psychological Test and Assessment Modeling*, 53 (2), 214-257.
- Kovas, Y., Haworth, C. M. A., Petrill, S. A. & Plomin, R. (2007). Mathematical ability of 10-year-old boys and girls: Genetic and environmental etiology of typical and low performance. *Journal of Learning Disabilities*, 40, 554-567.
- Largo, R. H., Jenni, O. G. (2005). Das Zürcher Fit-Konzept. *Familiendynamik*, 30, 111-127.
- Lee, K. (2010). Do early academic achievement and behaviour problems predict long-term effects among Head Start children? *Children and Youth Services Review*, 32, 1690-1703.
- Lorenz, J. H. (2009). Diagnose und Prävention von Rechenschwäche als Herausforderung im Elementar- und Primarbereich. In A. Heinze & M. Grüssing (Hrsg), *Mathematiklernen vom Kindergarten bis zum Studium* (S. 35-45). Münster: Waxmann.
- Lummis, M. & Stevenson, H. W. (1990). Gender differences in beliefs and achievement: A cross-cultural study. *Development Psychology*, 26, 254-263.
- Malhi, P., Kashyap, S. & Dua, S. (2005). Maternal estimates of mental age in developmental assessment. *Indian Journal of Pediatrics*, 72, 931-934.

- Marx, P. & Lenhard, W. (2010). Diagnostische Merkmale von Screeningverfahren. In M. Hasselhorn & W. Schneider (Hrsg.), *Frühprognose schulischer Kompetenzen*. Göttingen : Hogrefe.
- Naglieri, J. A. (2000). Can profile analysis of ability test scores work? An illustration using the PASS theory and CAS with an unselected cohort. *School Psychology Quarterly*, *15*, 419-433. doi: 10.1037/h0088798
- Petermann, F., Stein, I. A. & Macha, M. (2006). *Entwicklungsdiagnostik mit dem ET 6-6*. (3. veränd. Aufl.). Frankfurt/M.:Harcourt Test Service GmbH.
- Pfeiffer, S. I., Reddy, L. A., Kletzel, J. E., Schmelder, E. R. & Boyer, L. M. (2000). The practitioner's view of IQ testing and profile analysis. *School Psychology Quarterly*, *15*, 376-385. doi:10.1037/h008795
- Phillipson, S. & Phillipson, S. N. (2007). Academic expectations, belief of ability, and involvement by parents as predictors of child achievement: A cross-cultural comparison. *Educational Psychology*, *27*, 329-348.
- Pomerantz, E. M. & Dong, W. (2006). Effects of mothers' preceptions of children's competence: The moderating role of mothers' theories of competence. *Developmental Psychology*, *24*, 950-961.
- Priller, J., (1996). *Anamnese, Exploration, Psychosoziale Diagnose: ein Lehrbuch*. Sankt Augustin: Asgard-Verlag Hippe.
- Pritchard, D. A., Livingston, R. B., Reynolds, C. R. & Moses, J. A. Jr. (2000). Modal profiles for the WISC-III. *School Psychology Quarterly*, *15*, 400-418. doi:10.1037/h0088797
- Pulsifer, M. B., Hoon, A. H., Palmer F. B., Gopalan, R. & Capute A. J. (1994). Maternal estimates of developmental age in preschool children. *The Journal of Pediatrics*, *125* (1), 18-24.

- Rätty, H., Vänskä, J., Kasanen, K. & Kärkkäinen R. (2002). Parents' explanations of their child's performance in mathematics and reading: A replication and extension of Yee and Eccles. *Sex Roles*, 46, 121-128.
- Rennen-Allhoff, B. (1991). Wie verlässlich sind Elternangaben? *Praxis der Kinderpsychologie und Kinderpsychiatrie*, 40, 333-338.
- Rennen-Allhoff, B., Allhoff, P., Bowi, U. & Laaser, U., (1993). *Elternbeteiligung bei Entwicklungsdiagnostik und Vorsorge: Verfahren zur Früherkennung von Krankheiten und Entwicklungsstörungen im Säuglings- und Kindesalter durch Eltern*. Juventa Verlag Weinheim und München.
- Schrader, F.-W. (2006). Diagnostische Kompetenzen von Eltern und Lehrern. In D.H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3., überarb. und erweit. Aufl., S. 95-100). Weinheim: Psychologie Verlag Union.
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R. & Nores, M., (2005). *Lifetime effects: The HighScope Perry Preschool Study Trough Age 40*. Ypsilanti, MI: HighScope Press.
- Semin, G. R. & Papadopoulou, K. (1989). In G. Duveen & B. Lloyd (Eds.), *Social Representation and the Development of Knowledge* (pp. 107-126). Cambridge: CUP.
- Storck, M., Webel, D., Steinmacher, J., Kupferschmidt, C. & Bode, H. (1998). Diagnostik umschriebener Entwicklungsstörungen im Vorschulalter. Zur Validität elterlicher Angaben. *Monatsschrift Kinderheilkunde*, 146, 988-992.
- Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology*, 92, 144-51.

- Tröster, H., Flender, J. & Reineke, D., (2005). Entwicklungsscreening. Dortmunder Entwicklungsscreening für den Kindergarten (DESK 3-6). *Kindheit und Entwicklung* 14 (3), 140-149.
- von Aster, M., Schweiter, M. & Weinhold Zulauf, M. (2007). Rechenstörungen bei Kindern: Vorläufer, Prävalenz und psychische Symptome. *Zeitschrift für Entwicklungspsychologie*, 39, 85-96. doi:10.1026/0049-8637.39.2.85
- Walker & Wiske (1981). In Rennen-Allhoff, B., Allhoff, P., Bowi, U., & Laaser, U., (1993). *Elternbeteiligung bei Entwicklungsdiagnostik und Vorsorge: Verfahren zur Früherkennung von Krankheiten und Entwicklungsstörungen im Säuglings- und Kindesalter durch Eltern.* (S. 9-21). Juventa Verlag Weinheim und München.
- Waschbusch, D. A., Daleiden, E. & Drabman, R. S. (2000). Are parents accurate reporters of their child's cognitive abilities? *Journal of Psychopathology and Behavioral Assessment*, 22, 61-77.
- Watkins, M. W. & Glutting, J. J. (2000). Incremental validity of WISC-III profile elevation, scatter, and shape information for predicting reading and math achievement. *Psychological Assessment*, 12, 402-408. doi:101037//1040-3590.12.4.402
- Willinger, U. & Eisenwort, B. (2005). Mothers' estimates of their children with disorders of language development. *Behavioral Medicine*, 31(3), 117-126.
- Willinger, U., Schaunig, I., Jantscher, S., Schmoeger, M., Loader, B., Kummer, Ch. et al. (2011), Mothers' estimates of their preschool children and parenting stress. *Psychological Test and Assessment Modeling*, 53 (2), 228-240.

Anhang

Studie 1

Frischknecht, M.-C., Reimann, G. & Grob, A. (submitted). Erkennen Eltern Entwicklungsdefizite im Vorschulalter? Zur Akkuratheit elterlicher Einschätzungen kindlicher Entwicklung. Eingereicht bei *Kindheit und Entwicklung* Version vom 08. April 2014

Studie 2

Frischknecht, M.-C., Reimann, G., Gut, J., Ledermann, T. & Grob, A., (2014). Wie genau können Mütter die Mathematik- und Sprachleistungen ihrer Kinder einschätzen? Ein Vergleich zwischen Müttereinschätzung und Testleistungen bei Kindern im Alter von 6-10 Jahren. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 46 (2), 67-78.

Studie 3

Reimann, G., Stoecklin, M., Lavalley, K., Gut, J., Frischknecht, M.-C. & Grob, A. (2013). Cognitive and motivational profile shape predicts mathematical skills over and above profile level. *Psychology in the Schools*, 50(1). doi: 10.1002/pits.21659

Publikationen

Curriculum Vitae

Selbständigkeitserklärung

Studie 1

Frischknecht, M.-C., Reimann, G. & Grob, A. (submitted). Erkennen Eltern Entwicklungsdefizite im Vorschulalter? Zur Akkuratheit elterlicher Einschätzungen kindlicher Entwicklung. Eingereicht bei *Kindheit und Entwicklung*.

Version vom 08. April 2014

Erkennen Eltern Entwicklungsdefizite im Vorschulalter?

Zur Akkuratheit elterlicher Einschätzungen kindlicher Entwicklung

Marie-Claire Frischknecht, Giselle Reimann und Alexander Grob

Universität Basel

Zusammenfassung

Es wurde untersucht, ob Elterneinschätzungen betreffend dem Entwicklungsstand ihrer 3- bis 5jährigen Kinder ($N = 539$; $M = 4;6$ Jahre, $SD = 0;10$ Jahre) die Kriterien für ein akkurates Screening erfüllen in Bezug auf hinreichende Sensitivität, Spezifität, RATZ-Index und AUC-Werte. Dazu wurden in einer nicht-klinischen Stichprobe die Elterneinschätzungen zum Entwicklungsstand der Kinder mit deren Testleistungen in den Intelligence and Development Scales – Preschool (IDS-P) verglichen. Es ergaben sich substantiell positive Korrelationen zwischen Elterneinschätzungen und Testleistungen. Allerdings waren die Elterneinschätzungen nicht genügend valide, um in einem spezifischen Bereich oder in der Gesamtentwicklung eine Verzögerung festzustellen. Die Elterneinschätzungen erfüllten damit die Kriterien eines akkuraten Screenings nicht. Um Entwicklungsdefizite möglichst frühzeitig zu erkennen, ist eine standardisierte Entwicklungsdiagnostik unabdingbar.

Schlüsselwörter: Entwicklungsscreening, Entwicklungsdiagnostik, Vorschulkinder, Elterneinschätzung

Do parents realize deficits in child development at preschool age?

Accuracy of parent estimation concerning child development

Abstract

This study investigated whether parent estimations of their children's ($N = 539$; $M = 4;6$ years, $SD = 0;10$ years) development are a suitable screening instrument in terms of specificity, sensitivity, RAZ-index, AUC measures. Parent estimations were compared to children's results in a standardized test, i.e. the Intelligence and Development Scales – Preschool (IDS-P). Although substantial positive correlations were found, parent estimations could not be proved to be valid enough to detect serious developmental delays with respect to a specific domain or with respect to the whole development. Parent estimations do not serve as a suitable screening instrument. Standardized assessment instruments are necessary in order to identify developmental delays at an early stage.

Key words: developmental screening, assessment of development, preschool children, parent estimation

Erkennen Eltern Entwicklungsdefizite im Vorschulalter?

Zur Akkuratheit elterlicher Einschätzung kindlicher Entwicklung

Die Bedeutung der Früherkennung entwicklungsgefährdeter Kinder wird in der Fachliteratur als zentral erachtet. So können Investitionen in die Frühförderung für einen Staat bedeutende Einsparungen nach sich ziehen (Schweinhart, Montie, Xiang et al., 2005). Eine Entwicklungsauffälligkeit wird am ehesten bei steigenden Anforderungen sichtbar, vielfach im Kontext normativer Transitionen wie dem Kindergarten- oder Schulübertritt (Rutter, 1994). Es ist daher wenig überraschend, dass die Hälfte der entwicklungsauffälligen Kinder erst mit Eintritt ins Schulsystem erkannt wird (Glascoe, 1999). Wenn förderbedürftige Kinder erst bei der Schuleingangsuntersuchung entdeckt werden, haben sie das Alter für eine optimale Frühförderung überschritten (Böhm & Kuhn, 2000). Entwicklungsdefizite sollten also möglichst bereits vor dem Kindergarten- oder Schuleintritt erkannt werden. In diesem Zeitraum sind es die Eltern, welche die meisten Gelegenheiten haben zu beobachten, ob Meilensteine der Entwicklung von ihren Kindern in etwa rechtzeitig erreicht werden. Die Tatsache, dass Entwicklungsauffälligkeiten zum Teil erst bei Schul- oder Kindergarteneintritt festgestellt werden (Glascoe, 1999; Tröster, Flender & Reineke, 2005), lässt vermuten, dass die elterliche Beurteilung des Entwicklungsstandes ihres Kindes verzerrt ist und Abweichungen von der Regelentwicklung von ihnen nicht mit genügender Zuverlässigkeit erkannt werden. Die vorliegende Studie prüft, wie akkurat Elterneinschätzungen sind und ob es den Eltern möglich ist, durch Beobachtungen im Alltag den Entwicklungsstand ihres Kindes adäquat zu beurteilen, dass sie im Sinne eines Screenings einen Beitrag zur

Früherkennung von Entwicklungsdefiziten leisten können.

Da für ein Entwicklungsscreening generell eine Sensitivität ab 70% (Prozentsatz der Kinder mit tatsächlich vorliegender Entwicklungsverzögerung und korrekter elterlicher Identifikation) als akzeptabel und eine Spezifität von 70-80 % (Prozentsatz der Kinder ohne Entwicklungsverzögerung, die von den Eltern korrekt identifiziert werden) als angemessen erachtet werden, sowie der Positiv Prädiktive Wert (PPV: Prozentsatz der Kinder, die von den Eltern als entwicklungsauffällig eingeschätzt wurden, und auch tatsächlich eine Entwicklungsverzögerung aufweisen) bei mindestens 50 % liegen sollte (American Psychological Association, 1985; Aylward, 1997), müssten Elterneinschätzungen diesbezüglich ebenso hohe Werte ausweisen, um als valides Screeningverfahren in Betracht gezogen zu werden.

In der Literatur werden unterschiedliche Werte bezüglich Korrelation, Sensitivität, Spezifität und PPV für Elterneinschätzungen berichtet. Da in aller Regel nicht standardisierte Fragebögen zur Anwendung kommen, bleibt unklar, ob dies mit den unterschiedlichen Formulierungen bei der Elternbefragung zusammenhängen könnte. Nicht nur die Item-Formulierung und die Antwortmöglichkeiten für die Eltern, sondern auch der Cut-off zur Unterscheidung von unauffälliger versus auffälliger Entwicklung bei den Elternantworten unterscheidet sich je nach Studie. Unterschiede gibt es weiters hinsichtlich der untersuchten Stichproben: In Bezug auf die obengenannten Masse kommen Untersuchungen mit nicht-klinischen Stichproben zu weniger befriedigenden Werten (Koch, Kastner-Koller, Deimann et al., 2011; Storck, Webel, Steinmacher et al., 1998) als Studien mit klinischen Stichproben (z.B. Chen, Lin, Wen et al., 2007; Chung, Liu, Chang et al., 2010; Glascoe &

Sandler, 1995; Malhi, Kashyap & Dua, 2005; Pulsifer, Hoon, Plamer et al., 1994; Willinger & Eisenwort, 2005). Bei Letzteren wurden allerdings — im Gegensatz zu Studien mit nicht-klinischen Stichproben — nur diejenigen Elterneinschätzungen erfasst, bei denen bereits eine gewisse Sensibilität auf mögliche Entwicklungsabweichungen vorhanden war.

Bislang gibt es wenige Studien bei 3- bis 5jährigen Kindern, welche Elterneinschätzungen auf ihre Akkuratheit und damit auf die korrelativen Zusammenhänge und die Sensitivität, Spezifität und PPV betreffend möglicher Entwicklungsdefizite in einem nicht-klinischen Kontext gezielt in mehreren Funktionsbereichen und bezüglich der Gesamtentwicklung überprüfen. Die bisherig publizierten Studien konnten mehrheitlich keine substantiellen Zusammenhänge zwischen Elterneinschätzung und Testergebnissen nachweisen. Koch et al. (2011) konnten lediglich in drei von insgesamt 14 untersuchten Bereichen sowie bei der Einschätzung zur Gesamtentwicklung statistisch signifikante Korrelationen von .41 bis .51 nachweisen. Storck et al. (1998) berichten Korrelationen von .23 bis .27 für zwei von insgesamt fünf untersuchten Funktionsbereichen. In Bezug auf die Sensitivität und Spezifität zeigte diese Studie, dass die Eltern unabhängig von der spezifischen Teilfunktion nur ca. 30% der entwicklungsauffälligen Kinder korrekt identifizierten. Glascoe (1994) berichtet bezüglich Einschätzung zur Gesamtentwicklung eine Sensitivität von 22% und eine Spezifität von 47%, obwohl davon ausgegangen werden kann, dass Eltern regelentwickelter Kinder den Entwicklungsstand weniger überschätzen als Eltern entwicklungsauffälliger Kinder (Deimann & Kastner-Koller, 2011; Willinger et al., 2011). Auch wenn eine Diskrepanz zur Regelentwicklung festgestellt wird, führt dies nicht zwangsläufig zu einer höheren Sensitivität, wie eine Untersuchung von Willinger und Eisenwort (2005) aufzeigt: Obwohl die Mütter

einen Verdacht auf Entwicklungsverzögerungen im sprachlichen Bereich geäußert hatten, lag die Sensitivität hierzu bei lediglich 4%. Bei einem Vergleich der Einschätzungen von Müttern sozial auffälliger Kinder mit den Einschätzungen von Müttern verhaltensunauffälliger Kinder zeigte sich, dass Erstere ihre Kinder in einzelnen Funktionsbereichen häufiger überschätzten (Deimann, Kastner-Koller, Benka et al., 2005). Störungen des Sozialverhaltens gehen häufig auch mit kognitiven, sprachlichen und motorischen Problemen einher. Die stärkere Überschätzung einzelner Funktionsbereiche bei verhaltensauffälligen Kindern kann darauf hinweisen, dass Entwicklungsverzögerungen nicht als solche erkannt werden. Dies könnte einer der Gründe sein, weshalb Kinder mit Entwicklungsdefiziten bis zum Schul- oder Kindergarteneintritt unerkant bleiben.

Die betreffenden Studien beschränkten sich entweder auf kleine Stichproben mit weniger als 50 Kindern (Deimann et al., 2005; Deimann & Kastner-Koller, 2011; Koch et al., 2011) oder auf ausgewählte Entwicklungsbereiche (Storck et al., 1998; Willinger et al., 2011). Ebenso wurde die elterliche Einschätzung zur Gesamtentwicklung des Kindes relativ selten und mehrheitlich in klinischen Stichproben untersucht (Chung et al., 2010; Glascoe & Sandler, 1995; Pulsifer et al., 1994).

Es sei abschliessend darauf hingewiesen, dass sich in diversen Studien keiner der möglichen Einflussfaktoren für die Akkuratheit mütterlicher Einschätzung, nämlich Alter und Geschlecht des Kindes, Anzahl Geschwister, Geburtsrang, Dauer des Kindergartenbesuchs sowie Alter und Bildungsstand der Mutter, als relevant erwies (Deimann et al., 2005; Glascoe, 1999; Malhi et al., 2005; Williger & Eisenwort, 2005).

Fragestellung

Im Fokus des Interesses steht in der vorliegenden Studie die Einschätzung der Eltern zum Entwicklungsstand ihrer Kinder einer nicht-klinischen Stichprobe. Es soll der Frage nachgegangen werden, ob Eltern von 3- bis 5jährigen Kindern basierend auf ihren täglichen Beobachtungen erkennen können, ob bei ihrem Kind bezüglich der Gesamtentwicklung oder in Teilbereichen ein Entwicklungsdefizit vorliegt.

Die grosse Heterogenität der Item-Formulierungen für die Elterneinschätzung erschwert die Vergleichbarkeit bisheriger Forschungsergebnisse. Es ist möglich, dass die unterschiedlichen Forschungsergebnisse zumindest teilweise auf unterschiedliche Frageformulierungen zurückzuführen sind. In der vorliegenden Studie werden die Eltern anhand von konkreten, alltagsnahen Beobachtungsbeispielen dazu aufgefordert, eine altersbezogene Einschätzung zum Entwicklungsstand ihres Kindes abzugeben. Diese Itemformulierung wurde gewählt, weil Eltern ihr Kind im natürlichen Alltagssetting beobachten und sein Verhalten mit jenem von Gleichaltrigen vergleichen (Chen et al., 2007; Glascoe, 1999; Glascoe & Dworkin, 1995; Glascoe & Sandler, 1995). Aufgrund der bisherigen Befunde (Glascoe & Dworkin, 1995; Glascoe & Sandler, 1995; Malhi et al., 2005) erwarten wir, dass mit dieser Itemformulierung positive Korrelationen zwischen Elterneinschätzung und Testleistungen nachgewiesen werden können. Zur Einschätzung der Gesamtentwicklung finden sich in der Literatur Hinweise auf signifikante Korrelationen (Glascoe & Sandler, 1995; Koch et al., 2011). Wir nehmen also an, dass

1. Elterneinschätzungen auf Basis von Alltagsbeispielen substantiell positiv mit Testresultaten zu den entsprechenden Bereichen korrelieren, und dass
2. die Elterneinschätzung zum Gesamtentwicklungsstand substantiell positiv mit dem

Testresultat zur Gesamtentwicklung korreliert.

Je nach Studie wurde bei den Elterneinschätzungen bisher unterschiedlich definiert, welche Elternantworten auf einen auffälligen Entwicklungsstand beim Kind hinweisen. Dies gilt sowohl für den Cut-off als auch bei der sprachlichen Formulierung in der Antwortskala. Hierbei besteht insbesondere die Schwierigkeit, dass Eltern den Wortlaut einer gewählten Antwortmöglichkeit eventuell nicht gleich interpretieren, wie dies Experten tun (Glascoe & Dworkin, 1995). In der vorliegenden Studie wird deshalb der Cut-off zur Definition von unauffälliger versus auffälliger Entwicklung bei einer Standardabweichung unterhalb des Mittelwertes der Elternantworten selbst festgelegt, womit die Interpretation der Elternantworten auf sprachlicher Ebene umgangen wird (Storck et al., 1998).

Da insbesondere bei nicht-klinischen Stichproben mehrheitlich niedrige Sensitivitätsmasse erreicht wurden, erwarten wir aufgrund bisheriger Evidenz, dass die Elterneinschätzung sowohl bezüglich einzelner Entwicklungsbereiche, als auch bezüglich des Gesamtentwicklungsstandes zu wenig sensitiv ist, um auf Verzögerungen hinzuweisen. Wir nehmen also an, dass:

3. die Elternbeobachtung anhand von Alltagsbeispielen bezüglich einzelner Bereiche und die Einschätzung bezüglich dem Gesamtentwicklungsstand nicht ausreichend ist (Sensitivitätsmass $< 70\%$) um zu erkennen, ob ein Entwicklungsbereich oder der Gesamtentwicklungsstand in der Bandbreite der Norm liegen.

Methode

Stichprobenbeschreibung

Die Stichprobe besteht aus einer Teilmenge der Normierungsstichprobe der IDS-P

(Intelligence and Development Scales–Preschool; Grob, Reimann, Gut & Frischknecht, 2013). Die IDS-P-Normierung wurde mit 713 Kindern im Alter von 3;0 bis 5;11 Jahren ($M = 4;6$ Jahre, $SD = 0;10$ Jahre) in der Schweiz, Deutschland und Österreich im Einzelsetting durchgeführt. Parallel dazu bearbeitete die grosse Mehrheit der Eltern einen Fragebogen zum Entwicklungsstand ihres Kindes. In die vorliegende Studie wurden die Kinder einbezogen, für die der von den Eltern beantwortete Fragebogen vorlag. Daraus resultierte eine Stichprobe von $N = 539$ (48 % Mädchen; Erhebungskontexte: 53 % Schweiz; 21 % Deutschland; 26 % Österreich) im Alter von 3;0 bis 5;11 Jahren ($M = 4;6$ Jahre, $SD = 0;10$ Jahre). Fünf Prozent der Mütter und 4 % Prozent der Väter gaben an, über keinen Berufsabschluss zu verfügen, 40 % der Mütter und 39 % der Väter hatten eine Berufsausbildung, 6 % der Mütter und 4 % der Väter eine Berufsbildung mit Zusatzausbildung, 4 % der Mütter und 5 % der Väter eine höhere Berufsbildung und 40 % der Mütter und 44 % der Väter einen Hochschulabschluss. Bei 5 % der Mütter und bei 4 % der Väter fehlten die Angaben zur Berufsausbildung.

Untersuchungsinstrumente

IDS-P. Die kognitive Entwicklung wird in den IDS-P mit sieben Untertests erfasst: Wahrnehmung (Visuell), Aufmerksamkeit (Selektiv), Gedächtnis (Phonologisch, Räumlich-Visuell, Auditiv) und Denken (Bildlich, Konzeptuell). Die Allgemeine Entwicklung wird mit Untertests für Psychomotorik, Sozial-Emotionale Kompetenz, Denken Logisch-Mathematisch und Sprachentwicklung gemessen. Zur Erhebung des Gesamtentwicklungsstandes wurde aus dem Mittelwert aller Untertests eine neue Variable „*Gesamtentwicklung*“ gebildet. Das Testverfahren unterscheidet zuverlässig zwischen normalentwickelten Kindern und verschiedenen klinischen Gruppen, insbesondere Kindern

mit Trisomie 21, Sprachentwicklungsauffälligkeiten nach ICD 10, Allgemeinen Entwicklungsauffälligkeiten nach ICD 10 (F83), frühgeborenen Kindern und fremdsprachigen Kindern. Zur Berechnung der Kriteriumsvalidität wurden Vergleiche mit den RIAS (Reynolds Intellectual Assessment Scales; Hagmann von Arx & Grob, 2014), dem K-ABC Rechentest (Kaufmann Assessment Battery for Children; Melchers & Preuss, 2009) und dem SETK 3-5 (Sprachentwicklungstest für drei- bis fünfjährige Kinder; Grimm 2001) durchgeführt. Genauere Angaben zum Nachweis der Validität und der Reliabilität der IDS-P finden sich im Manual (Grob et al., 2013). Eine Beschreibung der Untertests inklusive Kennwerte für die internen Konsistenzen für die vorliegende Teilstichprobe findet sich im *Anhang*.

Elterneinschätzung – Item-Formulierung. Für die Elterneinschätzung wurde für jeden Untertest ein entsprechendes Verhalten des Kindes beschrieben, welches die Eltern im Alltag beobachten können. Es wurde darauf geachtet, dass das Item mit den Alltagsbeispielen möglichst genau die geforderten Leistungen im entsprechenden Untertest abbildet. So wurden die Eltern zum Beispiel zur Einschätzung *Wahrnehmung Visuell* dazu aufgefordert, folgende Einschätzung zu geben: „Mein Kind kann Längen und Grössen von Gegenständen einschätzen und vergleichen. Beispiele: Beim Spielen mit Bauklötzen, Duplo- oder Legosteinen Längen und Grössen der Bausteine mit dem Auge einschätzen; Lücken mit dem richtigen Baustein füllen, ohne auszuprobieren; Bausteine gezielt nach Grösse auswählen; Fahrzeuge, Stifte, Figuren oder Puppen der Grösse oder Länge nach ordnen etc.“. Die Formulierungen wurden einem Expertenrating unterzogen. Zur Beantwortung stand eine siebenstufige Ratingskala zur Verfügung, wobei 7 = „mein Kind kann das aussergewöhnlich gut für sein Alter“, 6 = „mein

Kind kann das sehr gut für sein Alter“, 5 = „mein Kind kann das gut für sein Alter“, 4 = „mein Kind kann das so, wie ich es von seinem Alter her erwarte“, 3 = „mein Kind kann das eher nicht so gut für sein Alter“, 2 = „mein Kind kann das nicht gut für sein Alter“, 1 = „mein Kind kann das überhaupt nicht gut für sein Alter“ oder 0 = „das kann ich nicht beantworten“ bedeutete. Zusätzlich schätzten die Eltern den Gesamtentwicklungsstand wie folgt ein: „Ich glaube, in seiner Gesamtentwicklung ist mein Kind (7) aussergewöhnlich weit für sein Alter, (6) sehr weit für sein Alter, (5) eher weit für sein Alter, (4) gerade seinem Alter entsprechend, (3) etwas zurück für sein Alter, (2) ziemlich zurück für sein Alter, (1) stark zurück für sein Alter, (0) das kann ich nicht beantworten“.

Analysemethoden

Die Zusammenhangshypothesen wurden mit Pearson-Korrelationen getestet. Um die Sensitivität, Spezifität und PPV der Elterneinschätzung zu bestimmen, wurde die Stichprobe dichotomisiert, indem über jede Dimension der Eltern- und IDS-P-Skalen separat eine „entwicklungsauffällige“ und eine „nicht-auffällige“ Gruppe gebildet wurde. Entsprechend dem Manual der IDS-P (Grob et al., 2013), wonach Wertpunkte von 7-13 als durchschnittliche Leistung (altersstandardisierter Mittelwert = 10 Wertpunkte; 1 *SD* = 3 Wertpunkte) und Wertpunkte < 7 als unterdurchschnittliche, solche > 13 als überdurchschnittliche Leistung definiert sind, wurden die Variablen aller Untertests sowie die Variable *Gesamtentwicklung* neu in zwei Kategorien „entwicklungsauffällig“ = unterdurchschnittlich (Wertpunkte 1-6) und „nicht-auffällig“ = durchschnittlich und überdurchschnittlich (Wertpunkte 7-18) unterteilt. Die Elternantworten wurden anhand des Kolmogorow-Smirnov-Z-Tests auf Normalverteilung hin geprüft, und anschliessend wurde bei den siebenstufigen Antwortskalen der

Elterneinschätzungen der Mittelwert für jeden einzelnen erfragten Bereich bestimmt.

Elterneinschätzungen, die den Mittelwert um eine Standardabweichung unterschritten, wurden als „entwicklungsauffällig“ gewertet. Elterneinschätzungen, welche dem Mittelwert (+/- 1 SD) oder darüber entsprachen, galten als „nicht auffällig“. Die Sensitivität und Spezifität wurde zusätzlich mit dem RATZ-Index (Relativer Anstieg der Trefferquote gegenüber der Zufallstrefferquote, Marx & Lenhard, 2010) überprüft, wonach Werte ab 0.3 als eine leichte Verbesserung der Vorhersagegüte und Werte ab 0.6 als eine deutliche Verbesserung eingestuft werden. Zusätzlich wurde eine Receiver Operating Characteristic (ROC) Kurven-Analyse (Grenzwertoptimierungskurve, Köbberling, Richter, Trampisch et al., 1991) durchgeführt, bei welcher das Verhältnis von Sensitivität und Spezifität mittels dem Wert für den Bereich unter der Kurve (AUC-Wert = *area under the curve*) dargestellt wird und – im hier gegebenen Fall – die Akkuratheit der Elterneinschätzungen angezeigt werden kann. Swets (1988) empfiehlt, AUC-Werte von 0.5 bis ≤ 0.7 ($p < 0.05$) als gering, Werte von 0.7 bis ≤ 0.9 ($p < 0.05$) als moderat und Werte von 0.9 bis 1 ($p < 0.05$) als hoch/optimal zu interpretieren. Flächen, welche einen Wert unter 0.5 aufweisen, gelten als ungeeignet für ein Screening. Sämtliche Analysen wurden mit SPSS Version 20 durchgeführt.

Resultate

Die Korrelation zwischen Testleistung und elterlicher Einschätzung war in 10 der insgesamt 14 Entwicklungsbereiche signifikant (** $p < .01$, * $p < .05$) und zwar für *Wahrnehmung Visuell* ($r = .10^*$), *Aufmerksamkeit Selektiv* ($r = .12^{**}$), *Gedächtnis Phonologisch* ($r = .29^{**}$), *Denken Bildlich* ($r = .18^{**}$), *Denken Konzeptuell* ($r = .11^*$),

Grobmotorik ($r = .16^{**}$), *Feinmotorik* ($r = .16^{**}$), *Visuomotorik* ($r = .28^{**}$), *Denken Logisch-Mathematisch* ($r = .17^{**}$) und *Sprache Expressiv* ($r = .13^{**}$). Für *Sozial-Emotionale Kompetenz*, *Sprache Rezeptiv*, *Gedächtnis Räumlich-Visuell* und *Gedächtnis Auditiv* waren die Korrelationen nicht signifikant. Die Korrelationen zwischen elterlicher Entwicklungseinschätzung und tatsächlicher Leistung des Kindes können durchwegs als gering bezeichnet werden. Die Einschätzung der Gesamtentwicklung korrelierte signifikant ($r = .31$) mit der Leistung „*Gesamtentwicklung*“.

Die Resultate zur Sensitivität, Spezifität sowie zum Positiv Prädiktiven Vorhersagewert (PPV) und RAZ-Index sind in Tabelle 1 aufgeführt. „Tabelle 1 hier einfügen“

Alle Sensitivitätsmasse lagen unter dem für ein akzeptables Screening geforderten Wert von 70 % (American Psychological Association, 1985) und können deshalb als durchwegs ungenügend bezeichnet werden.

In allen Bereichen wurden die für ein akzeptables Screening geforderten Spezifitätswerte ($\geq 70\%$, American Psychological Association, 1985; Aylward, 1997) erzielt.

Ein Positiv Prädiktiver Wert von mindestens 50 % konnte in keinem Bereich nachgewiesen werden. Während bei der Einschätzung der einzelnen Funktionsbereiche kein RAZ-Index $> .3$ identifiziert wurde, konnte bei der Einschätzung der Gesamtentwicklung mit einem RAZ-Index von .529 ein akzeptabler Wert festgestellt werden. Die ROC-Analyse ergab bei allen einzelnen Funktionsbereichen AUC-Werte, die tiefer als der von Swets (1988) als akzeptabel vorgeschlagenen Wert ausfielen. In Bezug auf die Gesamtentwicklung wurde mit einem AUC-Wert von .825 (KI = .683 - .966, $p < .001$) ein Wert erreicht, der im

moderaten Bereich (Swets,1988) liegt und akzeptabel ist.

Diskussion

Es wurde untersucht, ob zwischen Elterneinschätzung und Testleistungen des Kindes statistisch bedeutsame Korrelationen bestehen und ob Elterneinschätzungen die Kriterien eines Screeninginstrumentes für relevante Entwicklungsbereiche erreichen können.

Die Elterneinschätzung korrelierte in dieser grossen nicht-klinischen Stichprobe fast durchweg positiv mit der Testleistung des Kindes. Dieses Resultat ist insofern überraschend, als dass bisher bei nicht-klinischen Studien nur vereinzelt signifikante Korrelationen zwischen Elterneinschätzungen und Testresultaten bei Kindern in diesem Alter aufgezeigt werden konnten (Koch et al., 2011; Storck et al., 1998). Hingegen lagen in den Bereichen *Sozial-Emotionale Kompetenz* und *Sprache Rezeptiv, Gedächtnis Auditiv und Gedächtnis Räumlich-Visuell* keine signifikanten Korrelationen vor. Die interne Konsistenz der Untertests *Sozial-Emotionale Kompetenz* und *Sprache Rezeptiv* ist zwar geringer als jene der anderen Untertests, liegt aber im akzeptablen Bereich (Evers, 2001). Im Hinblick darauf, dass eine altersentsprechende sozial-emotionale Entwicklung als Grundlage für weitere Entwicklungsschritte in anderen Entwicklungsbereichen angesehen wird (Wiedebusch & Petermann, 2011), wäre eine akkurate Einschätzung in diesem Bereich jedoch besonders relevant. Die Schwierigkeit, den Stand der rezeptiven Sprache oder auch das auditive Gedächtnis einzuschätzen, könnte auch damit zusammenhängen, dass Eltern nicht unbedingt an ein Entwicklungsdefizit denken, wenn ihr Kind sprachlich formulierte Anweisungen nicht befolgt, sondern dies eher auf fehlenden Willen oder mangelnde Kooperation zurückführen (Glascoe, 1994).

Des Weiteren wurde in der vorliegenden Studie untersucht, wie hoch die Sensitivität von Elterneinschätzungen ist. Eine zentrale Schlussfolgerung unserer Studie ist, dass Elterneinschätzungen kein valides Screening zur Erkennung von Defiziten in relevanten Entwicklungsdomänen ist. Diese Schlussfolgerung ziehen wir, weil alle Sensitivitätsmasse deutlich unter dem Schwellenwert von 70 % (Swets, 1988) liegen. Dieses Resultat geht mit Befunden anderer Studien einher, wonach Eltern keine guten Diagnostiker von Entwicklungsauffälligkeiten sind, wenn es um einzelne Funktionsbereiche geht (Malhi et al., 2005; Storck et al., 1998; Willinger & Eisenwort, 2005; Willinger et al. 2011). Auch die Nennung von Alltagsbeispielen (Glascoe & Dworkin, 1995) kann Eltern nur beschränkt helfen, allfällige Auffälligkeiten akkurat wahrzunehmen. Der einzige akzeptable RAZ-Index konnte in der Beurteilung zur Gesamtentwicklung ausgewiesen werden, wenn auch nur in bescheidenem Masse. Das Resultat wird durch den AUC-Wert der ROC-Analyse gestützt, welcher auf eine moderate Verbesserung gegenüber einer Zufallseinschätzung hinweist. Bezogen auf diese Masse könnten elterliche Sorgen bezüglich des Gesamtentwicklungsstandes ein Hinweis auf eine globale Entwicklungsverzögerung sein. Im Gegensatz zu den Befunden aus dem nicht-klinischen Bereich von Glascoe (1994), kann den Eltern infolge der mehrheitlich hohen Spezifitätsmasse (> 70 %) sowohl bei der Gesamtentwicklung als auch in einzelnen Entwicklungsbereichen zudem eine gute Wahrnehmung bezüglich nicht-auffälliger Entwicklung bescheinigt werden.

Einschränkungen und Stärken der Studie

Erstens sei darauf hingewiesen, dass es sich in der vorliegenden Studie um einen querschnittlichen Vergleich zwischen der elterlichen Einschätzung und der tatsächlichen

Leistung des Kindes handelt. Damit können Veränderungen in der elterlichen Wahrnehmung, im Sinne einer zunehmenden Passung bei zunehmender Kompetenz des Kindes nicht erfasst werden (Hunt & Paraskevopoulos, 1980). Zweitens ist zu beachten, dass der elterlich eingeschätzte bereichsspezifische Entwicklungsstand mit einem einzigen Item erfasst wurde. Eine allenfalls mangelnde Reliabilität in dieser Einschätzung würde sich drastisch auf die Korrelationen mit dem standardisierten Test auswirken. Da jedoch in 10 von 14 Bereichen die Korrelationen signifikant ausfielen, ist davon auszugehen, dass die Items zur Elterneinschätzung zumindest eine befriedigende Reliabilität aufweisen. Es bleibt unklar, ob die Korrelationen mit einer allenfalls reliableren Operationalisierung höher ausfallen würden. Drittens ist zu erwähnen, dass zur Erfassung des Entwicklungsstandes des Kindes die Testleistung in den IDS-P diente. Die IDS-P entsprechen gängigen Kriterien hinsichtlich Reliabilität, Validität und Objektivität. Trotzdem ist jede Leistungsmessung auch dann noch mit einem Mass an Unpräzision behaftet: die Tagesform des Kindes, die Vertrautheit einer Testsituation, Persönlichkeitsfaktoren des Kindes etc. mögen das Testresultat mit bedingen.

Als Stärke der Studie ist hervorzuheben, dass es sich um eine grosse repräsentative Normstichprobe von Kindern im Altersbereich von 3 bis 5 Jahren handelt. Mit den sorgfältig ausgewählten Beobachtungsbeispielen wurden einerseits die Eltern an ihre eigenen Alltagsbeobachtungen herangeführt, andererseits wurde darauf geachtet, dass die Beispiele möglichst diejenigen Kompetenzen repräsentieren, welche in den entsprechenden Untertests vorrangig eine Rolle spielen. Zur Testung des Kindes wurde eine umfangreiche, gut validierte Testbatterie eingesetzt, welche sowohl Daten zum kognitiven als auch zum allgemeinen Entwicklungsstand lieferte und damit eine gut abgestützte Aussage zur Gesamtentwicklung

erlaubte.

Fazit für die Praxis

Der positive Zusammenhang zwischen der Einschätzung der Eltern und der Testleistung des Kindes in der Mehrheit der Bereiche weist darauf hin, dass elterliche Beobachtungen ein Hinweis auf den Rangplatz des Kindes innerhalb seiner Altersgruppe sein können und zwar nicht nur, was die Globaleinschätzung angeht, sondern auch in einzelnen Entwicklungsbereichen. Die vorliegenden Resultate bestätigen die Relevanz der Erfassung und Beachtung von Elterneinschätzungen. So konnte beispielsweise in einer Längsschnittstudie von Gut, Reimann und Grob (2013) aufgezeigt werden, dass die Einschätzung der Eltern langfristig die Schulnoten vorhersagen kann. Auch zeigt der Einsatz von validen Screeningverfahren wie beispielsweise im sprachlichen Bereich ELFRA (Grimm & Doil, 2006) oder DaZ-E (Keller & Grob, 2013), welche auf Elternbeobachtungen und -einschätzungen basieren, dass diese genutzt werden können für die Erkennung von Sprachauffälligkeiten. Will man mit einer frühen Förderung langfristige Auswirkungen eines defizitären Ausgangszustandes vermindern, braucht es eine standardisierte Diagnostik, bei welcher valide und reliable Screenings mit Hilfe von Elterneinschätzungen durchaus eine wichtige Rolle spielen. Hingegen sollte nicht davon ausgegangen werden, dass Eltern generell in der Lage sind zu erkennen, ob ihr Kind in einem spezifischen Bereich eine Verzögerung aufweist, welche weiterführende professionelle Abklärungen erfordern würde.

Literatur

American Psychological Association. *Standards for Educational and Psychological Tests.*

Washington, DC: American Psychological Association; 1985.

Aylward, G.P. (1997). Conceptual issues in developmental screening and assessment. *Journal of Developmental and Behavioral Pediatrics, 18*, 340-349.

Böhm, A. & Kuhn, J. (2000). Soziale Ungleichheit und Gesundheit bei Kindern. Ergebnisse von Einschulungsuntersuchungen im Land Brandenburg. *Soziale Arbeit 49*, 343-346.

Chen, Ch., Lin, Ch.-H., Wen, S.-H. & Wu, Ch.-H. (2007). How effectively do parents discern their children's cognitive deficits at a preschool age? *Journal of the Chinese Medical Association 70 (10)*, 445-450.

Chung, C.-Y., Liu, W.-Y., Chang, Ch.-J., Chen, C.-L., Tang, S. F.-T. & Wong, M.-K. (2010). The relationship between parental concerns and final diagnosis in children with developmental delay. *Journal of Child Neurology, 26*, 413-419.

Deimann, P. & Kastner-Koller, U. (2011). Maternal evaluations of young children's developmental status: A comparison of clinic- and non-clinic-groups. *Psychological Test and Assessment Modeling, 53 (2)*, 214-227.

Deimann, P., Kastner-Koller, U., Benka, M., Kainz, S. & Schmidt, H. (2005). Mütter als Entwicklungsdiagnostikerinnen. Der Entwicklungsstand von Kindergartenkindern im Urteil ihrer Mütter. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 37*, 122-134.

- Evers, A. (2001) Improving test quality in the Netherlands: Results of 18 years of test ratings. *International Journal of Testing, 1*, 137-153.
- Glascoe, F. P. (1994). It's not what it seems. The relationship between parents' concerns and children with global delays. *Clinical Pediatrics, 33*, 292-296.
- Glascoe, F. P. (1999). Using parents' concerns to detect and address developmental and behavioural problems. *Journal of the Society of Pediatric Nurses, 4 (1)*, 24-35.
- Glascoe, F. P. & Dworkin, P. H. (1995). The role of parents in the detection of developmental and behavioural problems. *Pediatrics, 95*, 829-836.
- Glascoe, F. P. & Sandler, H. (1995). Value of parents' estimates of children's developmental ages. *The Journal of Pediatrics, 127 (5)*, 831-835.
- Grimm, H. & Doil, H. (2006). *ELFRA Elternfragebögen für die Früherkennung von Risikokindern* (2. Aufl.). Göttingen: Hogrefe.
- Grimm, H. (2001), *Sprachentwicklungstest für drei bis fünfjährige Kinder (SETK- 3-5)*. Göttingen: Hogrefe.
- Grob, A., Reimann, A., Gut, J. & Frischknecht M.-C. (2013). *Intelligence and Development Scales – Preschool (IDS-P)*. Bern: Hans Huber.
- Gut, J., Reimann, G. & Grob, A. (2013). A contextualized view on long-term predictors of academic performance. *Journal of Educational Psychology, 105 (2)*, 436-443.
- Hagmann-von Arx, P. & Grob, A. (2014). *Reynolds Intellectual Assessment Scales (RIAS). Deutschsprachige Adaptation der Reynolds Intellectual Assessment Scales (RIAS™) von Cecil R. Reynolds und Randy W. Kamphaus*. Bern: Hans Huber.

- Hunt, J. McV. & Paraskevopoulos, J. (1980). Children's psychological development as a function of inaccuracy of their mothers' knowledge of their abilities. *The Journal of Genetic Psychology*, 136, 285-298.
- ICD-10 (2007). International statistical classification of diseases and related health problems (10th revision, version for 2007).
- Keller, K. & Grob, A. (2013). Elternfragebogen zu den Deutschkenntnissen mehrsprachiger Kinder. *Zeitschrift für Pädagogische Psychologie*, 27 (3), 169-180.
- Koch, H., Kastner-Koller, U., Deimann, P., Kossmeier, Ch., Koitz, C. & Steiner, M. (2011). The development of kindergarten children as evaluated by their kindergarten teachers and mothers. *Psychological Test and Assessment Modeling*, 53 (2), 214-257.
- Köbberling, J., Richter, K., Trampisch, HJ., & Windeler, J. (1991). *Methodologie der medizinischen Diagnostik. Entwicklung, Beurteilung und Anwendung von Diagnoseverfahren in der Medizin*. Berlin Heidelberg New-York: Springer-Verlag.
- Malhi, P., Kashyap, S. & Dua, S. (2005). Maternal estimates of mental age in developmental assesment. *Indian Journal of Pediatrics*, 72, 931-934.
- Marx, P. & Lenhard, W. (2010). Diagnostische Merkmale von Screeningverfahren. In M. Hasselhorn & W. Schneider (Hrsg.), *Frühprognose schulischer Kompetenzen*. Göttingen : Hogrefe.
- Melchers, P. & Preuss, U. (2009), *Kaumann Assessment Battery for Children (K-ABC). Deutsche Version* (8., unveränderte Aufl.). Frankfurt a.M.: Pearson Assessment.

- Pulsifer, M. B., Hoon, A. H., Palmer F. B., Gopalan, R. & Capute A. J. (1994). Maternal estimates of developmental age in preschool children. *The Journal of Pediatrics*, 125 (1), 18-24.
- Rutter, M. (1994). Continuities, transitions and turning points in development. In M. Rutter & D. Hay (Eds.), *Development through life* (pp. 1-25). London: Blackwell.
- Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W.S., Belfield, C.R. & Nores, M., (2005). *Lifetime effects: The HighScope Perry Preschool Study Trough Age 40*. Ypsilanti, MI: HighScope Press.
- Storck, M., Webel, D., Steinmacher, J., Kupferschmidt, C., Bode, H. (1998). Diagnostik umschriebener Entwicklungsstörungen im Vorschulalter. Zur Validität elterlicher Angaben. *Monatsschrift Kinderheilkunde*, 146, 988-992.
- Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, 240, 1285-1293.
- Tröster, H., Flender, J. & Reineke, D., (2005). Entwicklungsscreening. Dortmunder Entwicklungsscreening für den Kindergarten (DESK 3-6). *Kindheit und Entwicklung* 14 (3), 140-149.
- Wiedebusch, S. & Petermann, F. (2011). Förderung sozial-emotionaler Kompetenz in der frühen Kindheit. *Kindheit und Entwicklung*, 20 (4), 209-218.
- Willinger, U. & Eisenwort, B. (2005). Mothers' estimates of their children with disorders of language development. *Behavioral Medicine*, 31(3), 117-126.
- Willinger, U., Schaunig, I., Jantscher, S., Schmoeger, M., Loader, B., Kummer, Ch. & et al. (2011), Mothers' estimates of their preschool children and parenting stress. *Psychological Test and Assessment Modeling*, 53 (2), 228-240.

Autorenhinweis

Lic.phil. Marie-Claire Frischknecht

Institut für Psychologie der Universität Basel

Missionsstraße 62 a

CH-4055 Basel

marie-claire.frischknecht@unibas.ch

Dr. phil. Giselle Reimann

Institut für Psychologie der Universität Basel

Missionsstraße 62 a

CH-4055 Basel

giselle.reimann@unibas.ch

Prof. Dr. Alexander Grob

Fakultät für Psychologie

Missionsstrasse 60/62

CH-4055 Basel

alexander.grob@unibas.ch

Anhang. Testbeschreibung IDS-P

Untertest	Testbeschreibung	Anzahl Items	α
Kognition			.91
Wahrnehmung Visuell	Serien mit Abbildungen unterschiedlich langer Buntstifte bilden	8	.84
Aufmerksamkeit Selektiv	Entenbilder mit einem bestimmten Merkmal aus einer Menge verschiedener Entenbildern möglichst schnell richtig aussortieren	3	.93
Gedächtnis Phonologisch	Kunstwörter mit ansteigender Silbenlänge unmittelbar wiedergeben.	15	.70
Gedächtnis Räumlich- Visuell	Geometrische Figuren unter mehreren wiedererkennen.	10	.82
Gedächtnis Auditiv	Eine vorgelesene Geschichte nach 20 Minuten wiedererzählen können.	10	.81
Denken Bildlich	Geometrische Figuren mit Hilfe von Klötzchen nachlegen.	12	.85
Denken Konzeptuell	Gemeinsames Konzept von drei Bildern erkennen und mit zwei passenden Bildern ergänzen.	11	.81
Allgemeine Entwicklung			
Psychomotorik		45	.92
Grobmotorik	Auf einem auf dem Boden liegenden flachen Seil balancieren, Seil mit den Füßen seitlich	4	.77

	überspringen, auf einem Bein stehen		
Feinmotorik	Perlen und Würfel nach Bildvorlage	5	.65
	schnellstmöglichst auffädeln		
Visuomotorik	Geometrische Figuren abzeichnen	36	.90
Sozial-Emotionale	Gefühle von Kindern anhand ihres	16	.77
Kompetenz	Gesichtsdrucks auf einem Foto erkennen; Soziale Situationen auf Bildern verstehen und erklären		
Denken Logisch-	Mathematische Vorläuferfertigkeiten	17	.88
Mathematisch	(Mengenbegriff 3-10, Rangplatzerkennung, Kardinalität und Invarianz)		
Sprache		31	.84
Sprache Expressiv	Bildinhalt korrekt wiedergeben	8	.88
Sprache Rezeptiv	Vorgesprochene Sätze inhaltlich korrekt mit Holzfiguren nachspielen	12	.79

Tabelle 1. Sensitivität, Spezifität, PPV und RATZ-Index für die Elterneinschätzungen

Funktionsbereich	Sensitivität in %	Spezifität in %	PPV in %	RATZ- Index
Wahrnehmung Visuell	37	71.6	13.5	0.109
Aufmerksamkeit Selektiv	21.7	87.9	19.1	0.097
Gedächtnis Phonologisch	38.8	80	18.4	0.215
Gedächtnis Räumlich-Visuell	15.2	83	11.6	-0.019
Gedächtnis Auditiv	13.3	87.4	12.9	0.007
Denken Bildlich	17.1	95	33.3	0.239
Denken Konzeptuell	23.9	78	14.4	0.020
Grobmotorik	45	81.4	25	0.297
Feinmotorik	44.4	76.5	21.1	0.248
Visuomotorik	39.1	82	24	0.232
Sozial-Emotionale Kompetenz	21.1	76.5	10.4	-1.807
Denken Logisch-Mathematisch	29	79.3	16.5	0.093
Sprache Expressiv	27.9	82.2	17.9	0.109
Sprache Rezeptiv	14.8	83.4	11.1	-0.02
Gesamtentwicklung	55	98	38	0.529

Anmerkungen. PPV: Positiv prädiktiver Vorhersagewert, RATZ-Index: Relativer Anstieg der Trefferquote gegenüber der Zufallstrefferquote; N variiert zwischen 461 und 504, da für einige Fragen die Antwort 0 = „das kann ich nicht beantworten“, gewählt wurden.

Studie 2

Frischknecht, M.-C., Reimann, G., Gut, J., Ledermann, T. & Grob, A., (2014). Wie genau können Mütter die Mathematik- und Sprachleistungen ihrer Kinder einschätzen? Ein Vergleich zwischen Mütterereinschätzung und Testleistungen bei Kindern im Alter von 6-10 Jahren. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 46 (2), 67-78.

Wie genau können Mütter die Mathematik- und Sprachleistungen ihrer Kinder einschätzen?

Ein Vergleich zwischen Mütterereinschätzung und Testleistungen bei Kindern im Alter von 6–10 Jahren

Marie-Claire Frischknecht, Giselle Reimann, Janine Gut,
Thomas Ledermann und Alexander Grob

Institut für Psychologie der Universität Basel

Zusammenfassung. Es wurde geprüft, wie genau Mütter die Mathematik- und Sprachleistungen ihrer 6- bis 10-jährigen Kinder ($N = 382$) einzuschätzen vermögen. Die mütterliche Einschätzung wurde verglichen mit der Testleistung in den Intelligence and Development Scales (IDS). Die Einschätzungen der Mütter korrelierten substantiell positiv mit den Testleistungen der Kinder. Mütter überschätzten die Leistungen der Kinder. Die Mütter schätzten die Mathematikleistungen von Jungen höher ein als jene von Mädchen, wogegen sie die Deutschleistungen von Mädchen höher einschätzten als jene von Jungen. Gemessen an den Testleistungen wurden Mädchen und Jungen von den Müttern gleich häufig über-, unter- und genau eingeschätzt. Die vorliegenden Resultate bestätigen die Überschätzung seitens der Eltern, welche in der Literatur zu allgemeinen Entwicklungsbereichen häufig berichtet und wiederholt repliziert wurde. Mit dieser Untersuchung kann dieses Wissen auf die schulischen Leistungsfächer erweitert werden.

Schlüsselwörter: Diagnostische Kompetenz, Akkuratheit von Elternurteilen, Elterneinschätzung, Entwicklungsdiagnostik, Leistungseinschätzung Mathematik und Sprache

The accuracy of maternal reports regarding their children's mathematical and language abilities. A comparison of maternal estimates and actual test scores of children between 6 and 10 years of age

Abstract. In this study maternal estimations of child achievement in mathematics and language were compared to the children's achievement in a standardized mathematics and language test using the Intelligence and Development Scales (IDS 5–10, $N = 382$ children 6 to 10 years of age). A significant positive correlation was observed between mothers' estimations and actual test achievement. Mothers generally overestimated their children's achievement. Boys were estimated to score higher in mathematics compared to girls, and girls were estimated to score higher in language compared to boys. Mothers overestimated, underestimated, or correctly estimated the achievement of girls and boys equally often. The present results confirm maternal overestimation of child abilities, which has often been found and replicated with respect to general development. The present study extended this finding to the field of scholastic achievement. Key words: diagnostic competence, accuracy of parent judgment, parental assessment, parental beliefs, developmental assessment, conceptions of abilities, gender differences

Eltern sind wichtige Informationsquellen bei der Anamnese in schul- und lernpsychologischen Abklärungen, da sie das Verhalten des Kindes außerhalb der Schule beobachten können. Sie sind meist die primären Beobachter einer Entwicklungsauffälligkeit und melden ihr Kind aufgrund dieser Beobachtungen für eine medizinische, sprachtherapeutische oder entwicklungspsychologische Abklärung oder Behandlung an. Der Einbezug von Eltern bei der Planung von schulischen Maßnahmen fördert deren Akzeptanz und führt bei den Eltern zu einer

höheren Identifikation mit den getroffenen Maßnahmen (Schrader, 2006). Gleichzeitig stellen Eltern einen bedeutsamen Teil der Lernumwelt des Kindes dar, welche Anforderungen und Erwartungen an seine Kompetenzen stellt. Kenntnisse über die Kompetenzen des Kindes können Eltern nicht nur über tägliche Beobachtungen zuhause und das Aufgabenverhalten der Kinder, sondern auch über schriftliche Rückmeldungen von oder Zeugnisgespräche mit den Lehrpersonen sowie über den berichteten oder wahrgenommenen Vergleich mit den Leis-

tungen von Kindern derselben Klasse gewinnen. Bei Problemstellungen werden die Einschätzungen der Eltern zu den Kompetenzen des Kindes zunehmend in die entwicklungsdiagnostische Abklärung miteinbezogen und bilden so einen Bestandteil des diagnostischen Prozesses. Damit treffen Diagnoseleistungen aus verschiedenen Quellen zusammen. Schrader (2006) unterscheidet in diesem Kontext formelle von informellen Diagnoseleistungen. Formelle Diagnosen beruhen auf wissenschaftlich erprobten Methoden mit vorgegebenen Qualitätsstandards. Eltern hingegen geben informelle Diagnoseleistungen ab. Diese entstehen meist beiläufig und unsystematisch im Rahmen des alltäglichen erzieherischen Handelns. Sie setzen sich aus impliziten subjektiven Urteilen, Einschätzungen und Erwartungen zusammen. Wenn nun die Elterneinschätzung als informelle Diagnoseleistung Bestandteil des diagnostischen Prozesses sein soll, ist vorab zu prüfen, wie die Entsprechung zwischen kindlicher Testleistung und Elternurteil in relevanten Entwicklungsbereichen ausfällt.

Vergleich von Elterneinschätzung und Testleistung

Vergleiche zwischen Elterneinschätzungen und objektiven Testergebnissen zeigen, dass Eltern in der Regel das Leistungsniveau ihres Kindes überschätzen (Buch, Sparfeldt & Rost, 2006; Deimann, Kastner-Koller, Benka, Kainz & Schmidt, 2005; Petermann, Stein & Macha, 2006; Rennen-Allhoff, 1991). Die schulischen Leistungen betreffend konnte nachgewiesen werden, dass Eltern eher positive Einschätzungskategorien für die Leistungen ihrer Kinder bevorzugen. Sie wählen Kategorien, welche für eine – bezogen auf die Klasse – unterdurchschnittliche Leistung der Kinder steht, deutlich seltener als Kategorien, die für eine überdurchschnittliche Leistung stehen (Räty, Vänskä, Kasanen & Kärkkäinen, 2002). Des weiteren scheint der Bildungsstand der Mutter mit der Leistungseinschätzung und den Leistungen der Kinder zusammenzuhängen. So bestätigen Studien wiederholt, dass Mütter mit höherer Schulbildung höhere Erwartungen hinsichtlich der Schulleistungen ihrer Kinder hatten und diese höheren Erwartungen positiv mit den darauffolgenden Leistungen in Mathematik und Lesen korrelierten (z. B. Davis-Kean, 2006; Lee, 2010).

Bei einem Vergleich zwischen den elterlichen Erwartungen und schulischen Leistungen der Kinder in Mathematik und Sprache zeigten sich substantielle Korrelationen zwischen $r = .45$ bis $r = .62$ im sprachlichen Bereich und zwischen $r = .40$ bis $r = .51$ für Mathematik (Phillipson & Phillipson, 2007). In einer Untersuchung von Waschbusch, Daleiden und Drabman (2000) beantworteten Eltern nebst Fragen zum Verhalten auch Fragen zu verbalen Kompetenzen (Sprach- und Textverständnis, Gedächtnis für verbale Instruktionen, Verbales Verhalten

gegenüber anderen Kindern). Bezüglich der verbalen Kompetenzen konnten signifikante Korrelationen zwischen Elternurteil und Testleistungen der Kinder nur für die Jungen nachgewiesen werden; für Mädchen waren diese Korrelationen nicht bedeutsam. Bei Jungen wird von Eltern ein erfolgreiches Testergebnis vielfach mit Begabung in Verbindung gebracht, während bei Mädchen Erfolg sowohl bei Mathematik- als auch bei Sprachleistungen eher auf Anstrengung zurückgeführt wird (Lumms & Stevenson, 1990; Räty et al., 2002). Diese Ergebnisse geben Hinweise darauf, dass die Elterneinschätzungen bei Mädchen und Jungen in den Bereichen Mathematik und Sprache unterschiedlich ausfallen können.

Geschlechtsunterschiede in Elterneinschätzungen

Verschiedene Studien fanden Geschlechtsunterschiede hinsichtlich der elterlichen Einschätzung in Bezug auf Mathematik. Furnham (2000) ließ 112 Eltern ihre eigene und die Intelligenz ihrer Kinder bezogen auf Gardners (1983) sieben Funktionsbereiche der Intelligenz einschätzen. In Bezug auf die mathematischen Kompetenzen wurden die Jungen von beiden Elternteilen höher eingeschätzt als die Mädchen. Tiedemann (2000) untersuchte bei 489 Kindern aus der 3. und 4. Grundschule die Eigeneinschätzung der Kinder zu deren Mathematikleistungen, die Einschätzung ihrer Eltern bezüglich aktueller und zukünftiger Leistung sowie die Einschätzung der Lehrer bezüglich dieses Leistungsbereiches. Die Meinung, dass Jungen in Mathematik besser seien als Mädchen, wurde in allen befragten Gruppen signifikant vertreten, obwohl in vorangehenden Tests keine Leistungsdifferenz zwischen Mädchen und Jungen festzustellen waren.

Bei der Einschätzung von Sprachleistungen schreiben Eltern Mädchen gleiche oder höhere Kompetenzen zu als Jungen. Lumms und Stevenson (1990) verglichen Müttereinschätzungen bei Stichproben aus Japan, China und den USA. In allen drei Kulturen schätzten die Mütter die Lesekompetenzen bei den Mädchen höher ein als bei den Jungen, obwohl gezeigt werden konnte, dass Jungen ebenso gut lesen konnten wie Mädchen. In der Untersuchung von Räty et al. (2002) attestierten Eltern Mädchen im Lesen signifikant höhere Kompetenzen als Jungen.

Die Frage, ob tatsächlich Geschlechtsunterschiede in den mathematischen Kompetenzen vorliegen, ist nicht einfach zu beantworten, weil die Ergebnisse von den in einem Test geprüften Teilbereichen der Mathematik abhängen und in Abhängigkeit des Lebensalters verschieden ausfallen (Lumms & Stevenson, 1990; Lynn & Irwing, 2008). Für das Alter von 7–10 Jahren werden in einigen Studien signifikant bessere Leistungen bei Jungen in Mathematik nachgewiesen (Kovas, Haworth, Petrill &

Plomin, 2007; Swiatek, Lupkowski-Shoplik & O'Donoghue, 2000). Für die Sprachkompetenz konnten Hyde und Linn (1988) metaanalytisch generell keine substantiellen geschlechtsspezifischen Unterschiede nachweisen, aber einzelne Studien im Altersbereich von 7–10 Jahren wiesen nach, dass Mädchen im Lesen oder Schreiben signifikant bessere Ergebnisse erzielen als Jungen (Lummis & Stevenson, 1990; Schneider, Stefanek & Dotzler, 1997; Schwenck & Schneider, 2003; Swiatek et al. 2000).

Zusammenfassung des bisherigen Forschungsstandes und Fragestellung der vorliegenden Studie

Zusammenfassend gilt, dass die Elterneinschätzung von schulischen Leistungen mit objektiven Testergebnissen substantiell positiv korreliert (Phillipson & Phillipson, 2007), wobei sie bezüglich Mathematik und Sprachkompetenz je nach Geschlecht verschieden ausfallen kann und zum Teil stärkere Zusammenhänge für Jungen gefunden wurden (Furnham, 2000; Lummis & Stevenson, 1990; Rätty et al., 2002; Waschbusch et al. 2000). Darüber hinaus scheint der Bildungsstand der Mutter einen Einfluss auf die Leistungseinschätzung und die Leistung selbst zu haben (Davis-Kean, 2006; Halle, Kurtz-Costes & Mahoney, 1997; Lee, 2010).

Die Studien zu Elterneinschätzung und kindlichem Leistungsvermögen unterscheiden sich vor allem im Hinblick darauf, wie Elterneinschätzungen operationalisiert wurden. Dies hängt auch damit zusammen, dass keine standardisierten Instrumente zur Erfassung der Elterneinschätzung vorliegen. Die Erfassung der kindlichen Leistung ist ebenfalls uneinheitlich. Insbesondere bei den verbalen Kompetenzen kann dies beispielsweise nur das Lesen (Lummis & Stevenson, 1990; Rätty et al., 2002, Swiatek et al., 2000), nur die Rechtschreibung (Schneider et al., 1997) oder aber eine umfassende Erfassung von diversen verbalen Kompetenzen (Buch et al., 2006; Niklas & Schneider, 2012; Phillipson & Phillipson, 2007; Schwenck und Schneider, 2003; Waschbusch et al., 2000) beinhalten. Darüber hinaus unterscheiden sich die Studien dahingehend, dass die Elterneinschätzung nicht immer mit einer aktuellen Testleistung des Kindes verglichen wurde. So untersuchten beispielsweise Rätty et al. (2002) und Furnham (2000) die Einschätzung und allfällige Attributionsmuster der Eltern, ohne diese direkt mit Testleistungen zu vergleichen. Des Weiteren wurden nicht immer standardisierte Testverfahren zur Leistungsüberprüfung angewendet (Kovas et al., 2007; Lummis & Stevenson, 1990; Phillipson & Phillipson, 2007; Tiedeman, 2000). Sowohl für eine Leistungsüberprüfung von mathematischen und sprachlichen Kompetenzen, als auch um einen Vergleich der Leistungen von Mädchen und Jungen in diesen Bereichen machen zu können werden

Resultate von objektiven standardisierten Testverfahren benötigt. Die Forschung zur diagnostischen Kompetenz hinsichtlich der schulischen Leistungsfächer Mathematik und Deutsch untersuchte vorrangig die Diagnosegenauigkeit von Lehrkräften in Bezug auf die Leistungen von Schülerinnen und Schülern in der Grundschule (Karing, 2009). Es gibt kaum Studien, welche gezielt die Leistungsbereiche Mathematik und Deutsch bei Kindern der ersten Grundschuljahre auf die Übereinstimmung zwischen Elterneinschätzung und Testleistung des Kindes in standardisierten Tests prüfen.

Die vorliegende Studie untersucht erstens die Übereinstimmung der mütterlichen Einschätzung mit der kindlichen Leistung in Mathematik und Sprache und ob Überschätzungen häufiger vorkommen als Unterschätzungen. Zudem testen wir, ob es Unterschiede zwischen den Geschlechtern gibt. In den wenigen Studien, welche Mathematik- und Sprachleistungen der Kinder mit Elterneinschätzungen vergleichen, wird die Einschätzung der Eltern nicht direkt mit der Testleistung des Kindes in Verbindung gebracht außer in Form von Zusammenhängen (Phillipson & Phillipson, 2007; Waschbusch et al., 1999) was jedoch keine Informationen darüber gibt, ob Eltern die Leistungen der Kinder über- oder unterschätzen. Mit der ersten Fragestellung soll diese Forschungslücke geschlossen werden. Aufgrund der bisher festgestellten Tendenz, dass Eltern höhere Einschätzungskategorien bevorzugen (Rätty et al., 2000) wird angenommen, dass Mütter die Leistungen ihrer Kinder in den Schulfächern Mathematik und Deutsch häufiger überschätzen als genau ein- bzw. unterschätzen.

Zweitens wird der korrelative Zusammenhang zwischen mütterlich attribuerter Leistung der Schulfächer Mathematik und Deutsch und tatsächlicher Testleistung in Mathematik und Sprache erfasst. Explorativ überprüfen wir, ob diese Korrelationen geschlechtsspezifisch unterschiedlich ausfallen. Studien zur Übereinstimmung von Elterneinschätzung und kindlicher Leistungsfähigkeit erstrecken sich meist über verschiedene Entwicklungsbereiche oder Intelligenzeinschätzungen (Buch et al., 2006; Deimann et al., 2005; Petermann et al., 2006; Rennen-Allhoff, 1991). Zudem korrelieren Elterneinschätzungen mit objektiven Testresultaten substantiell positiv (Phillipson & Phillipson, 2007), wenn auch die Korrelationskoeffizienten für Mädchen und Jungen unterschiedlich ausfallen (Waschbusch et al., 2000). In der vorliegenden Studie wird deshalb gezielt nach den Einschätzungen im Leistungsbereich „Mathematik“ und „Deutsch“ gefragt. Auf Basis vorliegender Evidenzen erwarten wir, dass die mütterliche Leistungseinschätzung sowohl in Mathematik wie in Deutsch positiv mit der tatsächlichen Testleistung korreliert. Des weiteren testen wir, ob der Zusammenhang zwischen der tatsächlichen Testleistung und der mütterlichen Leistungseinschätzung durch das Geschlecht des Kindes moderiert wird, wobei der Bildungsstand der Mutter als

Kovariate einbezogen wird, da dieser signifikant mit den Leistungseinschätzungen und den Leistungen in Mathematik und Sprache zusammenhängen kann (Davis-Kean, 2006; Halle et al., 1997; Lee, 2010).

Drittens wird untersucht, ob die mütterliche Einschätzung in den Schulfächern Mathematik und Deutsch je nach Geschlecht des Kindes unterschiedlich ausfällt. Die Leistungen von Mädchen und Jungen werden von Eltern unterschiedlich eingeschätzt (Furnham, 2000; Lummis & Stevenson, 1990; Rätty et al., 2002; Tiedemann, 2000). Allerdings wurde diese Fragestellung selten spezifisch in Bezug auf Schulfächer untersucht, sondern meistens in Zusammenhang mit allgemeinen Entwicklungsbereichen oder Intelligenz. Auf Basis vorliegender Evidenz erwarten wir, dass Mütter die Leistungen von Jungen in Mathematik höher, jene in Deutsch jedoch niedriger einschätzen als die Leistungen von Mädchen.

Methoden

Stichprobenbeschreibung und Durchführung

Die Stichprobe besteht aus einer Teilmenge der Normierungsstichprobe IDS (Intelligence and Development Scales; Grob, Meyer & Hagmann-von Arx, 2009). Die IDS-Normierung wurde mit 1330 Kindern im Alter von 5 bis 10 Jahren ($M = 7.51$ Jahre, $SD = 1.63$ Jahre) in der Schweiz, Deutschland und Österreich im Einzelsetting durchgeführt. Parallel dazu bearbeitete die große Mehrheit der Eltern einen Fragebogen zum Verhalten und Entwicklungsstand ihres Kindes. In die vorliegende Studie wurden die Kinder einbezogen, welche die 1. bis 4. Klasse besuchten und für die der von der Mutter beantwortete Fragebogen vorlag. Die Einschränkung auf die mütterliche Einschätzung wurde zugunsten der Stichprobenhomogenität vorgenommen. Nach Ausschluss der Kinder mit fehlenden Angaben zur Leistung in den IDS-Untertests Mathematik und/oder Sprache sowie fehlenden Angaben im Elternfragebogen zur Einschätzung dieser Leistungsbereiche resultierte eine Stichprobe von $N = 382$ (Anteil Mädchen: $n = 190$, Anteil Jungen: $n = 192$; 93 % Deutsche Muttersprache; Erhebungskontexte: 51 % Schweiz; 28 % Deutschland; 21 % Österreich) im Alter zwischen 6 und 10 Jahren ($M = 7.98$ Jahre, $SD = 1.24$ Jahre). 82 Kinder besuchten die 1. Klasse, 101 Kinder die 2. Klasse, 89 Kinder die 3. Klasse und 110 Kinder die 4. Klasse. Sechs Prozent der Mütter gaben an, über keinen Berufsabschluss zu verfügen, 39 % gaben eine Berufsausbildung, 5 % eine Berufsausbildung mit Zusatzausbildung, 20 % eine höhere Berufsausbildung und 27 % einen Hochschulabschluss an. Bei 3 % fehlten die Angaben zur Berufsausbildung.

Untersuchungsinstrumente

Die Testleistung für Mathematik und Sprache wurde mit den entsprechenden IDS Untertests (Grob et al., 2009) erfasst. Mit den Aufgaben des Untertests Denken Logisch-Mathematisch werden unterschiedliche Entwicklungs- und Kompetenzbereiche mit aufsteigender Schwierigkeit operationalisiert: Mengenerfassung bis 5, Ordinalität, Invarianzaufgaben, Mentale Addition, Kenntnisse der Struktur des Zehnersystems, Proportionen, abstraktes räumliches Vorstellungsvermögen und Gleichungsaufgaben. Die interne Konsistenz dieses Untertests lag bei .85 (Cronbach's Alpha). Der Untertest Denken Logisch-Mathematisch korrelierte positiv mit den schulischen Leistungen, welche mit dem Schweizer Rechentest 1.–3. Klasse (SR 1–3; Lobeck & Frei, 1987) erhoben wurden. Die Aufgaben lassen sich inhaltlich den Richtzielen des Lehrplanes Mathematik zuordnen (Grob et al., 2009).

Der Bereich Sprache wurde anhand der beiden Untertests Sprache Rezeptiv und Sprache Expressiv erfasst. Im Untertest Sprache Rezeptiv muss ein vorgeschriebener Satz verstanden und in der richtigen Reihenfolge mit Spielfiguren nachgespielt werden (Bsp.: „Bevor der Junge auf dem rechten Bein hüpfte, bellte der Hund dreimal, nachdem er auf das Hundehaus gesprungen ist.“). Die Bearbeitung dieses Untertests erfordert das Verständnis von Wörtern sowie die syntaktische Analyse grammatikalischer Beziehungen. Im Untertest Sprache Expressiv sollen aus Wörtern (Nomen), die als Bild (beispielsweise ein Eis) dargeboten werden, semantisch sinnvolle und grammatikalisch korrekte Sätze gebildet werden (Bsp.: „Im Sommer esse ich gerne ein Eis.“). Die Bearbeitung dieses Untertests verlangt vom Kind Wortschatz, die angemessene Anwendung der Sprachproduktionsregeln, sprachliche Ausdrucksfähigkeit und Gewandtheit sowie die Herstellung von logischen Zusammenhängen. Sprache Expressiv und Rezeptiv korrelierten in der vorliegenden Stichprobe mit $r = .38$ ($p < .01$) miteinander. Für jedes Kind wurde der Mittelwert über die Summenwerte dieser beiden Untertests berechnet, um so eine neue Variable „Sprache“ zu bilden. Die interne Konsistenz der „Sprache“ lag bei .82 (Cronbach's Alpha). Die sprachlichen Untertests zeigten bedeutsame Zusammenhänge mit den Lese- und Rechtschreibleistungen, die bei der obengenannten Validierungsstichprobe mittels des Salzburger Lese- und Rechtschreibtestes (SLRT; Landerl, Wimmer & Moser, 2006) erhoben wurden.

Die durch die Mütter erfolgte Leistungseinschätzung wurde mit je einem Item für Mathematik und Deutsch erhoben („Wie schätzen Sie die Leistungen Ihres Kindes in Mathematik (Deutsch) ein im Vergleich zu einem Kind mit durchschnittlicher Leistung im selben Alter?“). Für die Beantwortung stand je eine fünfstufige Ratingskala zur Verfügung, wobei 1 = „mein Kind ist viel schlechter“, 2 = „mein Kind ist schlechter“, 3 = „mein Kind ist gleich gut“, 4 = „mein Kind ist besser“, 5 = „mein Kind ist viel bes-

ser“ (im Vergleich zu einem Kind mit durchschnittlicher Leistung im selben Alter) bedeutete.

Variablenbildung und Analysemethoden

Um die mütterliche Einschätzung mit der Testleistung des Kindes zu vergleichen, wurden je zwei neue dreistufige Variablen gebildet. Entsprechend den Angaben des Untersuchungsinstruments IDS 5–10 (Grob et al., 2009), wonach Wertpunkte von 7–13 als durchschnittliche Leistung (altersstandardisierter Mittelwert = 10 Wertpunkte ($SD = \pm 3$ Wertpunkte) und Wertpunkte unter 7 als unterdurchschnittliche, solche über 13 als überdurchschnittliche Leistung definiert sind, bestanden die beiden Variablen „Testleistung Mathematik“ und „Testleistung Sprache“ neu aus drei Kategorien 1 = unterdurchschnittlich (Wertpunkte 1–6), 2 = durchschnittlich (Wertpunkte 7–13) und 3 = überdurchschnittlich (Wertpunkte 14–19). Bei den Elterneinschätzungen wurde versucht, eine inhaltliche Korrespondenz zu diesen drei Kategorien zu finden, indem die Variablen „Einschätzung der Mütter in Mathematik“ und „Einschätzung der Mütter in Sprache“ neu entsprechend in die Stufen 1 = unterdurchschnittlich („mein Kind ist viel schlechter“ und „mein Kind ist schlechter“), 2 = durchschnittlich („mein Kind ist gleich gut“) und 3 = überdurchschnittlich („mein Kind ist besser“ und „mein Kind ist viel besser“) eingeteilt wurden.

Mittels Chi-Quadrat-Tests wurde geprüft, ob Überschätzungen häufiger vorkommen als Unterschätzungen. Für diese Analyse wurde eine neue Variable „Übereinstimmung“ gebildet, welche dreistufig ausfiel, nämlich 1 = Unterschätzung (Elterneinschätzung liegt tiefer als Testleistung), 2 = Übereinstimmung (Testleistung und Elterneinschätzung stimmen überein), 3 = Überschätzung (Elterneinschätzung liegt höher als Testleistung). Die Zusammenhangshypothesen wurden mit Pearson Korrelationen getestet und anhand von Fischer's Z-Transformationen auf Unterschiedlichkeit hin geprüft. Mittels einer hierarchischen Regressionsanalyse wurde getestet, ob der Zusammenhang zwischen der tatsächlichen Testleistung und der mütterlichen Leistungseinschätzung durch das Geschlecht des Kindes (0 = Junge; 1 = Mädchen) unter Kontrolle des Bildungsstandes der Mutter moderiert wird. Die Leistungseinschätzung durch die Mutter wurde zentriert, um dem Problem der Multikollinearität entgegenzuwirken (Aiken & West, 1991). Der Bildungsstand der Mutter wurde auf einer fünfstufigen Skala erfasst (1 = keinen Berufsabschluss, 2 = Berufsausbildung, 3 = Berufsbildung mit Zusatzausbildung, 4 = höhere Berufsbildung und 5 = Hochschulabschluss). Gruppenunterschiede zwischen Jungen und Mädchen wurden schließlich mittels t -Tests für unabhängige Stichproben überprüft. Sämtliche Analysen wurden mit SPSS Version 19 durchgeführt.

Ergebnisse

Übereinstimmung der mütterlichen Einschätzung mit der kindlichen Leistung. Die Anteile an genauer Einschätzung respektive Über- oder Unterschätzung im Vergleich mit der Testleistung werden in Abbildung 1 dargestellt. Der Anteil an genauer Einschätzung (Müttereneinschätzung und Testleistung korrespondieren) betrug in Mathematik bei den Jungen 50.1 % und bei den Mädchen 55.8 %. Bezüglich Sprache war der Anteil an genauer Einschätzung bei den Jungen 55.7 % und bei den Mädchen 44.8 %. Der Anteil an Überschätzung (Müttereneinschätzung lag in einer höheren Kategorie als die Testleistung) betrug in Mathematik bei den Jungen 43.2 % und bei den Mädchen 34.2 %. In Bezug auf Sprache war dieser Anteil bei den Jungen 33.3 % und bei den Mädchen 44.7 %. Der Anteil an Unterschätzung (Müttereneinschätzung lag in einer niedrigeren Kategorie als die Testleistung) betrug bei den Jungen bezüglich Mathematik 6.7 %, bei den Mädchen 10 % und bezüglich Sprache bei den Jungen 11 % und bei den Mädchen 10.5 %. (Für detaillierte Angaben zum Verhältnis der mütterlichen Einschätzung zur Testleistung s. Kreuztabellen 1–4 im Anhang.)

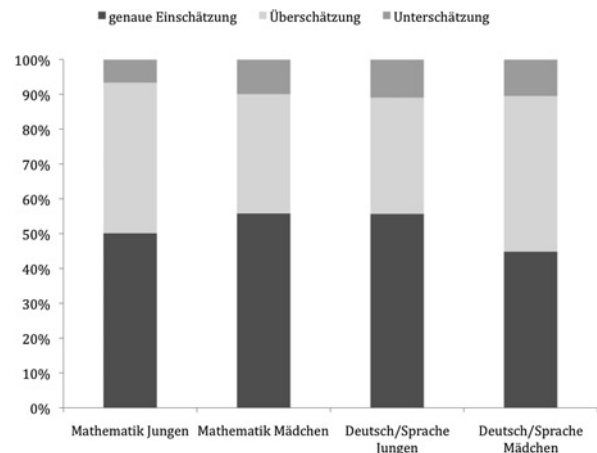


Abbildung 1. Anteil an genauer Einschätzung, Überschätzung und Unterschätzung den Müttern im Vergleich zu den Testleistungen der Kinder

Überschätzung versus Unterschätzung. In Bezug auf Mathematik zeigte sich, dass Überschätzungen signifikant häufiger vorkamen als Unterschätzungen, und zwar sowohl bei Jungen ($\chi^2_{(0.001;1;N = 192)} = 51.04$), als auch bei Mädchen ($\chi^2_{(0.001;1;N = 190)} = 25.19$). Auch in Bezug auf Sprache kamen Überschätzungen bei Jungen ($\chi^2_{(0.001;1;N = 192)} = 21.75$) und Mädchen ($\chi^2_{(0.001;1;N = 190)} = 40.24$) signifikant häufiger vor als Unterschätzungen.

Zusammenhang zwischen mütterlich attribuerter Leistung und Testleistung. Zur Beantwortung der Frage, ob die Elterneinschätzung mit der Testleistung zusammenhängt, wurden Pearson Korrelationen getrennt nach

Tabelle 1. Pearson-Korrelationen zwischen Testleistung und Einschätzung der Mütter in Mathematik und Sprache

	Testleistung Mathematik	Einschätzung Mathematik	Testleistung Sprache	Einschätzung Deutsch
Testleistung Mathematik		.42**	.43**	.11
Einschätzung Mathematik	.30**		.26**	.26**
Testleistung Sprache	.49**	.23**		.37**
Einschätzung Deutsch	.03	.35**	.16*	

Anmerkungen: ** $p < .01$ (2-seitig), * $p < .05$ (2-seitig). Mädchen ($n = 190$, unterhalb der Diagonalen) und Jungen ($n = 192$, oberhalb der Diagonalen). Fett markierte Werte bezeichnen die für die vorliegende Fragestellung relevanten Zusammenhänge.

Tabelle 2. Regressionskoeffizienten und Standardabweichungen für die Variable „Testleistung“ ($N = 382$)

Variable	Nichtstandardisierte Koeffizienten	Standardfehler	Standardisierte Koeffizienten
Modell mit Mathematik als Prädiktor			
Konstante	9.296***	.377	
Bildungsstand der Mutter	.358***	.098	.174
Geschlecht	-.475	.273	-.083
Einschätzung Mathematik	1.519***	.255	.394
Geschlecht 0/1 \times Einschätzung	-.434	.370	-.077
Modell mit Deutsch als Prädiktor			
Konstante	9.630***	.346	
Bildungsstand der Mutter	.216*	.089	.124
Geschlecht	-.051	.244	-.011
Einschätzung Deutsch	1.188***	.249	.349
Geschlecht 0/1 \times Einschätzung	-.690*	.345	-.144

Anmerkungen: Geschlecht: 0 = Junge, 1 = Mädchen. Variable Einschätzung wurde zentriert. Für Modell mit Mathematik als Prädiktor: $R^2 = .053$ ($p < .001$) für Schritt 1; $\Delta R^2 = .129$ ($p < .001$) für Schritt 2; $\Delta R^2 = .003$ ($p = .241$) für Schritt 3. Für Modell mit Deutsch als Prädiktor: $R^2 = .030$ ($p < .001$) für Schritt 1; $\Delta R^2 = .058$ ($p < .001$) für Schritt 2; $\Delta R^2 = .010$ ($p < .05$) für Schritt 3. *** $p < .001$; * $p < .05$. 3 % von $N = 382 =$ ohne Angaben zur Berufsbildung.

Funktionsbereich und Geschlecht der Kinder durchgeführt (Tabelle 1). Im Bereich Mathematik ergaben sich sowohl für Jungen ($r = .42, p < .01$) als auch für Mädchen ($r = .30, p < .01$) signifikante Korrelationen. Die Korrelationskoeffizienten für Mädchen und Jungen in Mathematik unterschieden sich nicht signifikant voneinander ($z = 1.4, p = .08$). Auch im Bereich Sprache fiel die Korrelation zwischen Elterneinschätzung und Testleistung sowohl für Jungen ($r = .37, p < .01$) wie Mädchen ($r = .16, p < .05$) signifikant aus, wobei der Korrelationskoeffizient bei den Jungen verglichen mit demjenigen der Mädchen signifikant höher ausfiel ($z = 2.3, p < .05$).

Die Regressionsanalyse zeigte, dass sowohl der Bildungsstand der Mutter, als auch die Einschätzung der Mutter signifikante Prädiktoren darstellten in Bezug auf die Testleistungen in Mathematik ($\beta = .17, p < .001$ resp. $\beta = .4, p < .001$) wie auch in Sprache ($\beta = .13, p < .05$ resp. $\beta = .35, p < .001$). Des Weiteren wurde deutlich, dass der Zusammenhang zwischen der mütterlichen Einschätzung (Deutsch) und der tatsächlichen Testleistung (Sprache) abhängig ist vom Geschlecht des Kindes ($\beta = .14, p < .05$). Dies war nicht der Fall bei der Einschätzung

in Mathematik ($\beta = .08, p = .241$). Insgesamt können 19 % der Varianz der Testleistungen in Mathematik und 10 % der Varianz der Testleistungen in Sprache aufgeklärt werden (s. Tabelle 2).

Geschlechtsunterschiede in der Testleistung und in der von Müttern eingeschätzten Leistung. Die mittels t -Tests für unabhängige Stichproben durchgeführten Analysen zu Gruppenunterschieden ergaben, dass für Mathematik der Geschlechtsunterschied sowohl in der Testleistung ($p < .01$, Cohen's $d = .30$) als auch in der mütterlichen Einschätzung ($p < .01, d = .30$) zugunsten der Jungen statistisch signifikant ausfiel. Bezüglich der Sprache zeigte sich kein Geschlechtsunterschied in der Testleistung. Allerdings fiel die Müttereinschätzung bei den Mädchen signifikant höher aus ($p < .01, d = -.28$) als bei den Jungen (s. Tabelle 3).

Diskussion

Die vorliegende Untersuchung verfolgte drei Ziele: Erstens sollte die Übereinstimmung der mütterlichen Ein-

Tabelle 3. Resultate der *t*-Tests und Effektstärken: Geschlechtsunterschiede in der Testleistung und in der Einschätzung der Mütter in Mathematik und Sprache/Deutsch

	Geschlecht	N	M	SD	t	df	d
Mathematik Testleistung	Mädchen	190	9.82	2.75	2.93**	380	.30
	Jungen	192	10.66	2.89			
Mathematik Mütterereinschätzung	Mädchen	190	3.27	0.71	3.02**	380	.30
	Jungen	192	3.49	0.75			
Sprache Testleistung	Mädchen	190	10.30	2.39	-0.30	380	-.03
	Jungen	192	10.23	2.44			
Deutsch Mütterereinschätzung	Mädchen	190	3.48	0.71	-2.82**	380	-.28
	Jungen	192	3.28	0.70			

Anmerkungen: ** $p < .01$ (2-seitig), d = Cohens' d , negative t - und d -Werte bezeichnen höhere Mittelwerte für Mädchen.

schätzung mit der kindlichen Testleistung in Mathematik und Sprache untersucht und geprüft werden, ob Überschätzungen häufiger vorkamen als Unterschätzungen und ob allfällige geschlechtsspezifische Effekte bei der Genauigkeitseinschätzung vorhanden sind.

Zweitens sollte der Zusammenhang zwischen mütterlicher Leistungseinschätzung in den Schulfächern Mathematik und Deutsch mit der Testleistung in denselben Bereichen unter Einbezug des Geschlechts als Moderator und des Bildungsstandes der Mutter als Kovariate analysiert werden.

Drittens sollte der Frage nachgegangen werden, ob die mütterlichen Einschätzungen der Leistungen ihrer Kinder in den Schulfächern Mathematik und Deutsch je nach Geschlecht unterschiedlich ausfallen, und ob sich in den Testresultaten geschlechtsspezifische Unterschiede finden lassen.

Bezüglich der ersten Fragestellung zeigte sich hypothesenkonform, dass bei der mütterlichen Einschätzung bei beiden Geschlechtern Überschätzungen häufiger vorkamen als Unterschätzungen. Im Vergleich zu den Testleistungen, die in etwa gleiche Anteile an über- und unterdurchschnittlichen Leistungen ausweisen, war bei der mütterlichen Einschätzung die Kategorie „überdurchschnittlich“ deutlich häufiger vertreten. Die Verzerrung in Richtung der besseren Einschätzungskategorien deckt sich mit den Beobachtungen von Rätty et al. (2002). Die vorliegende Studie konnte über bisherige Forschung hinausgehend aufzeigen, dass bei elterlichen Einschätzungen eine Überschätzung auch in den schulischen Fächern Mathematik und Deutsch vorhanden ist, da im Elternurteil eine unterdurchschnittliche Leistung viel weniger in Betracht gezogen wird. Hunt und Paraskevopoulos (1980) gehen davon aus, dass Mütter, die ihr Kind nicht gut einschätzen können, dieses entweder unter- oder überfordern. Eine Passung zwischen den an das Kind gestellten Anforderungen und seinen Möglichkeiten und Fähigkeiten beeinflusst das Erziehungsverhalten und die

kindliche Entwicklung positiv, während eine fehlende Passung zu einer gestörten Entwicklung führen kann (Grob & Jaschinski, 2003; Schrader, 2006). Eine Überschätzung seitens der Eltern kann hinderlich sein, weil Defizite beim Kind erst spät oder nicht erkannt werden oder weil Eltern aus Mangel an Akzeptanz stützende Maßnahmen nicht mittragen. Hingegen kann eine leichte Überschätzung entwicklungsförderlich wirken, wenn Erziehungspersonen dadurch dem Kind günstige Entwicklungsanreize anbieten (Deimann et al., 2005; Schrader, 2006). Bei den Müttern selbst kann die Überschätzung der kindlichen Leistungsfähigkeit stressreduzierend wirken (Deimann et al., 2005). Allerdings wurde für die Dichotomisierung von Mütterereinschätzung und Testresultat in der vorliegenden Studie der Cut-Off für die kategoriale Einteilung in „unterdurchschnittlich“, „durchschnittlich“ und „überdurchschnittlich“ posteriori festgelegt, weshalb die graduelle Komponente innerhalb der Einschätzungskategorien nicht eruiert werden kann. Somit muss davon ausgegangen werden, dass beispielsweise bei den rund 39 % der mütterlichen Überschätzungen ein Teil der Kinder von entwicklungsförderlichem Verhalten der Eltern profitieren kann, während bei denjenigen, welche von ihren Müttern graduell hoch überschätzt werden, das Risiko einer Überforderung besteht.

In Übereinstimmung mit der Literatur (Buch et al., 2006; Deimann et al., 2005; Rennen-Allhoff, 1991; Schrader, 2006) wurde ein positiver Zusammenhang zwischen Leistungseinschätzung und Testleistung in beiden Fächern gefunden, wobei die Korrelation im Bereich Sprache für die Jungen bedeutsam höher ausfiel als für Mädchen. Im Gegensatz zur Studie von Phillipson und Phillipson (2007), in welcher Korrelationen von .45 bis .62 berichtet wurden, sind die Korrelationen in der vorliegenden Studie mit .16 bis .42 eher im schwachen bis mittleren Bereich anzusiedeln. Dabei ist zu beachten, dass unterschiedliche Urteilsanker, unterschiedliche Informationsbasis und situationspezifisches Verhalten der Kinder (Döpfner, Lehmkuhl, Heubrock & Petermann, 2000;

Kuschel, Heinrichs, Bertram, Naumann & Hahlweg, 2007) eine Rolle spielen können. In der vorliegenden Untersuchung besteht die unterschiedliche Informationsbasis darin, dass die Eltern ihre Einschätzung in Bezug auf die Schulfächer Deutsch und Mathematik geben, wohingegen die Testleistung nicht die Schulleistung per se erfasst, sondern Kompetenzen, welche zwar für den Schulerfolg in Deutsch und Mathematik wesentlich sind, aber inhaltlich eher die Grundvoraussetzungen für die operative Umsetzung im Schulalltag prüfen. Dabei ist insbesondere für Deutsch darauf hinzuweisen, dass der Schwerpunkt in den IDS auf Wortschatz und Grammatik liegt, während im Schulfach Deutsch zusätzlich schrift- und metasprachliche Kompetenzen gewichtet werden. Auch das situationspezifische Verhalten der Kinder spielt eine Rolle, weil die Leistungserfassung auf dem Resultat einer einmaligen Testleistung beruht, wohingegen die Eltern für ihre Einschätzung möglicherweise die Rückmeldung der Schule und ihre eigenen Beobachtungen mit ein beziehen. Hervorzuheben ist, dass das vorliegende Untersuchungsdesign der gängigen Praxis im Beratungsbereich weitgehend entspricht. Das Resultat einer einmaligen Testung stellt einen wesentlichen Teil der professionell generierten Informationen über das Kind dar und fließt in die Beratungssituation ein. In dieser treffen demnach zwei auf verschiedene Weise generierte Informationen aufeinander. Es ist daher wichtig zu eruieren, inwiefern die Einschätzungen beider Informanten (Eltern und Beratungsperson) übereinstimmen, da die Gestaltung des Beratungsgesprächs wesentlich davon abhängt.

Der prädiktive Wert der Müttereinschätzungen unter Berücksichtigung des Bildungsstandes leistet einen signifikanten, wenn auch eher kleinen Beitrag bei der Vorhersage der Testleistungen. Dies könnte auch daran liegen, dass der Bildungsstand der Mütter in der vorliegenden Stichprobe nur einen kleinen Anteil von Müttern ohne Berufsbildung aufweist. Der stärkere Zusammenhang bei den Jungen könnte ein Hinweis darauf sein, dass die Leistungen bei Jungen insbesondere im Sprachbereich sensibler beobachtet werden als bei Mädchen.

Bezüglich der Geschlechtsunterschiede wurde angenommen, dass Mütter die Leistungen von Jungen in Mathematik höher einschätzen als die Leistungen von Mädchen. Tatsächlich konnte ein signifikanter Geschlechtsunterschied sowohl in der mütterlichen Einschätzung als auch in der Testleistung zugunsten der Jungen nachgewiesen werden, was mit bisherigen Forschungsergebnissen übereinstimmt (Furnham, 2000; Lummis & Stevenson, 1990; Rätty et al., 2002; Tiedemann, 2000). In Bezug auf die Elterneinschätzung konnte Tiedemann (2000) nachweisen, dass elterliche Stereotypen bezüglich mathematischer Kompetenzen im 3. und 4. Grundschuljahr das kindliche Selbstkonzept bezüglich des Leistungspotentials in Mathematik im Sinne einer selbsterfüllenden Prophezeiung beeinflussen können. Es wird davon ausgegangen, dass der Geschlechtereffekt in Zusammenhang

stehen könnte mit dem in der Gesellschaft etablierten Rollenverständnis. Rätty et al. (2002) sprechen von einem „klassisch geschlechtsbezogenen Attributionsmuster“ bei den Eltern, welches den Jungen grundsätzlich mehr Talent in Mathematik zugesteht als den Mädchen. Immerhin konnten Niklas und Schneider (2012) bereits bei Kindern ab der 1. Klasse moderate, aber doch signifikante Unterschiede in den Leistungsmaßen feststellen, und zwar im Bereich Mathematik zugunsten der Jungen und beim Lesen und Rechtschreiben zugunsten der Mädchen, ebenso hinsichtlich der diesbezüglichen Selbstkonzeptmaße. Es wird jedoch darauf hingewiesen, dass die geschlechtsspezifischen Entwicklungstrends im ersten Schuljahr nicht hinreichend durch die Bezüge zur differenziellen Selbstkonzeptentwicklung bei Mädchen und Jungen erklärt werden können.

Des Weiteren wurde angenommen, dass Mütter die Leistungen von Mädchen in Deutsch höher einschätzen als die Leistungen von Jungen. Die vorliegenden Resultate zeigen, dass Mädchen zwar im Schulfach Deutsch von den Müttern höher eingeschätzt wurden, in der Testleistung Sprache allerdings mit Jungen vergleichbare Leistungen erzielten. Ein ähnliches Resultat wird von Lummis und Stevenson (1990) bezüglich elterlicher Einschätzung der Lesekompetenz berichtet. Interessanterweise scheint der von Tiedemann (2000) festgestellte Effekt auf die mathematischen Kompetenzen nicht die gleichen Auswirkungen auf die Sprachkompetenzen des Kindes zu haben, obwohl auch in diesem Bereich stereotype Annahmen vorhanden sind. Weshalb im sprachlichen Bereich nicht die gleichen Mechanismen wirken wie in Mathematik, wäre zu prüfen.

Einschränkungen und Stärken der Studie

In der vorliegenden Studie konnte nicht eruiert werden, worauf die Mütter sich bei ihrer Einschätzung vorrangig gestützt haben. So ist die Einschätzung der Fähigkeiten des Kindes mittels einem curriculumsfreien Testverfahren nicht gleichzusetzen mit den Noten, welche das Kind nach einzelnen Leistungstests von der Schule nach Hause bringt und auf welche sich die Mütter möglicherweise bei ihrer Einschätzung ebenfalls gestützt haben. Da es sich in der vorliegenden Studie um einen querschnittlichen Vergleich von Einschätzungen der Mütter bezüglich der Leistung ihres Kindes handelt, werden Veränderungen in der Wahrnehmung der Mutter, die allenfalls mit zunehmender Passung eintreten könnten, nicht erfasst. Zudem stand für die Eltern für die Beantwortung pro Fach nur ein Item zur Verfügung, weshalb kein Kennwert für die interne Konsistenz berechnet werden konnte. Für zukünftige Studien empfiehlt es sich, die mütterliche Einschätzung mittels mehrerer Items zu erfassen, um die Reliabilität zu erhöhen. Als Operationalisierung der Kompetenz des Kindes diene dessen Testleistung in den IDS, womit an-

genommen wird, dass diese die Kompetenzen des Kindes widerspiegelt. Zu bemerken ist, dass bei Anwendung eines Testverfahrens immer auch Messfehler in Kauf genommen werden müssen. Gleichzeitig ist der Test gewissen Störeinflüssen unterworfen und die potentielle Leistungsfähigkeit des Kindes könnte unter Umständen höher sein, als im Test gezeigt wurde. Des Weiteren ist darauf hinzuweisen, dass die Einteilung der Elterneinschätzung in die drei Kategorien „unterdurchschnittlich“, „durchschnittlich“ und „überdurchschnittlich“, obwohl auf inhaltlichen Überlegungen beruhend, ein Stück weit einer Willkür unterworfen ist. In zukünftigen Studien könnten Elterneinschätzungen, wenn zulässig, aufgrund ihrer Normalverteilung in Kategorien anhand des Mittelwertes ± 1 Standardabweichung aufgeteilt werden, um damit inhaltlichen Überlegungen entgegenzuwirken.

Als Stärke der Studie ist hervorzuheben, dass es sich um eine repräsentative Norm-Stichprobe aus der Grundschule handelt. Im Gegensatz zu einer klinischen Stichprobe waren die Mütter in ihrer Einschätzung in der Regel nicht durch Resultate von bereits erfolgten standardisierten Leistungstests beeinflusst. Mit den Untertests der IDS wurden die Fähigkeiten in Mathematik und Sprache nicht mittels klassischer Schultests, sondern relativ curriculumsfrei mit einem standardisierten Verfahren erfasst. Dies lässt eine generelle Bewertung der Fähigkeiten in den erwähnten Funktionsbereichen zu, was der Einschätzung der Eltern, welche sich wie erwähnt auf verschiedene Informationsquellen stützt, mindestens teilweise entsprechen könnte. Für die Generierung der Daten zur mütterlichen Einschätzung bezüglich Schulleistungen wurde erstmals ein solcher Fragebogen eingesetzt.

Fazit für die Praxis

Als Fazit für die Praxis kann festgestellt werden, dass sich ein positiver Zusammenhang zwischen der Einschätzung der Mütter und der Testleistung zeigte, woraus zu schließen ist, dass die Einschätzung der Mütter ein Hinweis auf die Leistung des Kindes sein kann. Trotz positiver Korrelationen konnte aufgezeigt werden, dass ein hoher Anteil an Über- oder Unterschätzung vorhanden ist, wenn Müttereinschätzungen und Testleistungen des Kindes miteinander verglichen werden, womit aus professioneller Sicht die Einschätzung der Mutter in Bezug auf die schulischen Fächer Mathematik und Deutsch als nur beschränkt valide angesehen werden muss. Dies legt nahe, dass die Einschätzung der Mutter eine testpsychologische Einschätzung ergänzen, aber nicht ersetzen kann. Davon ausgehend, dass Mütter die Leistung ihres Kindes generell eher überschätzen, ist hingegen die Wahrscheinlichkeit groß, dass bei einem Kind tatsächlich gravierende Defizite vorliegen, wenn die Mutter eine Schwäche vermutet und die Leistung ihres Kindes im unterdurchschnittlichen Bereich ansiedeln würde.

Obwohl der Einbezug elterlicher Meinung im Schulalltag zunehmend gewichtet wird, wurden bis anhin zur Übereinstimmung von Elterneinschätzung und schulischen Leistungen ihrer Kinder kaum Studien durchgeführt. Der vorliegende Beitrag erweitert das bisherige Wissen betreffend elterlicher Wahrnehmung der Leistung ihrer Kinder und bildet wertvolles Hintergrundwissen für Beratungssituationen im entwicklungspsychologischen und pädagogischen Umfeld. Werden formelle und informelle Einschätzungen in einer Beratungssituation als sich ergänzende Informationsquellen genutzt können die Chancen für die Findung eines adäquaten Unterstützungssettings für das Kind erhöht werden.

Literatur

- Aiken, L. S. & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury, CA: Sage.
- Buch, S. R., Sparfeldt, J. R. & Rost, D. H. (2006). Eltern beurteilen die Entwicklung ihrer hochbegabten Kinder. Sonderdruck aus: *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 38, 53–61.
- Davis-Kean, P. E. (2005). The influence of parent education and family income of child achievement: The indirect role of parental expectations and the home environment. *Journal of Family Psychology*, 19, 294–304.
- Deimann, P., Kastner-Koller, U., Benka, M., Kainz, S. & Schmidt, H. (2005). Mütter als Entwicklungsdiagnostikerinnen. Der Entwicklungsstand von Kindergartenkindern im Urteil ihrer Mütter. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 37, 122–134.
- Döpfner, M., Lehmkuhl, G., Heubrock, D. & Petermann, F. (2000). *Diagnostik psychischer Störungen im Kindes- und Jugendalter: Leitfaden Kinder- und Jugendpsychotherapie*. Göttingen: Hogrefe.
- Furnham, A. (2000). Parents' estimates of their own and their children's multiple intelligences. *British Journal of Developmental Psychology*, 18, 583–594.
- Gardner, H. (1983). *Frames of mind: A theory of multiple intelligences*. New York: Basic Books.
- Grob, A. & Jaschinski, U. (2003). *Erwachsen werden. Entwicklungspsychologie des Jugendalters*. Weinheim: Beltz.
- Grob, A., Meyer, C. & Hagmann-von Arx, P. (2009). *Intelligence and Development Scales (IDS)*. Bern: Hans Huber.
- Halle, T. G., Kurtz-Costes, B. & Mahoney, J. L. (1997). Family influences of school achievement in low-income african american children. *Journal of Educational Psychology*, 89, 527–537.
- Hunt, J. McV. & Paraskevopoulos, J. (1980). Children's psychological development as a function of inaccuracy of their mother's knowledge of their abilities. *The Journal of Genetic Psychology*, 136, 285–298.
- Hyde, J. S. & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, 104, 53–69.
- Karing, C. (2009). Diagnostische Kompetenz von Grundschul- und Gymnasiallehrkräften im Leistungsbereich und im Bereich Interessen. *Zeitschrift für Pädagogische Psychologie*, 23, 197–209.

- Kovas, Y., Haworth, C. M. A., Petrill, S. A. & Plomin, R. (2007). Mathematical ability of 10-year-old boys and girls: Genetic and environmental etiology of typical and low performance. *Journal of Learning Disabilities, 40*, 554–567.
- Kuschel, A., Heinrichs, N., Bertram, H., Naumann, S. & Hahlweg, K. (2007). Wie gut stimmen Eltern und Erzieherinnen in der Beurteilung von Verhaltensproblemen bei Kindergartenkindern überein? *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 35*, 51–58.
- Landerl, K., Wimmer, H. & Moser, E. (2006). *Salzburger Lese- und Rechtschreibtest (SLRT)* (2., korr. u. akt. Aufl.). Bern: Huber.
- Lee, K. (2010). Do early academic achievement and behaviour problems predict long-term effects among Head Start children? *Children and Youth Services Review, 32*, 1690–1703.
- Lobeck, A. & Frei, M. (1987). *Schweizer Rechentest 1.–3. Klasse (SR 1–3)*. Göttingen: Hogrefe.
- Lummis, M. & Stevenson, H. W. (1990). Gender differences in beliefs and achievement: A cross-cultural study. *Development Psychology, 26*, 254–263.
- Lynn, R. & Irwing, P. (2008). Sex differences in mental arithmetic, digit span, and *g* defined as working memory capacity. *Intelligence, 36*, 226–235.
- Niklas, F. & Schneider, W. (2012). Die Anfänge geschlechtsspezifischer Leistungsunterschiede in mathematischen und schriftsprachlichen Kompetenzen. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 44*, 123–138.
- Petermann, F., Stein, I. A. & Macha, M. (2006). *Entwicklungsdiagnostik mit dem ET 6–6*. (3. veränd. Aufl.). Frankfurt/M.: Harcourt Test Service GmbH.
- Phillipson, S. & Phillipson, S. N. (2007). Academic expectations, belief of ability, and involvement by parents as predictors of child achievement: A cross-cultural comparison. *Educational Psychology, 27*, 329–348.
- Räty, H., Vänskä, J., Kananen, K. & Kärkkäinen, R. (2002). Parents' explanations of their child's performance in mathematics and reading: A replication and extension of Yee and Eccles. *Sex Roles, 46*, 121–128.
- Rennen-Allhoff, B. (1991). Wie verlässlich sind Elternangaben? *Praxis der Kinderpsychologie und Kinderpsychiatrie, 40*, 333–338.
- Schrader, F.-W. (2006). Diagnostische Kompetenzen von Eltern und Lehrern. In D. H. Rost (Hrsg.), *Handwörterbuch Pädagogische Psychologie* (3., überarb. und erweit. Aufl., S. 95–100). Weinheim: Psychologie Verlag Union.
- Schneider, W., Stefanek, J. & Dotzler, H. (1997). Erwerb des Lesens und Rechtschreibens: Ergebnisse aus dem SCHOLASTIK-Projekt. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter* (S. 113–130). Weinheim: Beltz.
- Schwenck, C. & Schneider, W. (2003). Einflussfaktoren für den Zusammenhang von Rechen- und Schriftsprachleistungen im frühen Grundschulalter. *Kindheit und Entwicklung, 12*, 212–221.
- Swiatek, M. A., Lupkowski-Shoplak, A. & O'Donogue, C. C. (2000). Gender differences in above-level EXPLORE scores of gifted third through sixth graders. *Journal of Educational Psychology, 92*, 718–723.
- Tiedemann, J. (2000). Parents' gender stereotypes and teachers' beliefs as predictors of children's concept of their mathematical ability in elementary school. *Journal of Educational Psychology, 92*, 144–151.
- Waschbusch, D. A., Daleiden, E. & Drabman, R. S. (2000). Are parents accurate reporters of their child's cognitive abilities? *Journal of Psychopathology and Behavioral Assessment, 22*, 61–77.

lic.phil. Marie-Claire Frischknecht
 Dr. Giselle Reimann
 Dr. Janine Gut
 Dr. Thomas Ledermann
 Prof. Dr. Alexander Grob

Entwicklungs- und Persönlichkeitspsychologie
 Institut für Psychologie der Universität Basel
 Missionsstrasse 62 a
 4055 Basel
 Schweiz
 E-Mail: marie-claire.frischknecht@unibas.ch

Anhang

Kreuztabelle 1. Einschätzung der Mütter und Testleistungen in Mathematik bei Jungen ($N = 192$)

		Testleistung			gesamt
		unterdurchschnittlich	durchschnittlich	überdurchschnittlich	
Einschätzung der Mütter	unterdurchschnittlich	<i>n</i> = 4 2.1 %	<i>n</i> = 6 3.1 %	<i>n</i> = 1 0.5 %	<i>n</i> = 11 5.7 %
	durchschnittlich	<i>n</i> = 12 6.2 %	<i>n</i> = 70 36.5 %	<i>n</i> = 6 3.1 %	<i>n</i> = 88 45.8 %
	überdurchschnittlich	<i>n</i> = 1 0.5 %	<i>n</i> = 70 36.5 %	<i>n</i> = 22 11.5 %	<i>n</i> = 93 48.5 %
	gesamt	<i>n</i> = 17 8.9 %	<i>n</i> = 146 76 %	<i>n</i> = 29 15.1 %	<i>n</i> = 192 100 %

Anmerkungen: Chi-Quadrat = 27.615, $df = 4$, sign. (2-seitig = .001)
(dunkelgrau = genaue Einschätzung; hellgrau = Überschätzung; grau = Unterschätzung)

Kreuztabelle 2. Einschätzung der Mütter und Testleistungen in Mathematik bei ($N = 190$)

		Testleistung			gesamt
		unterdurchschnittlich	durchschnittlich	überdurchschnittlich	
Einschätzung der Mütter	unterdurchschnittlich	<i>n</i> = 6 3.2 %	<i>n</i> = 10 5.3 %	<i>n</i> = 0 0.0 %	<i>n</i> = 16 8.5 %
	durchschnittlich	<i>n</i> = 11 5.8 %	<i>n</i> = 91 47.9 %	<i>n</i> = 9 4.7 %	<i>n</i> = 111 58.4 %
	überdurchschnittlich	<i>n</i> = 3 1.6 %	<i>n</i> = 51 26.8 %	<i>n</i> = 9 4.7 %	<i>n</i> = 63 33.1 %
	gesamt	<i>n</i> = 20 10.5 %	<i>n</i> = 152 80 %	<i>n</i> = 18 9.5 %	<i>n</i> = 190 100 %

Anmerkungen: Chi-Quadrat = 17.036, $df = 4$, sign. (2-seitig = .002)
(dunkelgrau = genaue Einschätzung; hellgrau = Überschätzung; grau = Unterschätzung)

Kreuztabelle 3. Einschätzung der Mütter „Deutsch“ und Testleistungen „Sprache“ Jungen ($N = 192$)

		Testleistung			gesamt
		unterdurchschnittlich	durchschnittlich	überdurchschnittlich	
Einschätzung der Mütter	unterdurchschnittlich	<i>n</i> = 4 2.1 %	<i>n</i> = 13 6.8 %	<i>n</i> = 0 0.0 %	<i>n</i> = 17 8.9 %
	durchschnittlich	<i>n</i> = 12 6.2 %	<i>n</i> = 95 49.5 %	<i>n</i> = 8 4.2 %	<i>n</i> = 115 59.9 %
	überdurchschnittlich	<i>n</i> = 2 1.0 %	<i>n</i> = 50 26 %	<i>n</i> = 8 4.2 %	<i>n</i> = 60 31.2 %
	gesamt	<i>n</i> = 18 9.4 %	<i>n</i> = 158 82.3 %	<i>n</i> = 16 8.3 %	<i>n</i> = 192 100 %

Anmerkungen: Chi-Quadrat = 9.664, $df = 4$, sign. (2-seitig = .046)
(dunkelgrau = genaue Einschätzung; hellgrau = Überschätzung; grau = Unterschätzung)

Kreuztabelle 4. Einschätzung der Mütter „Deutsch“ und Testleistungen „Sprache“ Mädchen ($N = 190$)

		Testleistung			gesamt
		unterdurchschnittlich	durchschnittlich	überdurchschnittlich	
Einschätzung der Mütter	unterdurchschnittlich	<i>n</i> = 2 1.1 %	<i>n</i> = 5 2.6 %	<i>n</i> = 0 0.0 %	<i>n</i> = 7 3.7 %
	durchschnittlich	<i>n</i> = 13 6.9 %	<i>n</i> = 74 38.9 %	<i>n</i> = 15 7.9 %	<i>n</i> = 102 53.7 %
	überdurchschnittlich	<i>n</i> = 3 1.6 %	<i>n</i> = 69 36.3 %	<i>n</i> = 9 4.7 %	<i>n</i> = 81 42.6 %
	gesamt	<i>n</i> = 18 9.5 %	<i>n</i> = 148 77.9 %	<i>n</i> = 24 12.6 %	<i>n</i> = 190 100 %

Anmerkungen: Chi-Quadrat = 9.038, $df = 4$, *n.s.*

(dunkelgrau = genaue Einschätzung; hellgrau = Überschätzung; grau = Unterschätzung)

Studie 3

Reimann, G., Stoecklin, M., Lavalée, K., Gut, J., Frischknecht, M.-C. & Grob, A. (2013).

Cognitive and motivational profile shape predicts mathematical skills over and above profile

level. *Psychology in the Schools*, 50(1). doi: 10.1002/pits.21659

COGNITIVE AND MOTIVATIONAL PROFILE SHAPE PREDICTS MATHEMATICAL SKILLS OVER AND ABOVE PROFILE LEVEL

GISELLE REIMANN, MARKUS STOECKLIN, KRISTEN LAVALLEE, JANINE GUT, MARIE-CLAIRE FRISCHKNECHT, AND ALEXANDER GROB

Faculty of Psychology, University of Basel, Switzerland

The interpretation of subtest profiles from intelligence testing remains popular among many practitioners who use subtest performance to draw diagnostic conclusions, in spite of criticism by some researchers, who point to the low reliability and predictive validity of subtest scores in predicting achievement outcomes. Prior research outlines two approaches to the study of subtest variation: the examination of interindividual variation in specific cognitive domains or subtests as compared to a standard sample, and the examination of intraindividual strengths and weaknesses, regardless of overall level. The present study seeks to add to knowledge in this field with data from 567 children ages 5 to 10 years who exhibit meaningful subtest variation on a new test of intellectual abilities. Results from the present sample point to statistically significant utility, with small to medium effects, of intraindividual cognitive and motivational profile shape over and above profile level in predicting mathematical skills. We discuss implications for school psychological and educational assessment research. © 2012 Wiley Periodicals, Inc.

The interpretation of factor and subtest scores has waxed and waned and waxed again in the field of school psychology over the past 30 years. As of 2000, 89% of U.S. school psychologist practitioners regularly made clinical use of Wechsler Intelligence Scale for Children-III (WISC-III; Wechsler, 1991) subtest profile analysis (Pfeiffer, Reddy, Kletzel, Schmelzer, & Boyer, 2000), with 70% of practitioners viewing profile analysis, or the interpretations of the pattern of intraindividual strengths and weaknesses regardless of level of functioning, as a particularly valuable tool. Contemporary psychologists often utilize a so-called “top-down-approach,” first interpreting global scores and then analyzing specific scale and subtest scores with the goal of more accurately describing individual strengths and deficits (Daniel, 2007; Kahana, Youngstrom, & Glutting, 2002; Smith & Watkins, 2004; Watkins, Glutting, & Youngstrom, 2005). At the same time, the value of profile analysis has been critiqued and deemed a “shared professional myth” unsupported by research evidence (Watkins, 2000), with resulting advice that psychologists steer clear of interpreting profiles or using them to make predictions about student achievement (Glutting, Watkins, Konold, & McDermott, 2006).

In this area of research, two aspects of substrate performance lend themselves to examination. The first is interindividual weakness on specific cognitive factors or subtests as compared with a reference group (i.e., normative comparison). The second is relative intraindividual weakness in one or more areas as compared with intraindividual strengths (i.e., ipsative comparison), regardless of level (Naglieri, 2000; Watkins, 2000).

Interindividual Differences in Cognitive Abilities as Predictors of Academic Achievement

Interindividual differences in substrate cognitive abilities predict academic achievement across several studies. In the area of mathematics, studies have implicated the Cattell-Horn-Carroll factors of comprehension-knowledge, fluid reasoning, short-term memory, working memory, processing speed, and long-term retrieval—but not visual-spatial thinking—as significant concurrent predictors of mathematics achievement (Floyd, Evans, & McGrew, 2003). Longitudinal studies indicated that

Correspondence to: Giselle Reimann or Alexander Grob, Faculty of Psychology, University of Basel, Missionstrasse 60/62, 4055 Basel, Switzerland. E-mail: giselle.reimann@unibas.ch or alexander.grob@unibas.ch

visual–spatial short-term memory and working memory predict math achievement over time, and executive functioning skills predict general achievement over time (Bull, Espy, & Wiebe, 2008). Notably, these studies did not control for general IQ, and, as Glutting et al. (2006) pointed out, there may be no explanatory utility in examining multiple subtest scores if a single parsimonious score (i.e., IQ) explains the variance just as well. However, one longitudinal study did control for general IQ, finding that working memory is a better predictor of achievement than general IQ, accounting for more unique variance when both are included in the same model (Alloway, 2009).

Contradictory evidence also exists, finding no or only little predictive utility in any composite or subtest score above that of IQ when IQ is accounted for on the first step of a regression analysis and the composite or subtest scores are entered second (Glutting et al., 2006; Kahana et al., 2002; Rohde & Thompson, 2007). However, regression analyses including both single cognitive subtest scores as well as full-scale IQ suffer from multicollinearity (Elliott, Hale, Fiorello, Dorvil, & Moldovan, 2010; Fiorello et al., 2007), with IQ entered on the first regression step leaving little variance left for remaining collinear predictors (Fiorello et al., 2007; Hale, Fiorello, Kavanagh, Hoepfner, & Gaither, 2001; Hale, Fiorello, Kavanagh, Holdnack, & Aloe, 2007; Haynes & Lench, 2003). Thus, some have argued that it may be inappropriate and misleading to use hierarchical regression analyses that include global IQ scores in the same analysis as subtest scores (Hale et al., 2008). Alternatives, such as commonality analysis (e.g., Fiorello, Hale, McGrath, Ryan, & Quinn, 2002), have been proposed but have also been heavily criticized (e.g., Watkins, Glutting, & Lei, 2007).

Intraindividual Differences in Cognitive Abilities as Predictors of Academic Achievement

Cognitive skill profiles can be described in terms of profile level, scatter, or shape. Profile level refers to interindividual differences and is defined as an individual's standardized mean score across subtests (or overall IQ when referring to cognitive scores), as compared with the reference group (i.e., standardization sample). In contrast to interindividual differences, intraindividual differences refer to within-person strengths and weaknesses and involve both profile scatter and shape. Profile scatter is defined as the extent to which scores deviate from an individual's mean. Profile shape (or pattern) refers to subtest peaks and valleys in the individual's score profile, or the pattern of individual strengths and weaknesses across subtests (Davison & Kuang, 2000; Ding, 2001). Profile shape can also be described as the rank order of subtest scores for an individual (Watkins & Glutting, 2000). Two profiles with the exact same level and scatter may have differing shape or patterns of high and low points (Watkins & Glutting, 2000; see Figure 1 for a graphical display of profile level, scatter, and shape). As opposed to the interpretation of single subtest or factor scores, the analysis

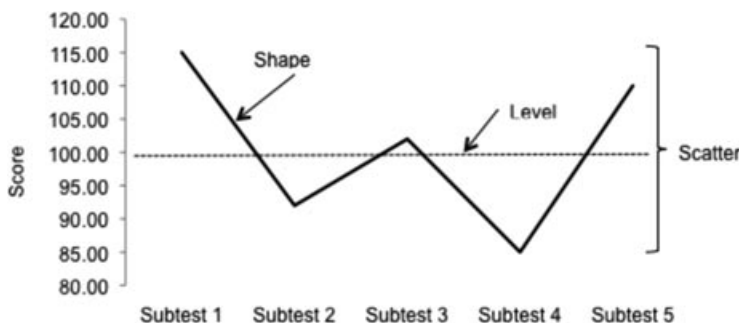


FIGURE 1. Graphical display of the three profile characteristics level, scatter, and shape. Level corresponds to the mean of all subtests. Scatter represents the variation of the subtests from the mean. Shape depicts the combination of strengths and weaknesses over the subtests.

of profile shape offers the possibility to interpret the combination of strengths and weaknesses as a whole, simultaneously and independently of the mean profile level, and to compare level and shape in concert without the problem of multicollinearity. Critics usually do not question the existence of profile shapes, but rather their diagnostic utility, reliability, and predictive validity (Davison & Kuang, 2000). Although profile level is accepted as an informative component in differentiating between groups of children with and without diagnosed learning disorders, profile shape does not typically provide additional information (Rizza, McIntosh, & McCunn, 2001; Robinson & Harrison, 2005). When it comes to predictive or incremental validity, research also indicates that profile level is the best predictor of educational outcomes, whereas profile shape adds minor incremental validity at best (Kahana et al., 2002; Watkins & Glutting, 2000). Further, Borsuk, Watkins, and Canivez (2006) demonstrated that membership in WISC-III subtest profile-type groups was generally not stable over a 3-year period (with the exception of two profiles, one of them containing a high frequency of learning-disabled children). The low reliability of subtest scores in comparison to full-scale IQ scores implies that profile analysis should not be used for classification purposes or for diagnostic labeling (Sattler & Saklofske, 2001).

Despite these arguments against the analysis of profile shape, some researchers have continued to argue in favor of it, positing that the analysis of profile shape may provide useful information about a child's cognitive strengths and weaknesses (Sattler & Saklofske, 2001). A number of studies have found evidence that profile shape information may actually be of additional value when predicting educational outcomes (Kim, Frisby, & Davison, 2004; Naglieri, 2000; Pritchard, Livingston, Reynolds, & Moses, 2000). Thus, although perhaps not as useful as IQ in predicting achievement, profile shape information could potentially be useful in explaining why an individual low-achieving student is so. Moreover, some studies demonstrated that profile level might not always be the best source of information, particularly when high discrepancies exist among subtest scores (Fiorello et al., 2002) or in cases of exceptionality, such as giftedness (Lohman, Gambrell, & Lakin, 2008), learning disabilities (Elliott et al., 2010; Fiorello et al., 2007; Hale et al., 2001), attention-deficit/hyperactivity disorder, and traumatic brain injury (Fiorello et al., 2007; Fiorello et al., 2002). Results reported by Fiorello et al. (2002) suggest that for children with a flat profile, the general profile level accurately reflects intellectual functioning and might therefore be the best predictor of educational outcomes. In the case of substantial subtest scatter, however, the interpretation of profile level might be inappropriate (Fiorello et al., 2002; Hale et al., 2001). Hale et al. (2001) offer a comparison: "If you put one hand in boiling water and the other in ice water, is it accurate to conclude that, on average, everything is fine?" (p. 34). The interpretation of subtest or factor scores in cases of high intraindividual subtest variability has also been hotly debated and contested, with results from Watkins et al. (2007) indicating that full-scale IQ scores are fully valid even in the face of high subtest variability (see also Daniel, 2007; Freberg, Vandiver, Watkins, & Canivez, 2008; Kotz, Watkins, & McDermott, 2008).

Achievement Motivation as Predictor of Academic Achievement

To our knowledge, only cognitive abilities have been included in profiles so far in the prediction of educational outcomes. However, achievement is not only predicted by cognitive abilities. Even a child with strong cognitive abilities will not necessarily meet his or her potential without motivation (Spinath, Spinath, Harlaar, & Plomin, 2006). Achievement motivation refers to specialized motivation "relevant to performance on tasks in which standards of excellence are operative" (Wigfield, Eccles, Roeser, & Schiefele, 2008, p. 406). Both typical classroom situations, as well as test sessions, are settings in which such standards can operate and, therefore, in which achievement motivation becomes relevant. Observable behavioral aspects of achievement motivation include persistence

(i.e., the tendency to continue trying, even as difficulty increases) and intrinsic motivation (i.e., the joy expressed when engaging in a task without the lure of external motivators).

Persistence, or self-disciplined effort, is generally believed to lead to greater success in young children, with older children understanding that the effects of effort are limited by ability (Folmer et al., 2008). With increasing task difficulty, more self-discipline is needed in order to persist and, eventually, succeed. Self-discipline seems to be crucial with respect to school achievement, with findings that various measures of self-discipline predicted academic performance even more robustly than IQ in a sample of adolescents (Duckworth & Seligman, 2005).

Intrinsic motivation, or the motivation to engage in an activity for its own sake (Pintrich & Schunk, 2002), also predicts academic achievement independently of intelligence in primary school children (Gottfried, 1990; Spinath et al., 2006), explaining 3% to 6% of incremental variance over and above general intelligence. A substantial amount of variance in achievement is predicted by a shared contribution of intelligence and motivation (Spinath et al., 2006), speaking in favor of examining the two constructs together when predicting educational outcomes to account for the overlap. Further, high intrinsic motivation may be especially important for low achieving children. For example, intrinsic motivation explains a significant amount of longitudinal growth in reading comprehension in children with low reading skills, but not in already-strong readers (Logan, Medford, & Hughes, 2011).

Persistence and intrinsic motivation seem to be interrelated in that children with high intrinsic motivation usually exhibit more persistence when engaging in a task (Ryan & Deci, 2000). To our knowledge, no study has investigated the effect of intrinsic motivation or persistence in combination with specific cognitive abilities on mathematical skills, neither in the form of regression analysis nor with profiles. Regarding intraindividual strengths and weaknesses, investigation of the buffering effects of motivation strengths against cognitive weaknesses, and vice versa, is needed and would be interesting. Further, that higher achievement motivation always translates to higher test performance should not be taken for granted. Extremely motivated students may, for example, attempt to process tasks with higher speed, which can in turn lead to more errors and, eventually, lower achievement (Brunstein & Heckhausen, 2006).

Summary and Purpose of the Study and Research Questions

In sum, *interindividual differences* in cognitive skills predicted academic achievement in past research. Deficits in specific cognitive abilities such as comprehension-knowledge, fluid reasoning, short-term memory, working memory, processing speed, and long-term retrieval are associated with lower performance in math (Alloway, 2009; Bull et al., 2008; Floyd et al., 2003). In addition, achievement motivation is also related to academic achievement, in the forms of persistence/self-discipline (Duckworth & Seligman, 2005) and intrinsic motivation (Gottfried, 1990; Spinath et al., 2006). Finally, overall profile level is consistently the strongest predictor of educational outcomes (Kahana et al., 2002; Watkins & Glutting, 2000).

Regarding *intraindividual differences*, profile shape generally provides little incremental value over profile level (Kahana et al., 2002; Watkins & Glutting, 2000). Despite these general findings, profile shape may better reflect intellectual functioning in children with substantial variability in subtest scores than a single global score alone (Fiorello et al., 2007; Fiorello et al., 2002; Lohman et al., 2008). Further, although achievement motivation has not been included in cognitive profiles in the past, evidence indicates that academic performance may be best predicted by variance shared by cognitive and motivational variables together (Spinath et al., 2006). In particular, high motivation appears to be especially important for children with low initial academic skills (Logan et al., 2011). Whether it is also important for children with low cognitive skills is an important remaining question.

The aims of the present study are twofold. The first aim is to investigate the relationship between *interindividual differences* in cognitive subskills and mathematical skills. The Intelligence and Development Scales for children aged 5 to 10 years (IDS; Grob, Meyer, & Haggmann-von Arx, 2009) offer an opportunity to address this question as the test separates fluid cognition from more crystallized abilities like language and mathematical skills, the latter which do not factor into the full-scale IQ. The IDS full-scale IQ represents more fluid aspects of intellectual ability, whereas mathematical skills are considered separately and are more dependent on educational experience. It was expected that all cognitive variables measured by the IDS would significantly predict mathematical skills. IQ was not included in this analysis in order to avoid multicollinearity problems. Further, it was also expected that achievement motivation would provide additional explanatory value in predicting mathematical skills.

The second aim was to extend previous research on the explanatory utility of *intraindividual differences*. The goal was to identify the most frequent profile shapes on the IDS (Grob et al., 2009) across seven cognitive and two motivational subtests, and to assess their incremental criterion validity (see Haynes & Lench, 2003, for a definition) with respect to mathematical skills over and above profile level, using Profile Analysis via Multidimensional Scaling (PAMS; e.g., Kim et al., 2004). Only children with high intraindividual variance in subtest scores were included in the study, as the purpose was to investigate the usefulness of profiles in the presence of meaningful subtest variance and because profile shape analyses do not make sense in children with little subtest variation (i.e., flat profiles). It was expected that profile level would be the strongest predictor of mathematical skills but that profile shape may add incremental validity to the prediction of mathematics. A further point of investigation was whether intraindividual cognitive weaknesses in combination with an intraindividually high amount of achievement motivation would translate to better mathematical skills as compared with any different profile shape, or whether this profile would be associated with lower mathematical skills, presumably because of faster but less thorough processing (Brunstein & Heckhausen, 2006).

In sum, the two major research questions are:

1. Do *interindividual differences* in cognitive substrates and motivational variables predict mathematical skills?
2. Do *intraindividual differences*, in the form of profile shape in cognitive and motivational variables, predict mathematical skills over and above overall profile level?

Contributions of this study to the literature include further attention to the debate on the utility of cognitive substrates, the inclusion of both cognitive and motivational variables in the same profile analysis, a test of the utility of profiles in a nonclinical population using data from a test battery other than the often-used Wechsler Scales, namely the German-language IDS (Grob et al., 2009), and a supplemental contribution of investigating the efficacy of the PAMS procedure.

METHOD

Participants

Participants were drawn from the IDS standardization sample ($n = 1,330$; 668 girls and 662 boys). Participants with any missing data in the cognitive and achievement motivation variables were excluded, as the statistical procedure required data to be complete for all cases, resulting in a sample of 1,145 children (581 girls and 564 boys) aged 5 to 10 years ($M = 7.99$, $SD = 1.62$). The 185 children who were excluded from the present study due to missing data did not differ from the 1,145 included children with respect to sex, age, IQ, maternal or paternal education. The standard deviation on all profile variables (i.e., the seven cognitive variables and the two achievement motivation variables

together) was computed separately for each child, resulting in a new variable termed intraindividual variability. Children with an intraindividual variability score above the median were included in the final sample ($n = 567$; 303 girls and 264 boys; mean age = 8.03 years, $SD = 1.63$). Of these 567 children, 54% were assessed in Switzerland, 27% in Germany, and 19% in Austria. A total of 87% were native German speakers, and 13% reported another native language but spoke German fluently enough to follow test instructions. All children belonged to a nonclinical sample, meaning that all children attended regular school (i.e., they did not attend special schools for children with severe learning delays or challenges). Parents were asked to fill in a background questionnaire. Based on parental reports, 2.8% of the children had a language impairment, 2.1% suffered from hyperactivity, and 6.5% suffered from attention-deficit disorders. A total of 4.2% of children were born preterm. We decided not to exclude these children from the analyses because they attended regular school classes. As children in the standardization sample were recruited from “regular” schools, children with more significant learning disabilities, developmental delays, or intellectual impairments (who typically attend special education schools in German-speaking regions) were by default not included in the standardization sample (although the IDS was later validated with such groups of children). Therefore, they were not included in this study.

Measures

The IDS (Grob et al., 2009) comprise a comprehensive assessment battery of child cognitive and general functioning. The IDS allow the examiner to assess general intellectual ability as well as general development across six developmental domains (Psychomotor, Cognition, Mathematical Skills, Language Skills, Social–Emotional Competence, and Achievement Motivation) for children aged 5 to 10 years (for more information see Hagmann-von Arx, Meyer, & Grob, 2008). For the purpose of the present study, only the domains of Cognition and Achievement Motivation were included in building the profiles, as the focus was on aspects of functioning hypothesized to have the most direct effect on academic achievement. Mathematical Skills served as the dependent variable. All subtest scores were age-standardized, with a mean of 10 points and a standard deviation of 3 points.

The test procedure and instructions of the IDS are highly standardized. Regarding reliability, internal consistency indices ranged between .57 (Gross Motor Skills) and .96 (Selective Attention) in the standardization sample, and for the variables under investigation in the present study all indices are generally satisfactory (Cronbach’s alphas for the present sample will be reported in the following sections). The test–retest reliability of the IDS was investigated in a subsample of the standardization sample ($n = 31$, mean age = 7.5 years, $SD = 1.5$ years at Time 1; Grob et al., 2009), with an average interval of 15 months between the two measurement points. The examiners were not the same at Time 2; therefore this measure can also be seen as a delayed form of interrater reliability. The retest reliability for the variables used in this study were satisfactory (Visual Perception: $r = .62$, $p < .001$; Selective Attention: $r = .75$, $p < .001$; Visual–Spatial Working Memory: $r = .48$, $p < .01$; Phonological Span: $r = .45$, $p < .05$; Verbal Long-Term Memory: $r = .81$, $p < .001$; Figural Reasoning: $r = .74$, $p < .001$; Conceptual Reasoning: $r = .52$, $p < .01$; Achievement Motivation I: $r = .71$, $p < .001$; Achievement Motivation II: $r = .41$, $p < .05$; Mathematical Skills: $r = .88$, $p < .001$; IQ: $r = .83$, $p < .001$).

Validity analyses indicate age trends, with clear improvements in performance with age across all subtests except Achievement Motivation II (intrinsic motivation), which was expected because young children can be as motivated as older children (Grob et al., 2009). Factor analysis indicated four factors: Cognition (explaining 15.50% of the variance, including Visual Perception, Selective Attention, Visual–Spatial Working Memory, Phonological Span, Verbal Long-Term Memory,

Table 1
 Correlations Between IDS Cognitive Subtests and HAWIK-IV Index Scores as Well as Full-Scale IQ for a Validation Sample ($n = 172$; Grob et al., 2009)

IDS Cognitive Subtests	HAWIK-IV index scores				
	Verbal Comprehension	Perceptual Reasoning	Working Memory	Processing Speed	HAWIK-IV Full-Scale IQ
Visual Perception (VP)	.11	.35***	.21**	.10	.28***
Selective Attention (SA)	.26**	.19*	.35***	.38***	.42***
Visual–Spatial Working Memory (VSWM)	.19*	.25**	.16*	.23**	.24**
Phonological Span (PS)	.31***	.33***	.47***	.20**	.45***
Verbal Long-Term Memory (VLTM)	.46***	.28***	.37***	.12	.45***
Figural Reasoning (FR)	.24**	.50***	.24**	.20*	.40***
Conceptual Reasoning (CR)	.28***	.43***	.31***	.10	.42***
IDS Full-Scale IQ	.48***	.62***	.56***	.34***	.69***

* $p < .05$. ** $p < .01$. *** $p < .001$.

Figural Reasoning, Conceptual Reasoning, Mathematical Skills, and Language Skills), Social–Emotional Competence (9.59% variance explained, including Understanding of Emotions, Regulation of Emotions, Understanding of Social Situations, Acting in Social Situations), Achievement Motivation (8.91% variance explained, including Achievement Motivation I and II), and Psychomotor (8.13% variance explained, including Gross Motor, Fine Motor, and Visual Motor Skills). Concurrent validity tests indicated correlations between cognitive subtests and the German version of the WISC-IV, Mathematical Skills and a Swiss mathematical achievement test, and Achievement Motivation and reports by parents and teachers (all results are reported in the following and can be found in Grob et al., 2009). Moreover, analyses examining differential validity indicated that most subtests differentiated controls from children with giftedness, learning disabilities, attention-deficit hyperactivity disorder (ADHD), Asperger’s syndrome, and behavior disorders (for more information see Grob et al., 2009).

Cognition. The cognitive domain consists of seven subtests, which can be averaged to produce a general IQ score ($M = 100$, $SD = 15$, Cronbach’s alpha = .92), which then correlates at .83 ($p < .001$) with the Full Scale IQ from the German version of the WISC-IV (HAWIK-IV; Petermann & Petermann, 2007) in a subset of 172 children from the validation sample (Grob et al., 2009). Correlations among the cognitive subtests of the IDS and the HAWIK-IV index scores are presented in Table 1.

The *Visual Perception* (VP) subtest assesses the ability to differentiate between visual stimuli by requiring the child to sequence cards with printed lines of different lengths (Cronbach’s alpha = .80). The *Selective Attention* (SA) subtest is a timed test assessing the ability to focus attention under speed by requiring the child to cross out printed ducks according to specified characteristics, ignoring other irrelevant characteristics (Cronbach’s alpha = .96). The *Visual–Spatial Working Memory* (VSWM) subtest assesses the child’s visual–spatial memory ability by requiring children to remember geometric forms and recognize them in a series of other forms, regardless of their changing color (Cronbach’s alpha = .73). The *Phonological Span* (PS) subtest assesses forward letter and digit span, requiring children to reproduce a series of orally presented letters and digits of increasing lengths (Cronbach’s alpha = .77). The *Verbal Long-Term Memory* (VLTM) subtest assesses long-term memory capacities, requiring children to reproduce a short story after several distractor tasks across 20 minutes (Cronbach’s alpha = .88). The *Figural Reasoning* (FR) subtest

Table 2
Bivariate Correlations for Study Variables (n = 567)

Variable	VP	SA	VSWM	PS	VLTM	FR	CR	AM I	AM II	MS	IQ
Visual Perception (VP)	–										
Selective Attention (SA)	.10	–									
Visual–Spatial Working Memory (VSWM)	.09	.01	–								
Phonological Span (PS)	.14	.06	.10	–							
Verbal Long-Term Memory (VLTM)	.13	.13	.15	.16	–						
Figural Reasoning (FR)	.21	–.01	.14	.08	.10	–					
Conceptual Reasoning (CR)	.10	–.01	.20	.05	.17	.16	–				
Achievement Motivation I (Persistence) (AM I)	.14	.15	.19	.19	.24	.11	.11	–			
Achievement Motivation II (Intrinsic Motivation) (AM II)	.18	.17	.09	.15	.23	.10	.14	.51	–		
Mathematical Skills (MS)	.33	.12	.28	.31	.28	.34	.24	.18	.22	–	
Full-Scale IQ (IQ)	.52	.38	.49	.47	.54	.50	.48	.33	.31	.56	–
Loadings on g-factor	.53	.23	.53	.44	.57	.53	.53	–	–	–	–

Note. Correlations printed in bold are significant ($p < .05$).

assesses the ability to reason about spatial information by requiring children to reproduce printed patterns with wooden rectangles and triangles (Cronbach's alpha = .75). The *Conceptual Reasoning* (CR) subtest assesses the ability to engage in inductive reasoning by requiring children to group pictures by concept (does not require explicit naming) (Cronbach's alpha = .78). Starting and ending points for each subtest are based on age and performance. Each of the seven subtests yields an age-standardized score ($M = 10$, $SD = 3$). Loadings on the g-factor and intercorrelations for all cognitive subtests are provided in Table 2.

Achievement Motivation. As self-report measures of motivation can be confounded by language abilities and self-awareness in young children (Fulmer & Frijters, 2009), achievement motivation was assessed via questionnaire by test examiners based on behavioral observations of task engagement or avoidance (Fulmer & Frijters, 2009) during testing. Behavioral observations tend to be more context-specific and less prone to social desirability and misinterpretation in young children. The Achievement Motivation questionnaire in the IDS is a standardized adaptation of the Achievement Motivation Inventory (*Leistungsmotivationsinventar, LMI*; Schuler & Prochaska, 2001) and contains eight items rated by the examiner after the testing session on a scale from 1 (low) to 4 (high). The first four items assess persistence and self-discipline (e.g., “The child forced him or herself to concentrate on the tasks even s/he didn't experience a great amount of pleasure or when s/he was tired”) while working on the IDS tasks, especially as tasks increase in difficulty (further referred to as *Achievement Motivation I* or AM I). In the standardization sample, examiner-rated persistence (AM I) was correlated with parent-reported school grades in mathematics ($r = .22$, $p < .01$) and German language ($r = .29$, $p < .001$) as an indication of criterion validity. The retest reliability in the previously mentioned subsample ($n = 31$; Grob et al., 2009) was $r = .71$ ($p < .001$), suggesting that self-discipline is measured as a relatively enduring, more trait-like characteristic.

The second four items assess intrinsic motivation, operationalized by the amount of joy and pride a child exhibits while working on the IDS tasks (e.g., “The child liked to engage in new tasks”), especially when difficult tasks were solved correctly (further referred to as *Achievement Motivation II* or AM II). In the standardization sample, examiner-rated intrinsic motivation (AM II) was correlated with school grades in mathematics ($r = .31$, $p < .001$) and German language ($r = .32$, $p < .001$) reported by parents as an indication of criterion validity. The retest reliability in the

previously mentioned subsample ($n = 31$; Grob et al., 2009) was $r = .41$ ($p < .05$), suggesting that intrinsic motivation was measured as a state rather than a trait.

The total scores for AM I and AM II, respectively, were age-standardized ($M = 10$, $SD = 3$). Cronbach's alphas were .86 for AM I and .76 for AM II, indicating satisfactory internal reliability. Analyses of concurrent validity indicated that the mean score of AM I and AM II taken together was positively correlated with teacher-reported achievement motivation (using the same scales, $r = .31$, $p < .001$) as well as with parent-reported achievement motivation ($r = .15$, $p < .01$) in the standardization sample (Grob et al., 2009).

Mathematical Skills. The subtest Mathematical Skills (MS) assesses mathematical thinking and calculation skills using 18 items that line up with the standard Swiss school curriculum. Starting and ending points are based on age and performance. The test measures arithmetic skills such as counting, assessing quantities, performing multiplication and division, inferring the number of beads in a two-dimensional picture of a three-dimensional pyramid, and solving more complex equations. The sum score for each child was translated into an age-standardized score, as on the other IDS subtests. Cronbach's alpha for the Mathematical Skills score was .88. The retest reliability in the previously mentioned subsample ($n = 31$; Grob et al., 2009) was $r = .88$ ($p < .001$). Moreover, this subset correlated with a standard Swiss math achievement test (Lobeck & Frei, 1987; $r = .41$, $p < .001$), suggesting a satisfactory concurrent validity.

Sociodemographic Variables: Parent Education, Child's Age, and Sex. Sociodemographic information was collected from permission forms (child birth date and sex) and a demographics questionnaire (parent education). In the present sample, parents of 418 children completed the demographics questionnaire. Parents' education was assessed for mothers and fathers separately using five categories (1 = obligatory school education only (equivalent to 10th-grade education in the United States; 9% of the mothers and 5% of the fathers who completed the questionnaire), 2 = vocational education (38% mothers and 33% fathers), 3 = advanced secondary education (equivalent to 1 year more than U.S. high school diploma; 6% mothers and 3% fathers), 4 = vocational university (17% mothers and 22% fathers), 5 = university or college (30% mothers and 37% fathers). Mothers' mean educational level was 3.20, $SD = 1.43$, and fathers' mean education level was $M = 3.55$, $SD = 1.39$, respectively. Child age was calculated from the child's birth date and the testing date.

PROCEDURE

Children were recruited from nursery schools, kindergartens, and primary schools in Germany, Austria, and the German speaking part of Switzerland. Children whose primary caretakers gave permission were tested individually by trained school psychologists as well as graduate and post-graduate students. Testing took place in a quiet room located at the institution of recruitment, the school psychology services building, or at the university of author affiliation, and lasted between 90 and 120 minutes.

Data Analysis

The value of each individual subtest in the prediction of Mathematical Skills in the *interindividual* approach was assessed using multiple regression. Whether and how the combination of weaknesses and strengths influences Mathematical Skills in the *intraindividual* approach was assessed using PAMS. Several authors recommend PAMS for processing profile level and profile shape independently (Davison, Kim, & Ding, 2001; Ding, 2001; Ding, 2005; Ding, 2006; Frisby & Kim, 2008; Kim, 2010; Kim et al., 2004). This technique is suitable when normative profiles are to be derived from data, rather than specified a priori according to theory (Ding, 2006). The utility and validity of this approach have been demonstrated in several studies (for more detail regarding the

Table 3
 Mean Standardized Points and Standard Deviations ($n = 567$)

Variable	<i>M</i>	(<i>SD</i>)
Visual Perception (VP)	10.09	3.45
Selective Attention (SA)	10.09	3.48
Visual-Spatial Working Memory (VSWM)	10.14	3.33
Phonological Span (PS)	10.09	3.47
Verbal Long-Term Memory (VLTM)	9.97	3.37
Figural Reasoning (FR)	9.90	3.53
Conceptual Reasoning (CR)	10.07	3.32
Achievement Motivation I (Persistence) (AM I)	10.02	3.07
Achievement Motivation II (Intrinsic Motivation) (AM II)	9.92	2.96
Mathematical Skills (MS)	9.94	3.00
IQ	103.34	11.61

PAMS method see Davison et al., 2001; Ding, 2001; Ding, 2005; Ding, 2006; Frisby & Kim, 2008; Kim et al., 2004). The Multidimensional Scaling and bootstrapping were conducted using R 2.11.0, with the Scaling by Majorizing a Complicated Function (SMACOF) procedure (de Leeuw & Mair, 2008). For all the other analyses, PASW Statistics 18.0 was used.

RESULTS

Preliminary Analyses

Table 3 provides an overview of the means and standard deviations of the standardized subtest and IQ scores. Table 2 provides bivariate correlations among study variables. Correlation analysis revealed small to medium associations among all cognitive variables (VP, SA, VSWM, PS, VLTM, FR, and CR), the two achievement motivation variables (AM I and AM II), Mathematical Skills (MS), and IQ, ranging between $r = -.01$ (SA and FR, SA and CR) and $.56$ (MS and IQ). As can be seen, Achievement Motivation I and II correlated significantly with all cognitive subtests, providing further justification for taking cognitive and achievement motivation variables together when building the profiles.

Interindividual Differences in Cognitive and Motivational Variables as Predictors of Mathematical Skills: Regressions

The regression with Mathematical Skills as the dependent variable, and the seven cognitive and two motivation variables as predictors, yielded significant effects for Visual Perception ($\beta = .20$, $p < .001$), Visual-Spatial Working Memory ($\beta = .17$, $p < .001$), Phonological Span ($\beta = .21$, $p < .001$), Verbal Long-Term Memory ($\beta = .14$, $p < .001$), Figural Reasoning ($\beta = .23$, $p < .001$), and Conceptual Reasoning ($\beta = .11$, $p < .01$). Although the zero-order correlations were significant, in the regression Selective Attention ($\beta = .07$, $p = .06$), Achievement Motivation I ($\beta = -.04$, $p = .33$), and Achievement Motivation II ($\beta = .08$, $p = .07$) were not statistically significant predictors of Mathematical Skills. A total of 34% of variance in Mathematical Skills was explained by the subtest scores.

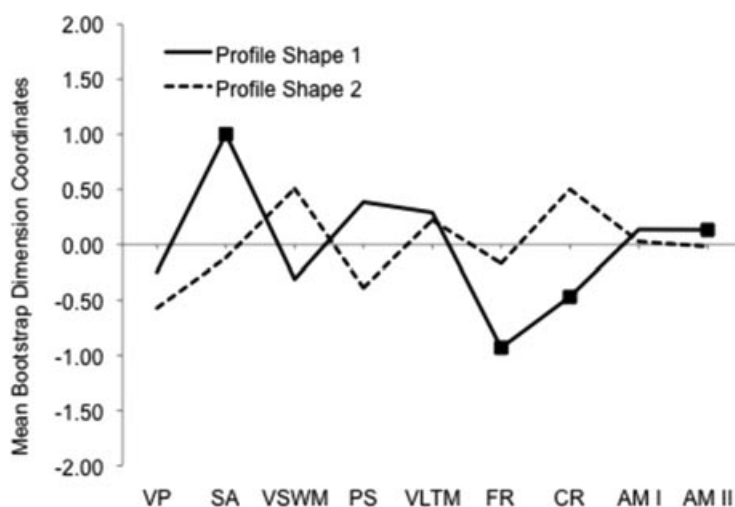


FIGURE 2. The two PAMS profile shapes. The continuous black line represents profile shape 1, the broken line represents profile shape 2. The squared markers represent the statistically significant dimension coordinates. VP = Visual Perception, SA = Selective Attention, VSWM = Visual-Spatial Working Memory, PS = Phonological Span, VLTM = Verbal Long-Term Memory, FR = Figural Reasoning, CR = Conceptual Reasoning, AM I = Achievement Motivation I (Persistence), AM II = Achievement Motivation II (Intrinsic Motivation).

Intraindividual Differences in Cognitive and Motivational Variables as Predictors of Mathematical Skills: Profile Analysis

In the first step of the PAMS procedure, the most common profile shapes are extracted. Therefore, Euclidian distances were computed from the seven cognitive and two achievement motivation subtests across all children in the dataset. Then, a nonmetric multidimensional scaling analysis (MDS) using the SMACOF procedure was performed on these distances. The goal was to obtain a two-dimensional solution, as only the most common profile shapes were to be extracted, facilitating interpretation ease (see Figure 2 for a graphical display of the two profile shapes). According to Kim et al. (2004), these two dimensions were interpreted as latent profile shapes. The fit index in the symmetric SMACOF model is a stress value, which was 0.01, a satisfactory value that is significantly different from a stress value that could be expected by chance (de Leeuw & Stoop, 1984). Scale values (further termed dimension coordinates to be consistent with Kim et al., 2004) were computed for all nine observed variables on the two extracted latent profile shapes, resulting in a $9 \text{ (variables)} \times 2 \text{ (profile shapes)}$ matrix.

In the second step of profile analysis, bootstrapping was used to determine the significance of these dimension coordinates with respect to the profile shape (see Ding, 2005; Kim, 2010). For this purpose, 1,000 bootstrap samples were drawn to generate the sampling distribution of the dimension coordinates. To make the bootstrap configuration more commensurable with the original MDS configuration, a Procrustes transformation was included in the bootstrap procedure (Peres-Neto & Jackson, 2001). By performing a Procrustes transformation, each bootstrap configuration is adjusted to the original MDS configuration. Mean bootstrap dimension coordinates and 95% confidence intervals around the simulated dimension coordinates are presented in Table 4. Profile shape 1, which represented the first dimension of the MDS solution, contained statistically significant high values on the subtests of Selective Attention (mean bootstrapped MDS dimension coordinate: 1.00) and Achievement Motivation II (mean bootstrapped MDS dimension coordinate: 0.13). As seen in Table 4, the lower bound 95% confidence interval of Achievement Motivation II lies only

Table 4

Mean MDS Dimension Coordinates Estimated From 1,000 Bootstrap Samples With 95% Bootstrap Confidence Interval

Variable	Profile Shape 1			Profile Shape 2		
	Dimension Coordinates	95% CI		Dimension Coordinates	95% CI	
		Lower	Upper		Lower	Upper
Visual Perception (VP)	-0.25	-0.60	0.29	-0.57	-0.96	0.34
Selective Attention (SA)	1.00*	0.45	1.45	-0.12	-0.72	0.50
Visual-Spatial Working Memory (VSWM)	-0.31	-0.70	0.01	0.51	-0.27	1.02
Phonological Span (PS)	0.39	-0.32	0.80	-0.39	-1.10	0.58
Verbal Long-Term Memory (VLTM)	0.29	-0.03	0.56	0.22	-0.21	0.70
Figural Reasoning (FR)	-0.93*	-1.31	-0.51	-0.17	-0.53	0.41
Conceptual Reasoning (CR)	-0.47*	-0.79	-0.11	0.50	-0.05	0.94
Achievement Motivation I (Persistence) (AM I)	0.14	-0.01	0.31	0.03	-0.12	0.21
Achievement Motivation II (Intrinsic Motivation) (AM II)	0.13*	0.01	0.29	-0.02	-0.14	0.13

* Statistically significant at .05 level.

slightly above zero, whereas the lower bound 95% confidence interval of Achievement Motivation I lies slightly below zero. This leads to a dimension coordinate significantly different from zero in case of Achievement Motivation II, and a statistically nonsignificant dimension coordinate in case of Achievement Motivation I, despite very similar 95% confidence intervals. Further, the first profile shape was characterized by statistically significant low values on the subtests of Figural Reasoning (mean bootstrapped MDS dimension coordinate: -0.93) and Conceptual Reasoning (mean bootstrapped MDS dimension coordinate: -0.47). In case of profile shape 2, no dimension coordinate was significantly different from zero.

The third step involved the estimation of person parameters, that is, person correspondence weights as a measure of correspondence to the two profile shapes using the PAMS syntax available in Ding (2006). Thereby, the original observed variable scores of each person function as dependent variables that are then regressed onto the variable dimension coordinates, which in this case form the independent variables (Kim et al., 2004), resulting in the estimation of individual person correspondence weights. These person correspondence weights are analogous to unstandardized regression coefficients and represent the degree to which an individual's observed scores correspond to the profile shape (Kim et al., 2004). In a separate step, profile level was computed for each individual, corresponding to the mean of all subtest scores. Profile level information was only used in this step of the calculation and was excluded from the profile shape analysis. In addition, an R^2 value was computed for every child as a measure of the proportion of variance in each child's observed profile shape explained by the two latent profile shapes (Frisby & Kim, 2008). The mean variance in the observed scores explained by the two profile shapes was $R^2 = .32$ ($SD = .23$). That is, there were a number of children whose observed scores were not well represented by the two profile shapes. Because the sample was selected on the basis of high intraindividual variability, it was supposed to be heterogeneous and hence comprise different types of profile shapes. Therefore, the two profile shapes extracted can only be representative for part of the whole sample, whereas the rest of the sample may exhibit different profile shapes less common in the whole sample. For these children, the depicted profile shapes cannot be predictive of anything because they don't correspond to these profile shapes. Therefore, R^2 was used to pick only those children who correspond to one of the

Table 5

Bivariate Pearson Correlations Between Profile Correspondence Weights, Profile Level, Sociodemographic Variables, IQ, and Mathematical Skills for the Reduced Sample (n = 133)

	Person Correspondence Weights			n
	Profile Shape 1	Profile Shape 2	Level	
Correspondence weights profile 1	–	0.19*	–.14	
Correspondence weights profile 2	.19*	–	.03	133
Sex	.30***	.05	–.07	133
Maternal education	–.17	–.13	.36***	102
Paternal education	–.17	–.20*	.28**	101
Age (in months)	.00	.01	.13	133
IQ	–.27**	–.02	.96***	133
Mathematical skills	–.37***	–.10	.58***	133

* $p < .05$. ** $p < .01$. *** $p < .001$.

two extracted profile shapes. For this purpose, children with an explained variance score (R^2) equal or below .50 were excluded from the subsequent analyses. The reduced sample consisted of 133 children (82 girls and 51 boys, mean age = 7.81 years, $SD = 1.60$ years, 2.3%, 1.5%, and 4.5% suffered from parent reported language impairment, hyperactivity, or attention deficit, and 5.3% were born preterm) with more than 50% variance in the observed scores explained by either profile shape 1 or 2. It is interesting to note that the profile level variable did not differ significantly when the reduced sample ($n = 133$, $M = 10.26$, $SD = 1.47$) was compared to the rest of the sample not selected ($n = 434$, $M = 9.96$, $SD = 1.63$), neither with respect to the mean ($p = .053$) nor the variance ($p = .138$).

In the fourth step of the profile analysis, these person correspondence weights (i.e., their similarity scores with respect to the two profile shapes) and profile levels were correlated with each other and with sex, maternal and paternal education, age, IQ, and Mathematical Skills in the profile shape-concordant sample. As seen in Table 5, correspondence weights on profile shape 1 were positively correlated with correspondence weights on profile shape 2 ($r = .19$, $p < .05$) and with sex ($r = .30$, $p < .001$), and negatively with IQ ($r = -.27$, $p < .01$) and Mathematical Skills ($r = -.37$, $p < .001$). In other words, children who showed a subtest pattern close to profile shape 1 were more likely to be girls than boys, and the higher the resemblance of individual scores to profile shape 1, the lower IQ and Mathematical Skills. The negative association between correspondence weights on profile shape 1 and IQ also held when the effect of sex was controlled ($r = -.24$, $p < .001$). Correspondence weights on profile shape 2 were significantly negatively related to paternal education ($r = -.20$, $p < .05$), that is, the higher the resemblance of individual scores to profile shape 2, the lower the parental education. Profile level was not significantly related to correspondence weights on the two profile shapes, which demonstrates that the PAMS model extracted profile level and profile shape independently. Profile level, however, was significantly positively correlated with IQ ($r = .96$, $p < .001$), which was expected, as the profiles were mainly based on cognitive variables. In addition, profile level was positively correlated with maternal education ($r = .36$, $p < .001$), paternal education ($r = .28$, $p < .01$), and Mathematical Skills ($r = .58$, $p < .001$).

Finally, a regression analysis was performed to predict Mathematical Skills. Profile level was entered in the first step, and correspondence weights on profile shape 1 and 2 were entered in the second step to determine incremental validity. The strongest predictor of Mathematical Skills was

profile level with a standardized regression coefficient of .58 ($p < .001$) in the first step and .54 ($p < .001$) in the second step. This predictor explained 34% of the variance in Mathematical Skills. When correspondence weights on profile shape 1 and 2 were entered in the second step, they were shown to explain another 8.8% ($p < .001$) of the variance in Mathematical Skills. Only correspondence weights on profile shape 1 emerged as a statistically significant predictor of Mathematical Skills ($\beta = -.28, p < .001$). This means that children who corresponded to profile shape 1 were more likely to score low on Mathematical Skills, irrespective of their overall profile level. Correspondence weights on profile shape 2 were not a statistically significant predictor of Mathematical Skills ($\beta = -.06, p = .39$).

Taken together, the two profile shapes explained an additional 8.8% in the outcome variable when profile level was accounted for. According to Cohen (1988), this increment is a small to medium effect ($R^2 = .03$ is a small effect, $R^2 = .10$ a medium effect, and $R^2 = .30$ a large effect). Multicollinearity was not a problem in any of the regression models as profile level and correspondence weights did not correlate at all (see Table 5), and the multicollinearity indices were all within the acceptable range.

DISCUSSION

The aim of the present study was to investigate the predictive validity of single subtest scores (i.e., interindividual differences) as well as profile shape (i.e., intraindividual differences) in the prediction of Mathematical Skills. All analyses were run with children showing high intraindividual variance in the subtest scores. Therefore, the present study also addressed the question of whether subtest interpretation is useful in samples with highly variable subtest scores (e.g., Fiorello et al., 2002). In addition, profile analysis was conducted using multidimensional scaling, a well-validated statistical approach that provides a promising alternative to other statistical approaches applied so far (Davison et al., 2001; Ding, 2001; Ding, 2005; Ding, 2006; Frisby & Kim, 2008; Kim, 2010; Kim et al., 2004). Further, achievement motivation was included in the profiles to predict educational outcomes not only by means of cognition, but also by motivational aspects.

Intraindividual Differences in Cognitive and Motivational Variables as Predictors of Mathematical Skills

In the first analysis, *interindividual differences* in single subtest scores were used to predict Mathematical Skills. It was expected that all seven cognitive subtests as well as the two achievement motivation subtests would predict Mathematical Skills. Although all zero-order correlations with Mathematical Skills were significant, results indicated differential predictive utility when the variables were included together in a regression, in that six cognitive subtests (Visual Perception, Visual-Spatial Working Memory, Phonological Span, Verbal Long-Term Memory, Figural Reasoning, and Conceptual Reasoning) were statistically significant predictors of Mathematical Skills. The strongest predictor of Mathematical Skills based on the size of the standardized regression coefficient was Figural Reasoning. Figural Reasoning involves spatial thinking, and it may be crucial for solving problems such as estimating quantity or amount in a two-dimensional picture of a three-dimensional structure. Visual-Spatial Working Memory was the fourth-strongest predictor, whereas Phonological Span represented the second-strongest predictor, which is surprising, because based on the results of Bull et al. (2008) it would have been expected that visual working memory would be a stronger predictor than phonological short-term memory. This finding also contradicts Floyd et al. (2003), who found that visual-spatial thinking was not a concurrent predictor of mathematics achievement. Selective Attention was not a statistically significant predictor of Mathematical Skills in this model.

Surprisingly, neither achievement motivation scale (persistence, intrinsic motivation) was predictive of Mathematical Skills when cognitive skills were in the same analysis. The present results indicate that children with high intrinsic motivation or more persistence do not necessarily have an advantage over children with lower achievement motivation and similar cognitive abilities. This is inconsistent with Duckworth and Seligman (2005), who found that self-discipline (persistence in the present study) was related to academic performance even when IQ was controlled. Results also contradict Spinath et al. (2006), who found that intrinsic motivation predicted achievement in elementary school children, even when accounting for self-perceptions and intelligence. However, Duckworth and Seligman (2005) studied adolescents, and academic performance was assessed using school grades, which makes their results difficult to compare to the present study. Inconsistencies in previous research could have arisen because of different assessment methodologies. Spinath et al. (2006) assessed intrinsic motivation via self-report, and Duckworth and Seligman (2005) used self-reports, parent reports, and teacher reports to assess self-discipline. The present study used reports by the examiner after the testing session. Our measure might therefore be more context-specific and less prone to social desirability, but may also be less generalizable to other situations. On the other hand, some consider motivation purely contextual and domain-specific rather than a stable trait (Linnenbrink & Pintrich, 2002). Therefore, further investigation of the present findings using different measures and in different contexts seems warranted.

It was expected that cognition and achievement motivation would share variance (Spinath et al., 2006). The evidence from the bivariate correlations points this direction: IQ correlated at $r = .33$ with persistence and at $r = .31$ with intrinsic motivation. It could therefore be that the cognitive variables in the regression analysis already tapped the variance that was shared with achievement motivation, leaving no variance left for achievement motivation itself to predict mathematical achievement.

Intraindividual Differences in Cognitive and Motivational Variables as Predictors of Mathematical Skills

Critics of the interpretation of intraindividual differences or profile shape analysis have pointed to the finding that profile shape or pattern does not add incremental validity to the prediction of educational outcomes, once profile level is accounted for (e.g., Kahana et al., 2002; Watkins & Glutting, 2000). Other studies showed that profile shape contributes to the prediction of educational outcomes, over and above the effect of profile level (Kim et al., 2004; Naglieri, 2000; Pritchard et al., 2000; Rohde & Thompson, 2007). In the present study, profile level was expected to be the strongest predictor of Mathematical Skills, but profile shape was hypothesized to add incremental validity, especially when meaningful subtest variation is present.

The first of the two dominant extracted profile shapes was characterized by intraindividual strengths in Selective Attention and Achievement Motivation II (intrinsic motivation) with concurrent intraindividual weaknesses in Figural Reasoning and Conceptual Reasoning. This profile shape seems to capture children with marked intraindividual deficits in two of seven cognitive subtests, who are at the same time highly intrinsically motivated to improve their achievement. Children this age may try to enhance achievement by making a greater effort (Folmer et al., 2008). It could be hypothesized that these children are aware of their weaknesses and might try to compensate by working extra hard or extra fast (Brunstein & Heckhausen, 2006). This latter assumption may explain the intraindividual strength in Selective Attention, as children who work fast, but not necessarily more accurate, tend to process more items and reach a higher score despite the higher amount of errors. As for all the other cognitive subtests, a faster but less accurate working style typically manifests in lower scores. An alternative explanation would be that during primary school, children are typically optimistic about their own competencies (e.g., Fredricks & Eccles, 2002; Jacobs, Lanza,

Osgood, Eccles, & Wigfield, 2002; Wigfield et al., 2008), and that this high belief leads to relatively high persistence and intrinsic motivation even when achievement is weak.

The first profile shape was more common in girls, as can be concluded from the statistically significant correlation of $r = .30$ between correspondence weight on profile shape 1 and sex. Moreover, this profile shape was associated with lower IQ, suggesting that the intraindividual weakness in two cognitive subtests might translate to an interindividual or normative weakness and affect overall IQ.

Furthermore, this profile shape accounted for an additional 8.8% variance in the prediction of Mathematical Skills over and above general profile level, which supports our hypothesis and is consistent with previous research (e.g., Kim et al., 2004; Naglieri, 2000; Pritchard et al., 2000). The more closely children matched this profile shape, the lower their performance in Mathematical Skills when profile level was controlled. The result that deficits in Figural Reasoning and Conceptual Reasoning are associated with lower Mathematical Skills is congruent with the findings reported by Rohde and Thompson (2007) with a sample of young adults who found that processing speed and spatial abilities added a statistically significant amount of variance to the prediction of mathematical achievement in the Scholastic Aptitude Test when general cognitive ability was controlled. In our study, profile level was the strongest predictor of Mathematical Skills with 34% variance explained, which was also expected.

The second profile shape did not demonstrate statistically significant strengths or weaknesses and was not related to Mathematical Skills. The lack of association with math skills may be related to the low reliability of the second profile shape, which results from this profile shape being the second MDS dimension and therefore by definition being less reliable when compared to the first dimension.

Limitations

The first limitation of the present study is the use of a purely exploratory method, namely PAMS. This fit the exploratory nature of the present study. Nevertheless, confirmatory techniques should be applied to the first profile shape found in this study (e.g., a Confirmatory Factor Analysis parameterization of the PAMS as recommended by Kim, Davison, & Frisby, 2007). It would also be informative to investigate the incremental predictive validity of profile shapes with longitudinal data (e.g., with a longitudinal PAMS as suggested by Ding, 2003).

Second, a two-dimensional solution was chosen for this study, resulting in exactly two latent profile shapes. The determination of number of profile shapes can be subjective (Ding, 2006). In the present study, the decision was based on interpretability and reliability of the estimation. The possibility that other meaningful profile shapes exist can thus not be excluded, especially not with respect to this heterogeneous group with high intraindividual variance. Moreover, the two profile shapes reported only reflect the most common profile shapes in the subsamples. Less frequently occurring profile shapes (that are perhaps more clinically relevant) were not considered.

Third, the present sample was chosen from the original sample based on participants' intraindividual variance, with subjects included only if their intraindividual variance was above the median. Such a median split could be viewed as arbitrary. Most studies investigating children with high variability selected children based on significant difference scores in cognitive subtests (e.g., Fiorello et al., 2002). However, our approach selected children based on their variability among all cognitive and motivational variables together, and not just differences between two variables. The sample was even further reduced in a next step, when only those children whose variance in the observed scores was well explained by one of the two profile shapes were selected. This led to a

seemingly artificial reduction of sample size, but it was also seen as reasonable to exclude those children who do not correspond to either one of the profile shapes and would therefore weaken the association between profile shapes and external criteria. Notably, this sample reduction did not lead to a reduction in variance in the profile level variable. Moreover, the sample was reduced according to profile shape similarity, which was computed independently from profile level. Therefore, the sample restriction did not imply a limitation of profile level as a possible predictor of Mathematical Skills.

Fourth, the outcome variable used in this study (i.e., Mathematical Skills) is not the only possibility. It could be that the profile shapes relate to other variables not studied here, such as reading, writing, grades, or behavior. It is reasonable to assume that the profile shapes of cognitive strengths and weaknesses affect not only mathematical skills, but broader learning skills, and would also affect other learning outcomes such as reading or writing. These possible relationships should be investigated in future studies.

Fifth, the sample used in this study was drawn from a German-speaking European population in three nations, Austria, Germany, and Switzerland. Therefore, the results do not necessarily generalize to other populations. The sample used in the current study was part of a standardization sample and was therefore nonclinical. It would be interesting to use PAMS to establish profile shapes in larger clinical samples (such as learning-disabled children), and thus address the differentiation of diagnostic groups.

Sixth, the assessment of achievement motivation as rated by the examiner is questionable. Despite satisfactory internal consistency and retest-reliability coefficients, the construct validity of this measure should be further investigated. Self-reported measures were not taken because of the many limitations of this methodology, especially with young children (Fulmer & Frijters, 2009).

Finally, it is unclear whether the profile shapes found in this study would be stable over time. For this purpose, a longitudinal study would be informative.

Implications for School Psychological and Educational Assessment Research

Taken together, the results support prior research findings that general ability or profile level is the most significant predictor of academic achievement (e.g., Watkins et al., 2007), even in a sample with high intraindividual variance. This does not come as a surprise, as a general score derived from several intercorrelated subtests can be more reliable than a combination of single subtest scores. However, profile shape information may not need to be ignored, at least in cases of substantial subtest variation. Advice for practical profile interpretation in case of subtest variability is to use a top-down approach; that is, to interpret general cognitive ability first and then move to the more specific combinations of subtest scores when formulating hypotheses about individual strengths and deficits (Kahana et al., 2002; Watkins et al., 2005). This approach does not necessarily imply that specific cognitive abilities have more predictive power than a single general cognitive factor, as Kahana et al. (2002) claim. On the contrary, general ability must still be seen as the most predictive factor of academic achievement. Nevertheless, the interpretation of individual profile shapes may lead to additional information that should not necessarily be neglected. It may be argued that the children whose skills resembled the profile shapes found in this study only represent a small fraction of the whole sample. However, children who show this specific pattern might be understood in more detail when all the information available is used. It can therefore be concluded that profile shape can be informative in children with high intraindividual variability, even over and above profile level. The challenging remaining question should be determining the minimum level of intraindividual variability needed to reliably interpret profile shape.

REFERENCES

- Alloway, T. P. (2009). Working memory, but not IQ, predicts subsequent learning in children with learning difficulties. *European Journal of Psychological Assessment, 25*, 92–98. doi:10.1027/1015-5759.25.2.92.
- Borsuk, E. R., Watkins, M. W., & Canivez, G. L. (2006). Long-term stability of membership in a Wechsler Intelligence Scale for Children-Third Edition (WISC-III) subtest core profile taxonomy. *Journal of Psychoeducational Assessment, 24*, 52–68. doi:10.1177/0734282905285225
- Brunstein, J., & Heckhausen, H. (2006). Leistungsmotivation [achievement motivation]. In J. Heckhausen & H. Heckhausen (Eds.), *Motivation und handeln [motivation and action]*. Heidelberg: Springer.
- Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology, 33*, 205–228. doi:10.1080/87565640801982312
- Cohen, J. (1988). *Statistical power for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Daniel, M. H. (2007). “Scatter” and the construct validity of FSIQ: Comment on Fiorello et al. (2007). *Applied Neuropsychology, 14*, 291–295. doi:10.1080/09084280701719401
- Davison, M. L., Kim, S. K., & Ding, S. (2001). Profile analysis via multidimensional scaling (PAMS): Exploring the predominant profile patterns in data. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.
- Davison, M. L., & Kuang, H. (2000). Profile patterns: Research and professional interpretation. *School Psychology Quarterly, 15*, 457–464. doi:10.1037/h0088801
- de Leeuw, J., & Mair, P. (2008). Multidimensional scaling using majorization: SMACOF in R. *UCLA Statistics Preprint Series, 537*, 1–29. Retrieved March 10, 2011, from <http://cran.r-project.org/web/packages/smactof/vignettes/smactof.pdf>
- de Leeuw, J., & Stoop, I. (1984). Upper bounds for kruskal’s stress. *Psychometrika, 49*, 391–402. doi:10.1007/BF02306028
- Ding, C. S. (2001). Profile analysis: Multidimensional scaling approach. *Practical Assessment, Research & Evaluation, 7*. Retrieved March 10, 2011, from <http://PAREonline.net/getvn.asp?v=7&n=16>.
- Ding, C. S. (2003). Exploratory longitudinal profile analysis via multidimensional scaling. *Practical Assessment, Research & Evaluation, 8*. Retrieved February 6, 2011, from <http://PAREonline.net/getvn.asp?v=8&n=12>
- Ding, C. S. (2005). Determining the significance of scale values from multidimensional scaling profile analysis using a resampling method. *Behavior Research Methods, 37*, 37–47. doi:10.3758/BF03206396
- Ding, C. S. (2006). Multidimensional scaling modeling approach to latent profile analysis in psychological research. *International Journal of Psychology, 41*, 226–238. doi:10.1080/00207590500412219
- Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science, 16*, 939–944. doi:10.1111/j.1467-9280.2005.01641.x
- Elliott, C. D., Hale, J. B., Fiorello, C. A., Dorvil, C., & Moldovan, J. (2010). Differential Ability Scales-II prediction of reading performance: Global scores are not enough. *Psychology in the Schools, 47*, 698–720. doi:10.1002/pits.20499
- Fiorello, C. A., Hale, J. B., Holdnack, J. A., Kavanagh, J. A., Terrell, J., & Long, L. (2007). Interpreting intelligence test results for children with disabilities: Is global intelligence relevant? *Applied Neuropsychology, 14*, 2–12. doi:10.1080/09084280701280338
- Fiorello, C. A., Hale, J. B., McGrath, M., Ryan, K., & Quinn, S. (2002). IQ interpretation for children with flat and variable test profiles. *Learning and Individual Differences, 13*, 115–125. doi:10.1016/S1041-6080(02)00075-4
- Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school-age years. *Psychology in the Schools, 40*, 155–171. doi:10.1002/pits.10083
- Folmer, A. S., Cole, D. A., Sigal, A. B., Benbow, L. D., Satterwhite, L. F., Swygert, K. E., et al. (2008). Age-related changes in children’s understanding effort and ability: Implications for attribution theory and motivation. *Journal of Experimental Child Psychology, 99*, 114–134. doi:10.1016/j.jecp.2007.09.003
- Freberg, M. E., Vandiver, B. J., Watkins, M. W., & Canivez, G. L. (2008). Significant factor score variability and the validity of the WISC-III Full Scale IQ in predicting later academic achievement. *Applied Neuropsychology, 15*, 131–139. doi:10.1080/09084280802084010
- Fredricks, J. A., & Eccles, J. S. (2002). Children’s competence and value beliefs from childhood through adolescence: Growth trajectories in two male-sex-typed domains. *Developmental Psychology, 38*, 519–533. doi:10.1037/0012-1649.38.4.519
- Frisby, C. L., & Kim, S. K. (2008). Using profile analysis via multidimensional scaling (PAMS) to identify core profiles from the WMS-III. *Psychological Assessment, 20*, 1–9. doi:10.1037/1040-3590.20.1.1
- Fulmer, S. M., & Frijters, J. C. (2009). A review of self-report and alternative approaches in the measurement of student motivation. *Educational Psychology Review, 21*, 219–246. doi:10.1007/s10648-009-9107-x
- Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *The Journal of Special Education, 40*, 103–114. doi:10.1177/00224669060400020101

- Gottfried, A. E. (1990). Academic intrinsic motivation in young elementary school children. *Journal of Educational Psychology*, 82, 525–538. doi:10.1037/0022-0663.82.3.525
- Grob, A., Meyer, C. S., & Hagmann-von Arx, P. (2009). *Intelligence and Development Scales (IDS)*. Bern: Huber.
- Hagmann-von Arx, P., Meyer, C. S., & Grob, A. (2008). Assessing intellectual giftedness with the WISC-IV and the IDS. *Zeitschrift fuer Psychologie [Journal of Psychology]*, 216, 172–179. doi:10.1027/0044-3409.216.3.172
- Hale, J. B., Fiorello, C. A., Dumont, R., Willis, J. O., Rackley, C., & Elliott, C. (2008). Differential Ability Scales – Second Edition (neuro)psychological predictors of math performance for typical children and children with math disabilities. *Psychology in the Schools*, 45, 838–858. doi:10.1002/pits.20330
- Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Hoepfner, J. A. B., & Gaither, R. A. (2001). WISC-III predictors of academic achievement for children with learning disabilities: Are global and factor scores comparable? *School Psychology Quarterly*, 16, 31–55. doi:10.1521/scpq.16.1.31.19158
- Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Holdnack, J. A., & Aloe, A. M. (2007). Is the demise of IQ interpretation justified? A response to special issue authors. *Applied Neuropsychology*, 14, 37–51. doi:10.1080/09084280701280445
- Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment*, 15, 456–466. doi:10.1037/1040-3590.15.4.456
- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grade one through twelve. *Child Development*, 73, 509–527. doi:10.1111/1467-8624.00421
- Kahana, S. Y., Youngstrom, E. A., & Glutting, J. J. (2002). Factor and subtest discrepancies on the Differential Ability Scales – Examining prevalence and validity in predicting academic achievement. *Assessment*, 9, 82–93. doi:10.1177/1073191102009001010
- Kim, S. K. (2010). Evaluating the invariance of cognitive profile patterns derived from profile analysis via multidimensional scaling (PAMS): A bootstrapping approach. *International Journal of Testing*, 10, 33–46. doi:10.1080/15305050903454549
- Kim, S. K., Davison, M. L., & Frisby, C. L. (2007). Confirmatory factor analysis and profile analysis via multidimensional scaling. *Multivariate Behavioral Research*, 42, 1–32. doi:10.1080/00273170701328973
- Kim, S. K., Frisby, C. L., & Davison, M. L. (2004). Estimating cognitive profiles using profile analysis via multidimensional scaling (PAMS). *Multivariate Behavioral Research*, 39, 595–624. doi:10.1207/s15327906mbr3904_2
- Kotz, K. M., Watkins, M. W., & McDermott, P. A. (2008). Validity of the general conceptual ability score from the differential ability scales as a function of significant and rare interfactor variability. *School Psychology Review*, 37, 261–278. Retrieved February 6, 2011, from [http://edpsychassociates.com/Papers/GCAvalidity\(2008\).pdf](http://edpsychassociates.com/Papers/GCAvalidity(2008).pdf)
- Linnenbrink, E. A., & Pintrich, P. R. (2002). Motivation as an enabler for academic success. *School Psychology Review*, 31, 313–327. Retrieved March 10, 2011, from <http://psycnet.apa.org/psycinfo/2002-18945-003>
- Lobeck, A., & Frei, M. (1987). *Schweizer rechenstest 1.–3. Klasse [Swiss test of arithmetics grades 1 to 3]*. Basel: Beltz.
- Logan, S., Medford, E., & Hughes, N. (2011). The importance of intrinsic motivation for high and low ability readers' reading comprehension performance. *Learning and Individual Differences*, 21, 124–128. doi:10.1016/j.lindif.2010.09.011
- Lohman, D. F., Gambrell, J., & Lakin, J. (2008). The commonality of extreme discrepancies in the ability profiles of academically gifted students. *Psychology Science Quarterly*, 50, 269–282. Retrieved March 10, 2011, from http://faculty.education.uiowa.edu/dlohman/pdf/Lohman_Gambrell_Lakin.pdf
- Naglieri, J. A. (2000). Can profile analysis of ability test scores work? An illustration using the PASS theory and CAS with an unselected cohort. *School Psychology Quarterly*, 15, 419–433. doi:10.1037/h0088798
- Peres-Neto, P. R., & Jackson, D. A. (2001). How well do multivariate data sets match? The advantages of a Procrustean superimposition approach over the Mantel test. *Oecologia*, 129, 169–178. doi:10.1007/s004420100720
- Petermann, F., & Petermann, U. (2007). *Hamburg Wechsler Intelligenztest fuer Kinder-IV (HAWIK-IV)*. Bern: Huber.
- Pfeiffer, S. I., Reddy, L. A., Kletzel, J. E., Schmelzer, E. R., & Boyer, L. M. (2000). The practitioner's view of IQ testing and profile analysis. *School Psychology Quarterly*, 15, 376–385. doi:10.1037/h0088795
- Pintrich, P. R., & Schunk, D. (2002). *Motivation in education: Theory, research, and applications* (2nd ed.). Upper Saddle, NJ: Prentice-Hall, Inc.
- Pritchard, D. A., Livingston, R. B., Reynolds, C. R., & Moses, J. A., Jr. (2000). Modal profiles for the WISC-III. *School Psychology Quarterly*, 15, 400–418. doi:10.1037/h0088797
- Rizza, M. G., McIntosh, D. E., & McCunn, A. (2001). Profile analysis of the Woodcock-Johnson III Tests of Cognitive Abilities with gifted students. *Psychology in the Schools*, 38, 447–455. doi:10.1002/pits.1033
- Robinson, B. R., & Harrison, P. L. (2005). WISC-III core profiles for students referred or found eligible for special education and gifted programs. *School Psychology Quarterly*, 20, 51–65. doi:10.1521/scpq.20.1.51.64194
- Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence*, 35, 83–92. doi:10.1016/j.intell.2006.05.004

- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25, 54–67. doi: 10.1006/ceps.1999.1020
- Sattler, J. M., & Saklofske, D. H. (2001). Interpreting the WISC-III. In J. M. Sattler (Ed.), *Assessment of children – cognitive applications* (4th ed.). (pp. 298–334). San Diego: Jerome M. Sattler, Publisher, Inc.
- Schuler, H., & Prochaska, M. (2001). *Leistungsmotivationsinventar (LMI)*. Göttingen: Hogrefe.
- Smith, C. B., & Watkins, M. W. (2004). Diagnostic utility of the Bannatyne WISC-III pattern. *Learning Disabilities Research & Practice*, 19, 49–56. doi:10.1111/j.1540-5826.2004.00089.x
- Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence*, 34, 363–374. doi:10.1016/j.intell.2005.11.004
- Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly*, 15, 465–479. doi:10.1037/h0088802
- Watkins, M. W., & Glutting, J. J. (2000). Incremental validity of WISC-III profile elevation, scatter, and shape information for predicting reading and math achievement. *Psychological Assessment*, 12, 402–408. doi: 10.1037//1040-3590.12.4.402
- Watkins, M. W., Glutting, J. J., & Lei, P. W. (2007). Validity of the full-scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology*, 14, 13–20. doi:10.1080/09084280701280353
- Watkins, M. W., Glutting, J. J., & Youngstrom, E. A. (2005). Issues in subtest profile analysis. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 251–268). New York: The Guilford Press.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children – Third Edition*. San Antonio, TX: The Psychological Corporation.
- Wigfield, A., Eccles, J. S., Roeser, R. W., & Schiefele, U. (2008). Development of achievement motivation. In W. Damon & R. M. Lerner (Eds.), *Child and adolescent development: An advanced course*. Hoboken, NJ: Wiley.

Publikationen

Frischknecht, M.-C., Reimann, G., Gut, J., Ledermann, T. & Grob, A., (2014). Wie genau können Mütter die Mathematik- und Sprachleistungen ihrer Kinder einschätzen? Ein Vergleich zwischen Mütterereinschätzung und Testleistungen bei Kindern im Alter von 6-10 Jahren. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 46 (2), 67-78.

Frischknecht, M.-C., Reimann, G. & Grob, A. (submitted). Erkennen Eltern Entwicklungsdefizite im Vorschulalter? Zur Akkuratheit elterlicher Einschätzungen kindlicher Entwicklung. Eingereicht bei *Kindheit und Entwicklung* Version vom 08. April 2014

Grob, A., Reimann, G., Gut, J. & Frischknecht, M.-C. (2013). *Intelligence and Development Scales – Preschool (IDS-P)*. Bern: Huber

Reimann, G., Stoecklin, M., Lavalley, K., Gut, J. Frischknecht, M.-C. & Grob A. (2013). Cognitive and motivational profile shape predicts mathematical skills over and above profile level. *Psychology in the Schools*, 50(1), 37-56. doi: 10.1002/pits.21659

Reimann, G., Gut, J., Frischknecht, M.-C., & Grob, A. (2013). Memory abilities in children with mathematical difficulties: Comorbid language difficulties matter. *Learning and Individual Differences*, 23, 108-113.

Selbständigkeitserklärung für Dissertation

Hiermit erkläre ich, dass ich die Dissertation „Einschätzung kindlicher Entwicklung: Zur Aussagekraft von informellen und formellen Diagnoseleistungen“ selbstständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wortwörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Die zur Promotion eingereichten Zeitschriftenbeiträge wurden in Zusammenarbeit mit den jeweiligen Co-Autorinnen und Co-Autoren angefertigt.

Rapperswil, im Mai 2014

Marie-Claire Frischknecht-Brunner