

Autorinnen:
Eva Stumpf & Susanne Trotter

Titel:
Validierung von Auswahlverfahren für gymnasiale Begabtenklassen

Pre-Print Version

Diese Artikelfassung entspricht nicht vollständig dem in der Zeitschrift veröffentlichten Artikel. Dies ist nicht die Originalversion des Artikels und kann daher nicht zur Zitierung herangezogen werden

Originalartikel 2015 erschienen in Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, Hogrefe Verlag, DOI 10.1026/0049-8637/a000119

Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie
Validierung von Auswahlverfahren für gymnasiale Begabtenklassen Validation of the
selection procedure for special gifted classes
 --Manuskript-Entwurf--

Manuskriptnummer:	ZEPP-D-13-00020R3
Vollständiger Titel:	Validierung von Auswahlverfahren für gymnasiale Begabtenklassen Validation of the selection procedure for special gifted classes
Artikeltyp:	Originalarbeit
Schlüsselwörter:	Begabtenförderung, Hochbegabung, Auswahlverfahren, Leistungsprognose selection, fostering, giftedness, prediction of school achievement
Korrespond. Autor:	Eva Stumpf, PD Dr. Psychologie IV Würzburg, GERMANY
Korrespondierender Autor, Zweitinformationen:	
Korrespondierender Autor, Institution:	Psychologie IV
Korrespondierender Autor, zweite Institution:	
Erstautor:	Eva Stumpf, PD Dr.
Erstautor, Zweitinformationen:	
Reihenfolge der Autoren:	Eva Stumpf, PD Dr. Susanne Trottler
Reihenfolge 'Zweite Informationen' von Autoren:	
Zusammenfassung:	<p>Für die Auswahl von Schülerinnen und Schülern für Begabtenfördermaßnahmen werden gemeinhin mehrdimensionale Verfahren gefordert und durchgeführt. Validierungsstudien, die diese Entscheidungen wissenschaftlich untermauern, liegen bislang nur wenige vor. In dieser Studie wurden die zentralen Elemente (Intelligenz, Zeugnisnoten, Verhalten im Probeunterricht) der Auswahlverfahren für gymnasiale Begabtenklassen hinsichtlich ihrer Eignung für die Leistungsprognose regressionsanalytisch überprüft. Die Ergebnisse bestätigen die Validität der Intelligenzwerte und Zeugnisnoten der vierten Jahrgangsstufe für die Prognose der Schulleistungen in der siebten Jahrgangsstufe. Die Daten des Probeunterrichts eignen sich darüber hinaus weder zur Prognose später erworbener Zensuren noch zur Prognose standardisiert erhobener Leistungskriterien. Es werden praktische Konsequenzen für die Gestaltung der Auswahlverfahren gezogen.</p> <p>Consensually, identifying children for gifted educational programs requires the use of multiple measures, but empirical results that validate these selection decisions are still rare. In this study, the selection processes for special classes for gifted children are assessed. The results confirm the prognostic validity of intelligence test scores and school marks measured in Grade 4 for school achievement in Grade 7. Ratings of the children's behavior during a trial lesson did neither improve the prediction of school marks, nor the prediction of standardized achievement measurements. Implications for the selection processes are discussed.</p>

Titel:

Validierung von Auswahlverfahren für gymnasiale Begabtenklassen

Validation of the selection procedure for special gifted classes

Abstract

Für die Auswahl von Schülerinnen und Schülern für Begabtenfördermaßnahmen werden gemeinhin mehrdimensionale Verfahren gefordert und durchgeführt. Validierungsstudien, die diese Entscheidungen wissenschaftlich untermauern, liegen bislang nur wenige vor. In dieser Studie wurden die zentralen Elemente (Intelligenz, Zeugnisnoten, Verhalten im Probeunterricht) der Auswahlverfahren für gymnasiale Begabtenklassen hinsichtlich ihrer Eignung für die Leistungsprognose regressionsanalytisch überprüft. Die Ergebnisse bestätigen die Validität der Intelligenzwerte und Zeugnisnoten der vierten Jahrgangsstufe für die Prognose der Schulleistungen in der siebten Jahrgangsstufe. Die Daten des Probeunterrichts eignen sich darüber hinaus weder zur Prognose später erworbener Zensuren noch zur Prognose standardisiert erhobener Leistungskriterien. Es werden praktische Konsequenzen für die Gestaltung der Auswahlverfahren gezogen.

Schlagworte: Begabtenförderung, Hochbegabung, Auswahlverfahren, Leistungsprognose

Consensually, identifying children for gifted educational programs requires the use of multiple measures, but empirical results that validate these selection decisions are still rare. In this study, the selection processes for special classes for gifted children are assessed. The results confirm the prognostic validity of intelligence test scores and school marks measured in Grade 4 for school achievement in Grade 7. Ratings of the children's behavior during a trial lesson did neither improve the prediction of school marks, nor the prediction of standardized achievement measurements. Implications for the selection processes are discussed.

Key words: selection, fostering, giftedness, prediction of school achievement

Im Jahr 2002 wurde durch das Bundesministerium für Bildung und Forschung (BMBF) eine Bestandsaufnahme zur schulischen Begabtenförderung in Europa in Auftrag gegeben. Die Autoren (Mönks, Peters & Pflüger, 2003) resümierten daraus für Deutschland unter anderem den Bedarf einer Modifizierung der Auswahlkriterien für die Zuweisung von Schülerinnen und Schülern an Begabtenfördermaßnahmen, um eine höhere Passung von Fähigkeits- und Anforderungsprofil zu erreichen und die Objektivität der Entscheidungen zu erhöhen. Wie die Auswertungen der Erhebung für Deutschland zeigten, wurden hier multimethodale und multimodale Auswahlverfahren für die Begabtenförderung bevorzugt. Das Einbeziehen standardisierter Fähigkeitstests wurde nicht generell, sondern speziell für die Überprüfung konkreter Fragestellungen oder für die Auswahl für Sommercamps gefordert (Mönks et al., 2003).

Während Begabtenfördermaßnahmen seitdem bundesweit enorm ausgeweitet wurden und es in diesem Zuge auch zu einem breiteren Angebot an Fähigkeitsgruppierungen kam, wurden die Auswahlkriterien auch in den vergangenen 15 Jahren nur selten kritisch hinterfragt. Deutlich wird dies an einer aktuelleren Studie, wonach auch die Auswahlverfahren für die Fördermaßnahmen für Studierende und Promovenden der elf Begabtenförderwerke des BMBF inhaltlich von jedem Standort selbst definiert und ausgestaltet werden (Middendorff, Isserstedt & Kandulla, 2009). Ein Mangel an klar definierten und einheitlichen Auswahlkriterien wurde auch für die Talentsuchen für die Summer School der National Academy for the Gifted and Talented Youth in England aufgezeigt (Hartas, Lindsay & Muijs, 2008). Diese Unschärfe und Heterogenität der Auswahlkriterien wird jedoch meist nicht als problematisch angesehen, sondern gar als Garant für eine adäquate Vielfalt der Begabtenförderung interpretiert (Middendorff et al., 2009).

Da jedoch davon auszugehen ist, dass der Erfolg von Begabtenfördermaßnahmen entscheidend von der Eignung der Auswahlstrategie beeinflusst wird (Stumpf, 2012a), sollte diesem Aspekt der Begabtenförderung deutlich mehr Aufmerksamkeit zuteilwerden. Die vorliegende Studie soll dazu beitragen, durch eine Validierung praktischer Auswahlverfahren für gymnasiale Begabtenklassen diese Form der Begabtenförderung weiter empirisch zu fundieren. Solche Begabtenklassen, die in der Regel einzügig ab der Jahrgangsstufe 5 an regulären Gymnasien angeboten werden, sind unseren aktuellen

Recherchen¹ zufolge derzeit in fünf Bundesländern (Baden-Württemberg, Bayern, Berlin, Brandenburg, Rheinland-Pfalz) etabliert. In einigen Bundesländern (z. B. Baden-Württemberg, Sachsen) werden zudem spezielle Gymnasien mit angegliederten Internaten für überdurchschnittlich begabte Schülerinnen und Schüler angeboten, die meist ab der siebten Jahrgangsstufe beginnen. In den Begabtenklassen gelten üblicherweise die regulären Curricula, die akzeleriert durchgeführt werden, sodass die frei werdende Zeit für Enrichmentmaßnahmen genutzt werden kann (vgl. Stumpf, 2012b). Die Aufnahme in die gymnasialen Begabtenklassen bzw. -schulen erfolgt nach erfolgreichem Durchlaufen der Auswahlverfahren, für das an einigen Standorten die Schulen, andernorts psychologische Beratungsstellen verantwortlich sind. Aufgrund der hohen Intensität der Begabtenklassen sowie der Tatsache, dass aus Kapazitätsgründen meist nur eines von drei vorgeschlagenen Kindern aufgenommen werden kann (vgl. Stumpf & Schneider, 2009), scheint der Einsatz möglichst treffsicherer Auswahlstrategien hier in besonderem Maße geboten. Der mehrdimensionale Auswahlprozess umfasst in der Regel Ergebnisse aus standardisierten diagnostischen Verfahren (z. B. Intelligenz), Zeugnisnoten und subjektiven Eindrücken aus Aufnahmegesprächen und/oder einem Probeunterricht. Wie genau die konkreten Entscheidungen für oder gegen die Aufnahme eines bestimmten Kindes getroffen werden, bleibt jedoch häufig intransparent. Nachfolgend werden die verfügbaren wissenschaftlichen Empfehlungen zur Gestaltung von Auswahlverfahren für Begabtenfördermaßnahmen dargestellt, bevor der aktuelle Stand der Forschung zur Validität solcher Aufnahmeverfahren resümiert wird.

Einige Autoren weisen auf die Notwendigkeit hin, die konkreten Auswahlkriterien aus den zuvor definierten Zielsetzungen und der Zielgruppe des Förderprogramms herzuleiten (Stumpf, 2012a; Vock, Preckel & Holling, 2007). Für die konkrete Umsetzung wird konsensuell das Einbeziehen möglichst vielfältiger Informationen (z. B. Intelligenz, Schulleistung), Messverfahren (z. B. Testverfahren, Checklisten) und Quellen (Schüler, Lehrkräfte, Eltern) in einem sequenziellen Vorgehen gefordert (z. B. Feldhusen & Jarwan, 2002; Heller & Perleth, 2007; Renzulli, 1993; Rost, 1991). Während Heller (2001) eine Vorauswahl potenziell geeigneter Schülerinnen und Schüler anhand von Checklisten oder Ratings durch die Lehrkräfte und erst im zweiten Schritt die Durchführung einer standardisierten Testdiagnostik an der vorselegierten Gruppe vorschlägt, plädieren andere

¹ Quelle: <http://www.fachportal-hochbegabung.de/bundeslaender/>; Zugriff am 13.06.2014; noch nicht enthalten sind darin die Angaben zu den Ländern Mecklenburg-Vorpommern und Thüringen.

Autoren dafür, die standardisierte Testdiagnostik den weniger objektiven Messinstrumenten voranzustellen, um subjektive Verzerrungen gering zu halten (Feldhusen & Jarwan, 2002; Rost, 1991; Stumpf, 2012a). Wie Lehmann und Jüling (1999) herausstellen, sollten im Auswahlprozess insbesondere solche Kriterien berücksichtigt werden, die die Erfolgskriterien des Förderprogramms möglichst valide prognostizieren.

Stand der Forschung

Bislang liegen noch wenige empirische Studien zur Validität von Auswahlverfahren für gymnasiale Begabtenklassen vor. In der Untersuchung von Lehmann und Jüling (1999; s. auch Jüling & Lehmann, 1996) gelang die Prognose späterer Zensuren von Schülerinnen und Schülern in mathematisch-naturwissenschaftlich-technischen Spezialklassen am besten anhand der Kombination aus Intelligenztestwert², Zeugnisnoten und dem Ergebnis in einer von der Schule erstellten Mathematiklausur Ende der vierten Jahrgangsstufe. Durch die Kombination dieser Prädiktoren konnte eine Varianzaufklärung von 37 % für den Gesamtnotendurchschnitt der sechsten Jahrgangsstufe erzielt werden. Der Vorhersagewert der einzelnen Prädiktoren für dieses Kriterium betrug hingegen nur $r^2 = .15$ für die Mathematiklausur, $r^2 = .10$ für den Intelligenztestwert und $r^2 = .08$ für die Zeugnisnoten (Lehmann & Jüling, 1999). Auch für die Prognose des Notendurchschnitts der achten Jahrgangsstufe wurde durch die Aggregation der Prädiktoren der höchste Validitätskennwert von $R^2 = .24$ erzielt (Jüling & Lehmann, 1999).

Hany (2002) untersuchte das Auswahlverfahren für mathematisch-naturwissenschaftliche Spezialklassen in Thüringen, die ab der neunten Jahrgangsstufe beginnen. Die Ergebnisse zeigten insgesamt nur vergleichsweise geringe, nicht signifikante Korrelationen³ ($r = -.01$ bis $r = -.28$) zwischen den jeweils zwei Intelligenztestaufgaben des Kognitiven Fähigkeitstests (Heller, Gaedike & Weinländer, 1985) zum quantitativen bzw. nonverbalen Denken und verschiedenen Leistungskriterien am Ende der neunten Jahrgangsstufe (Zensuren in mathematisch-naturwissenschaftlichen Fächern, Zensuren in sprachlichen Fächern, Leistungspunkte in einer regionalen Mathematikolympiade) (Hany, 2002). Deutlich höhere Zusammenhänge zu den mathematisch-naturwissenschaftlichen Zensuren der neunten Jahrgangsstufe zeigten sich hingegen für die Anstrengungsbereitschaft ($r = -.51$), für naturwissenschaftliche Aktivitäten ($r = -.47$), das Fähigkeitsselfkonzept ($r =$

² Erhoben mittels Leistungsprüfsystem von Horn (1983).

³ In diesem Beitrag werden die Korrelationskoeffizienten für Zensuren und weitere Variablen einheitlich nach nicht-invertierter Notenskala dargestellt, unabhängig davon, wie diese Darstellung in der Originalquelle gerichtet war.

-.44), die mathematisch-naturwissenschaftlichen Interessen ($r = -.40$) sowie für den Leistungsstand in Physik ($r = -.34$) zum Zeitpunkt der Bewerbung. Mit den sprachlichen Zensuren wie auch den Olympiadepunkten Ende der neunten Jahrgangsstufe korrelierten von den zwölf Einzelvariablen der Auswahlverfahren jeweils nur die Anstrengungsbereitschaft ($r = -.48$ bzw. $r = -.41$) sowie der Leistungsstand im Fach Mathematik ($r = -.35$ bzw. $r = -.34$) signifikant. Ergo hat sich in dieser Studie die Anstrengungsbereitschaft als wichtigstes Korrelat der später erhobenen Leistungskriterien erwiesen.

Auch für Spezialklassen ab der fünften Jahrgangsstufe mit Förderschwerpunkt Sprachen zeigten sich nur geringe Korrelationen zwischen den im Auswahlverfahren am Ende der Grundschulzeit erhobenen kognitiven Fähigkeiten und den Schulleistungen zu Beginn der sechsten Jahrgangsstufe (von $r = .02$ bis $r = -.36$) (Hany, 2004). Hier waren die späteren Zensuren in Deutsch, Englisch und Mathematik besonders eng (von $r = -.35$ bis $r = -.63$) mit komplexen muttersprachlichen Fertigkeiten (Fantasieaufsatz, Sachtextanalyse) und etwas geringer (von $r = -.05$ bis $r = -.56$) mit einfachen sprachlichen Fertigkeiten (z. B. Rechtschreibung, Grammatik) assoziiert. Wie anhand der Ergebnisse dieser Studie (Hany, 2004) deutlich wird, wurden die späteren Zensuren der Schülerinnen und Schüler von den verwendeten Maßen der Auswahlverfahren in vergleichbarer Höhe prognostiziert wie ergänzend erhobene subjektive Lehrkraftratings der fachspezifischen und allgemeinen Leistungsfähigkeit der Schülerinnen und Schüler. Die in beiden Studien (Hany, 2002, 2004) überraschend gering ausgeprägte Assoziation zwischen den intellektuellen Fähigkeiten der Schülerinnen und Schüler und den Leistungskriterien erklärte Hany (2004) anhand der Schwellenwerttheorie, wonach Leistungsunterschiede in den zugrunde liegenden Stichproben im oberen Fähigkeitsbereich nicht weiter anhand der kognitiven Fähigkeitsunterschiede erklärt werden können (vgl. auch Schneider, 2008).

Für eine adäquate Diskussion der Relevanz einer Intelligenztestung für die Aufnahme in Begabtenklassen sind zudem die Befunde zu den akzelerierten Gymnasialklassen in Rheinland-Pfalz anzuführen, deren Schülerinnen und Schüler gemeinsam eine Jahrgangsstufe der Mittelstufe überspringen. Um die Aufnahme entsprechend „dem komplexen Begabungsbegriff“ (Kaiser, 1997, S. 112) zu gestalten, wurden insbesondere das Arbeitsverhalten und die Motivation der Schülerinnen und Schüler für diese Entscheidungen berücksichtigt, wohingegen keine Intelligenztestung vorgenommen wurde. Wie die spätere Begleitstudie verdeutlichte, übertrafen die so gebildeten akzelerierten Klassen reguläre Gymnasialklassen nicht signifikant hinsichtlich der durchschnittlichen Intelligenzausprägung

(Kaiser, 1997). In Anbetracht der Tatsache, dass die Begriffe Begabung und Intelligenz in der psychologischen Fachliteratur weitgehend synonym verwendet werden (s. Heller, 2000; Rost, 2000), sollten Begabtenklassen allerdings auch eine durchschnittlich höhere Intelligenzausprägung ihrer Schülerschaft erwarten lassen. Insofern kann dieser Befund als Fehlleistung der diagnostischen Strategie interpretiert werden. Da insgesamt noch wenige Studien zur Validierung von Auswahlverfahren für gymnasiale Begabtenklassen vorliegen, sollen nachfolgend auch Validierungsstudien zu den Selektionskriterien für andere Begabtenfördermaßnahmen kurz dargestellt werden.

Perleth und Collier (2009) überprüften die prädiktive Validität der Auswahlverfahren für die sogenannten Hector⁴-Seminare in Baden-Württemberg. Dabei handelt es sich um Förderkurse aus den MINT-Fächern (Mathematik, Informatik, Naturwissenschaften, Technik) für Schülerinnen und Schüler der Sekundarstufe. Die Auswahl wird in einer ersten Stufe über die Nominierung der 10 % besten Schülerinnen und Schüler durch ihre Lehrkräfte operationalisiert, aus denen in der zweiten Stufe wiederum die besten 10 % mittels Intelligenzdiagnostik aufgenommen werden. Hier erwiesen sich die Intelligenzwerte mit $r = -.30$ (Median) als moderat valide zur Vorhersage des acht Jahre später erzielten Notendurchschnitts der Fächer Mathematik, Physik, Chemie, Biologie, Deutsch sowie der ersten Fremdsprache (Perleth & Collier, 2009).

Auch in einer Untersuchung zur Validierung der Auswahlverfahren für die Teilnahme am Frühstudium bestätigte sich der Einfluss intellektueller Fähigkeiten auf den langfristigen Erfolg (Stumpf & Schneider, 2013; Stumpf, 2011). Ebenso kam es hier durch die Prädiktorenkombination von Intelligenzmaßen und Zensuren zur Erhöhung des Anteils aufgeklärter Varianz, der allerdings mit maximal 10 % insgesamt sehr gering ausfiel (Stumpf & Schneider, 2013). Dies könnte auf mäßig geeignete Erfolgskriterien zurückzuführen sein, da keine Leistungspunkte bzw. Zensuren herangezogen werden konnten, sondern die Teilnahmedauer als Erfolgskriterium fungierte. Die Gültigkeit der Schwellenwerttheorie der Intelligenz hat sich in diesem Zusammenhang zur Prognose der Teilnahmedauer nicht bestätigen lassen (Stumpf, 2011). Letztgenannter Befund korrespondiert mit Ergebnissen der „Study of Mathematically Precocious Youth“ (SMPY), wonach Leistungsunterschiede im Erwachsenenalter innerhalb der Gruppe mathematisch Hochbegabter ebenfalls noch

⁴ Diese Förderkurse wurden nach dem Gründerehepaar Josefine und Dr. h.c. Hans-Werner Hector der Stiftung benannt, die die Seminare initiiert hat und die Ressourcen zur Durchführung zur Verfügung stellt.

teilweise auf Fähigkeitsunterschiede zurückzuführen waren (Ferriman-Robertson, Smeets, Lubinski & Benbow, 2010).

Befunde einer groß angelegten US-amerikanischen Studie zu einem dem Frühstudium ähnlichen Förderprogramm zeigen wiederum eine moderate prognostische Validität der Ergebnisse eines im Auswahlverfahren durchgeführten kognitiven Fähigkeitstests (Preliminary Scholastic Aptitude Test, PSAT). In sämtlichen untersuchten Studienfächern war die inkrementelle Validität des PSAT in Kombination mit den Zeugnisnoten zur Erklärung der Varianz der 19 Monate später im Studienfach erzielten Leistungen beträchtlich und lag zwischen $R^2 = .19$ (Studienfach Analysis) und $R^2 = .43$ (Studienfach Englische Literatur) (Ewing, Camara & Willsap, 2006).

Auch die Ergebnisse der umfangreichen Studien von VanTassel-Baska und Koautoren ($N = 30\,526$ bzw. $N = 14\,290$) zum Vergleich intelligenz- vs. leistungsorientierter Auswahlstrategien für diverse Begabtenförderprogramme in Südkalifornien zeigen, dass die ausgewählten Gruppen später bessere Leistungen erzielten, wenn für die Auswahlentscheidungen Intelligenzwerte berücksichtigt und strengere Intelligenz- bzw. Leistungsschwellen gesetzt wurden (VanTassel-Baska, Feng & de Brux, 2007; VanTassel-Baska, Feng & Evans, 2007).

Zusammenfassend ergibt sich damit bislang kein einheitliches Bild darüber, welche Prädiktoren für die Leistungsentwicklungen in Begabtenklassen im Auswahlverfahren zu operationalisieren sind. Vor allem die Tatsache, dass die Ergebnisse der vorliegenden deutschsprachigen Studien zu genau dieser Thematik hinsichtlich der prädiktiven Validität der Intelligenzwerte stark divergieren, wirft Probleme für diese praktisch relevante Fragestellung auf. In den genannten deutschsprachigen Untersuchungen wurden neben der Intelligenz auch Zensuren und Vorwissensmaße einbezogen. Unterschiede zeigen sich hinsichtlich der eingesetzten Intelligenztestverfahren: Während in den Studien von Lehmann und Jüling (1999; s. auch Jüling & Lehmann, 1996, 1999) das Leistungsprüfsystem von Horn (1983) zum Einsatz kam, setzte Hany (2002, 2004) einzelne Subtests des KFT (Heller & Perleth, 2000) ein. Darüber hinaus unterscheiden sich die Studien hinsichtlich der untersuchten Altersgruppe, und aufgrund vielfacher früherer Befunde sind geringere Zusammenhänge zwischen Intelligenz- und Leistungsdaten in höheren Jahrgangsstufen zu erwarten (z. B. Rost, 2009). Teilweise könnten die geringen Zusammenhänge auch auf eingeschränkte Varianzen der Prädiktoren und Kriterien zurückgeführt werden, da den Untersuchungen zu Begabtenklassen in der Regel keine repräsentativen Stichproben

zugrunde liegen. Weiterhin liegen den drei Studien vergleichsweise kleine Stichproben für längsschnittliche Analysen zugrunde (Hany, 2002⁵: $N \leq 90$; Hany, 2004: $N = 33$; Jüling & Lehmann, 1996: $N = 72$), sodass die Ergebnisse nicht als ausreichend repräsentativ anzusehen sind. Unter Einbeziehen des internationalen Forschungsstands lassen die oben referierten Befunde insgesamt erwarten, dass sowohl Zensuren als auch Intelligenzmaße geeignete Prädiktoren für die Leistungsentwicklung in gymnasialen Begabtenklassen darstellen und dass deren Kombination die Prognosegüte erhöht. Kritisch zu betrachten ist die Tatsache, dass in den bisherigen Studien zu Fördermodellen im deutschen Bildungssystem fast ausschließlich Zensuren als Leistungskriterien angewendet wurden, Befunde zur Entwicklung objektiv gemessener Leistungen liegen bislang unserer Kenntnis nach hierzulande nicht vor.

Auswahlverfahren für gymnasiale Begabtenklassen

Vorliegende Studie stellt einen Ausschnitt des Projekts⁶ zur Untersuchung des Lernens in der Sekundarstufe (PULSS) dar, in dessen Rahmen je vier Gymnasien mit Begabtenklassen in Bayern und Baden-Württemberg von Anfang der fünften bis Ende der siebten Jahrgangsstufe wissenschaftlich begleitet wurden (s. u.). In diesen Ländern wurden die Begabtenklassen in den vergangenen zehn Jahren auf acht (Bayern) bzw. 17 Standorte (Baden-Württemberg) ausgeweitet, die mehrheitlich mit dem Übertritt in die Sekundarstufe I beginnen. Entsprechend den zuvor erläuterten wissenschaftlichen Empfehlungen wird die Auswahl über einen mehrstufigen und multimethodalen Prozess vorgenommen, in dessen genauer Ausgestaltung deutliche Unterschiede zwischen den Schulen bestehen (vgl. Trottler, 2011). Einzig eine Intelligenztestung wird einheitlich an allen Standorten durchgeführt. Hierbei kommen allerdings unterschiedliche Testverfahren zum Einsatz und es greifen unterschiedlich strenge Schwellenwerte zwischen $IQ \geq 120$ und $IQ \geq 130$. Die anhand der Intelligenztestung vorselegierte Gruppe vorgeschlagener Schülerinnen und Schüler durchläuft in der Folge die weiteren Bausteine der Auswahlverfahren, das in je fünf der acht Gymnasien die Erfassung der Zeugnisnoten der vierten Jahrgangsstufe bzw. die Teilnahme

⁵ Aus der Studie wird der genaue Stichprobenumfang nicht klar ersichtlich. Hierzu lautet die Beschreibung wie folgt: „Im Rahmen des erteilten Forschungsauftrages konnten Daten von ca. 90 Bewerbern für das Schuljahr 2000/2001 gesammelt werden. Weitere Daten konnten von den 40 aufgenommenen Schülern erhoben werden, die zunächst bis zum Ende des Kalenderjahres 2001 untersucht wurden. Auf diese Weise war es möglich, zumindest für die aufgenommenen Schüler festzustellen...“ (Hany, 2002, S. 263).

⁶ Die Studie wurde mit Unterstützung der Kultusministerien Bayern und Baden-Württemberg sowie der Karg-Stiftung und in Kooperation der Universitäten Würzburg, Erlangen-Nürnberg und Trier durchgeführt.

an einem Probeunterricht zur Beobachtung des Arbeits- und Sozialverhaltens der Kinder beinhaltet (vgl. Trottler, 2011). Dieser Probeunterricht ist mit hohem zeitlichem Aufwand verbunden, da die vorselektierte Gruppe vorgestellter Kinder in Kleingruppen aufgeteilt über mehrere Stunden beobachtet wird. Einige Schulen binden in ihre Auswahlverfahren darüber hinaus noch weitere Elemente mit ein, wie z. B. standardisierte Fragebögen zum akademischen Selbstkonzept oder Beurteilungen der Grundschullehrkraft.

In Passung mit den eingangs vorgestellten wissenschaftlichen Empfehlungen wird die Auswahl der Schülerinnen und Schüler für diese Begabtenklassen also in einem sequenziellen Prozess vorgenommen, in dem – wie u.a. von Feldhusen und Jarwan (2002) sowie Rost (1991) empfohlen - die Ergebnisse der standardisierten Intelligenzdiagnostik als erste Schwelle greifen und die weitere Auswahl durch Einbeziehen weiterer Informationen, Messverfahren und Quellen vorgenommen wird. Dieses Vorgehen spiegelt die bereits erwähnte Heterogenität der Gestaltung von Auswahlverfahren wider. Rekurrierend auf die erläuterte Befundlage zur Validität von Auswahlverfahren sollten Intelligenz- mit Leistungsmaßen kombiniert werden, um valide Leistungsprognosen zu erlauben. Die Zeugnisnoten werden jedoch bislang nicht einheitlich in die Auswahlverfahren für die gymnasialen Begabtenklassen einbezogen. Hinsichtlich des Probeunterrichts ist kritisch festzustellen, dass dessen Durchführung für die Schulen mit hohem zeitlichem Aufwand verbunden ist und bislang der Nutzen für die Aufnahmeentscheidungen noch unerforscht zu sein scheint.

Fragestellungen

Für die Überprüfung der Validität der in der Praxis zur Anwendung kommenden Auswahlverfahren für die gymnasialen Begabtenklassen ist daher zu überprüfen,

1. welchen Beitrag die Prädiktoren Intelligenz und Zeugnisnoten der vierten Jahrgangsstufe zur Erklärung von Leistungsunterschieden in der siebten Jahrgangsstufe leisten, und
2. ob die Validität der Leistungsprognose anhand von Zeugnisnoten und Intelligenzwerten durch die im Zuge des Probeunterrichts erhobenen Daten erhöht werden kann.

Vornehmliches Ziel dieser Untersuchung stellt die Validierung der Auswahlverfahren für die Begabtenklassen dar, wie sie in der Praxis der meisten beteiligten Gymnasien zur Anwendung kommen. Es sollen solche Prädiktorvariablen und deren Kombinationen

identifiziert werden, die die Leistungsentwicklungen der Schülerinnen und Schüler möglichst valide prognostizieren, und darüber hinaus möglichst ökonomisch operationalisierbar sind. Letztgenannter Aspekt ist vor allem für diejenigen Gymnasien von Bedeutung, bei denen die Nachfrage die verfügbare Kapazität um ein Vielfaches übersteigt.

Methode

Die vorliegenden Ergebnisse beinhalten die Validierung der Auswahlverfahren der an PULSS teilnehmenden Gymnasien, wobei es sich um die Eingangsklassen der Schuljahre 2008/2009 und 2009/2010 handelt. Ursprünglich strebten wir an, alle für die Begabtenklassen vorgestellten Kinder in die Untersuchung einzubeziehen. Allerdings wurde die Aufnahme der Daten der abgelehnten Kinder sowie deren längerfristige wissenschaftliche Begleitung von den beteiligten Ministerien nicht genehmigt. Die hier vorgestellten Ergebnisse sind daher nur für die in die Begabtenklassen aufgenommenen Schülerinnen und Schüler gültig. Die Gestaltung und Durchführung der Auswahlverfahren oblag den einzelnen Schulen, weshalb sie nicht identisch aufgebaut waren. Die Durchführung eines Intelligenztests wurde an allen teilnehmenden Schulen vorausgesetzt.

Stichprobe

In die Analysen einbezogen wurden diejenigen fünf Gymnasien der PULSS-Studie, in denen im Auswahlverfahren neben der Intelligenzmessung auch die aktuellen Zeugnisnoten (uneinheitlich entweder Zwischen- oder Übertrittszeugnis) der Fächer Deutsch, Mathematik und Heimat- und Sachunterricht (bayerische Grundschulen) bzw. Mensch, Natur und Kultur (baden-württembergische Grundschulen) erhoben und außerdem ein Probeunterricht durchgeführt wurde. Die Auswahlverfahren wurden im Frühjahr 2008 und 2009 vor dem Übertritt in die fünfte Jahrgangsstufe von $N = 460$ Schülerinnen und Schülern absolviert (316 männlich, 144 weiblich), wovon $n = 241$ für die Begabtenklassen zugelassen wurden. Da die verschiedenen eingesetzten Intelligenztests (z. B. KFT, PSB, HAWIK-IV) zu unterschiedlichen Ergebnissen führen können (z. B. Schlagheck & Petermann, 2006), wurden zur Gewährleistung der Vergleichbarkeit in diese Auswertungen ausschließlich die Daten derjenigen Schülerinnen und Schüler einbezogen, die während der Auswahlverfahren den gleichen Intelligenztest bearbeitet hatten. Aufgrund der größten verfügbaren Stichprobe entschieden sich die Autorinnen für den Kognitiven Fähigkeitstest (KFT 4-12, Heller & Perleth, 2000). Die Stichprobe wurde zusätzlich um einzelne

Schülerinnen und Schüler verringert, die nicht an allen für die Analysen relevanten Erhebungen (z. B. Probeunterricht, Leistungstests in Jahrgangsstufe 7) teilgenommen hatten. Für die vorliegende Studie resultiert somit eine Gesamtstichprobe aus $N = 136$ Schülerinnen und Schülern (85 männlich, 51 weiblich) aus 9 Begabtenklassen der 5 bayerischen und baden-württembergischen Gymnasien, die zu Beginn des fünften Schuljahres zwischen 8.94 und 11.17 ($M = 10.33$, $SD = 0.49$) Jahre alt waren.

Prädiktoren

Für die Erfassung der *Intelligenz* wurden die IQ-Werte des KFT für die vierte Klasse (Heller & Perleth, 2000) herangezogen, die von den Schulen uneinheitlich mittels der Lang- bzw. der Kurzform erfasst wurden. Laut Manual korrelieren beide Versionen sehr hoch miteinander ($r = .98$), sodass die Daten für die vorliegende Studie zusammengefasst werden konnten.

Die *Zensuren* der Schülerinnen und Schüler wurden anhand der durchschnittlichen Zeugnisnoten (Mathematik, Deutsch, Heimat- und Sachunterricht bzw. Mensch, Natur und Kultur) zum Halbjahr der vierten Jahrgangsstufe als Prädiktoren einbezogen.

An jedem teilnehmenden Gymnasium wurde eine Lehrkraft als Kontaktperson benannt, die u. a. das Leistungs-, Arbeits- und Sozialverhalten der vorgestellten Schülerinnen und Schüler während des *Probeunterrichts* auf einer siebenstufigen Skala (von 1 = „stark unterdurchschnittlich“ bis 7 = „stark überdurchschnittlich“) einschätzte.

Kriterien

Als Leistungskriterien wurden sowohl standardisierte Testverfahren als auch Zensuren einbezogen, die jeweils Mitte der siebten Jahrgangsstufe erhoben wurden. Die Feststellung der *mathematischen Fähigkeiten* erfolgte mithilfe eines für PULSS entwickelten Verfahrens (Weiß & Schneider, 2009), da bisher kein für die Stichprobe geeignetes Messinstrument vorlag. Der Mathematiktest wurde auf Basis bereits bestehender lehrplanabhängiger sowie selbst entworfener Aufgaben aufgebaut, welche aus der Schnittmenge der bayerischen und baden-württembergischen Lehrpläne resultierten. Die meisten Aufgaben entstammen den gängigen Lehrbüchern „Delta 7“ (Schätz & Eisentraut, 2007) und „Elemente der Mathematik 7“ (Griesel, Postel & Suhr, 2005) und bezogen sich auf folgende Inhaltsbereiche: Achsen- und Punktsymmetrie, Konstruktion von Drei- und Vierecken, Winkelbetrachtungen, besondere Dreiecke, Kongruenzsätze, Termbegriff und Rechnen mit Termen, Vertiefen des Prozentrechnens, statistische Erhebung und arithmetisches Mittel. Das Instrument enthält 56 Items und erwies sich als intern konsistent

(Cronbachs $\alpha = .84$). Im Fach *Deutsch* kam als *Leseverständnistest* der Subtest zum Textverständnis aus der Lesetestbatterie für die Klassenstufen 8-9 (LESEN 8-9; Bäuerlein, Lenhard & Schneider, 2012) zum Einsatz. Dieser besteht aus 38 Items. Da sich in Vortests mit Jahrgangsstufen entsprechenden Instrumenten Deckeneffekte ergeben hatten, musste auf einen Test zurückgegriffen werden, der für höhere Jahrgangsstufen konzipiert war. Die interne Konsistenz ist ebenfalls als gut einzustufen (Cronbachs $\alpha = .84$). Zudem wurde ein *Englisch Leistungstest* für PULSS entwickelt (Harder & Ziegler, 2009), da auch in diesem Fach für die sehr leistungsfähige Stichprobe kein passendes Instrument vorlag. Der neu entworfene Test basierte ebenfalls auf den Lehrplänen der beiden Bundesländer Bayern und Baden-Württemberg sowie den Lehrbüchern für die siebte Jahrgangsstufe. Inhaltlich bezog er sich in 190 Items auf die Bereiche Wortproduktion, Vokabelwissen, Orthografie, Textverständnis, Grammatik und Aussprachewissen. Die interne Konsistenz fiel hierbei sehr überzeugend aus (Cronbachs $\alpha = .96$). Die Ergebnisse dieser drei standardisierten Testverfahren wurden sowohl einzeln als auch aggregiert als Leistungskriterien in die statistischen Analysen einbezogen. Dafür wurden sie zunächst z-transformiert, um die Vergleichbarkeit zu erhalten und einen Mittelwert aus den drei Leistungstests bilden zu können.

Die Zensuren der Schülerinnen und Schüler wurden in den mit den Testverfahren verwandten Hauptfächern Mathematik, Deutsch und Englisch ebenfalls zum Halbjahr der siebten Jahrgangsstufe als Leistungskriterien herangezogen und klassenweise z-standardisiert in die Auswertungen aufgenommen. Die Standardisierung wurde vorgenommen, um die durch Leistungsunterschiede zwischen den Klassen sowie Urteilstendenzen der Lehrkräfte verursachten Varianzanteile in den Prädiktorvariablen gering zu halten.

Statistische Auswertung

Die Fragestellungen wurden mittels Regressionsanalysen untersucht. Die Daten der Auswahlverfahren (Intelligenztestergebnis, Notendurchschnitt, Einschätzungen aus dem Probeunterricht) wurden hierfür als Prädiktoren und die akademischen Leistungen (Zensuren sowie standardisierte Testergebnisse) der Schülerinnen und Schüler Mitte der siebten Jahrgangsstufe als Kriterien einbezogen. Zur Überprüfung der beiden Fragestellungen wurden blockweise Regressionsanalysen mit dem Intelligenztestergebnis (1. Block), der Durchschnittsnote in der vierten Jahrgangsstufe (2. Block) und den Einschätzungen aus dem Probeunterricht (3. Block, innerhalb des Blocks: Einschluss) als Prädiktoren eingesetzt. Der

korrigierte Determinationskoeffizient R^2_{kor} wird als Effektstärkemaß berichtet. Fälle mit fehlenden Werten in den einbezogenen Variablen wurden aus den Analysen ausgeschlossen (listwise). Aufgrund dieser Regelung variiert die Stichprobengröße zwischen den Auswertungen.

Ergebnisse

Die $n = 136$ in die Studie aufgenommenen Schülerinnen und Schüler erzielten im Mittel einen IQ-Wert von $M = 129.12$ ($SD = 7.16$) im KFT. Sie wiesen im Zeugnis Mitte der vierten Jahrgangsstufe einen Notendurchschnitt von $M = 1.44$ ($SD = 0.39$) auf. Auf der Skala von 1-7, mit welcher die Einschätzungen für den Probeunterricht abgegeben werden konnten, erreichten die aufgenommenen Kinder für das Leistungs-, das Arbeits- und das Sozialverhalten recht hohe Mittelwerte, die zwischen $M = 5.26$ ($SD = 1.25$) und $M = 5.46$ ($SD = 0.97$) variierten. Zum Zeitpunkt der Messung der Kriterien (Zensuren Mitte der 7. Jahrgangsstufe, Ergebnisse der Leistungstests) waren die Stichprobengrößen ($n = 69$ im Englischleistungstest bis $n = 132$ bei den Noten⁷) geringer, da nicht von allen teilnehmenden Schülerinnen und Schülern ein vollständiger Datensatz vorlag (z. B. wegen Krankheit während eines Erhebungszeitpunktes).

Ergebnisse zu Fragestellung 1

Die Ergebnisse zu Fragestellung 1 geben Aufschluss zur prädiktiven Validität von Intelligenz und Zeugnisnoten für die Leistungsprognose und sind in den Tabellen 1 und 2 dargestellt.

(Ungefähr hier Tabelle 1 einfügen)

Prädiktion der Zensuren Mitte der 7. Jahrgangsstufe

Es zeigte sich, dass das KFT-Ergebnis aus den Auswahlverfahren allein betrachtet für die Aufklärung der Varianzen der Englisch- und Deutschzensur der siebten Jahrgangsstufe keinen signifikanten Beitrag leistete, jedoch für die Aufklärung der Varianz der Mathematikzensur bzw. des Zensurendurchschnitts aus Mathematik, Deutsch und Englisch mit geringen bis moderaten Effektstärken (vgl. Tabelle 1). Durch die Hinzunahme des Notendurchschnitts der vierten Jahrgangsstufe als weiteren Prädiktor konnte die

⁷ Die auffällig kleine Stichprobe im Englischleistungstest resultiert daraus, dass während der Erhebungen lediglich die erste Fremdsprache mittels Testverfahren erhoben wurde, z. T. auch Latein, was nicht in diese Untersuchung mit einging.

Varianzaufklärung der Zensuren in Mathematik, Deutsch und Englisch sowie deren Durchschnitt Mitte der siebten Jahrgangsstufe signifikant erhöht werden. Dabei erwies sich der Notendurchschnitt der vierten Jahrgangsstufe für drei der vier verwendeten Kriterien als bedeutsamer für die Leistungsprognose als das KFT-Ergebnis.

(Ungefähr hier Tabelle 2 einfügen)

Prädiktion der Testleistungen Mitte der 7. Jahrgangsstufe

Die Prognoseleistung des KFT-Ergebnisses liegt für das Abschneiden im standardisierten Leseverständnistest der siebten Jahrgangsstufe im moderaten bis starken Bereich (vgl. Tabelle 2), der Durchschnitt aus dem Leseverständnis-, dem Englisch- und dem Mathematiktest wird vom KFT-Ergebnis gering bis moderat aufgeklärt. Die Varianzen im Englisch- und Mathematiktest der siebten Jahrgangsstufe ließen sich durch die KFT-Werte jedoch nicht aufklären. Die Hinzunahme des Notendurchschnitts der vierten Jahrgangsstufe als weiteren Prädiktor verbesserte die Vorhersage der standardisierten Leistungstestergebnisse der siebten Jahrgangsstufe bedeutsam im Leseverständnis, in Mathematik sowie dem Mittelwert der Leistungstests. Dabei zeigte sich, dass der Notendurchschnitt der vierten Jahrgangsstufe zur Varianzaufklärung des Mathematiktestergebnisses sowie des Durchschnitts der Testergebnisse bedeutsamer als die KFT-Werte war, wohingegen sich der KFT zur Varianzaufklärung des Leseverständnistest-Ergebnisses als bedeutender erwies. Die beiden Prädiktoren KFT-Ergebnis und Notendurchschnitt der vierten Jahrgangsstufe konnten die Varianz der Englischtestleistung der siebten Jahrgangsstufe allerdings nicht aufklären.

Insgesamt betrachtet zeigen die Ergebnisse zu Fragestellung 1, dass die Kombination aus dem KFT-Ergebnis und dem Notendurchschnitt der vierten Jahrgangsstufe die standardisierten Leistungskriterien der siebten Jahrgangsstufe in etwas höheren Anteilen vorhersagen konnten als die Zensuren. Die neben dem KFT zusätzliche Berücksichtigung des Notendurchschnitts der vierten Jahrgangsstufe erlaubte zudem in den meisten Fällen eine Verbesserung der Varianzaufklärung der Leistungskriterien.

Ergebnisse zu Fragestellung 2

Wie aus Tabelle 1 und 2 deutlich wird, kann der Anteil aufgeklärter Varianz der untersuchten Leistungskriterien durch eine zusätzliche Berücksichtigung der Einschätzungen aus dem Probeunterricht weder für die Zensuren noch für die Leistungstestergebnisse der siebten Jahrgangsstufe erhöht werden. Die Modelle mit den 5 Prädiktoren erlangen zwar mit

einer Ausnahme (Prognose des Englischleistungstests) statistische Bedeutsamkeit, jedoch tragen nur der Notendurchschnitt der vierten Jahrgangsstufe sowie das KFT-Ergebnis signifikant dazu bei, nicht aber die Einschätzungen aus dem Probeunterricht.

Diskussion

Die hier vorgestellten Ergebnisse zur prognostischen Validität der Auswahlverfahren für gymnasiale Begabtenklassen verdeutlichen eine Abhängigkeit vom gewählten Leistungskriterium. Insgesamt konnten die gewählten Prädiktoren der Auswahlverfahren die Leistungsvarianz der siebten Jahrgangsstufe in höheren Anteilen aufklären, wenn standardisierte Leistungstests anstelle von Zeugnisnoten als Kriterien verwendet wurden. In Bezug auf die Fragestellungen bleibt zu resümieren, dass sich sowohl die im Auswahlverfahren erhobenen Intelligenzwerte als auch die Zeugnisnoten der vierten Jahrgangsstufe als bedeutsam zur Prognose späterer Leistungen erwiesen haben, wobei den Zeugnisnoten insgesamt betrachtet eine höhere prädiktive Rolle zur Vorhersage der Leistungskriterien zukommt. Die ergänzende Berücksichtigung der Daten aus dem Probeunterricht führte hingegen nicht zu einem höheren Anteil aufgeklärter Varianz der Leistungskriterien in der siebten Jahrgangsstufe.

Um die Schulleistungen der Schülerinnen und Schüler möglichst valide und ökonomisch prognostizieren zu können, sollten im Auswahlverfahren daher die Intelligenzwerte *und* die Zeugnisnoten der vierten Jahrgangsstufe einbezogen werden. Insofern stimmen die hier vorgestellten Ergebnisse mit den eingangs referierten Studien (Lehmann & Jüling, 1999; Stumpf & Schneider, 2013) überein. Wie zuvor dargestellt, beziehen bislang nicht alle Gymnasien neben den Intelligenztestergebnissen auch die Zensuren der vierten Jahrgangsstufe in ihre Auswahlentscheidung für die Begabtenklassen ein, was auf Basis unserer Ergebnisse dringend anzuraten wäre. Für die zeitaufwendige Durchführung des Probeunterrichts konnte hingegen kein zusätzlicher Nutzen bezüglich der Leistungsprognose bestätigt werden. Wie die insgesamt mangelnde Eignung der Daten aus dem Probeunterricht als Entscheidungsgrundlage verdeutlicht, gelingt es mittels dieser Methode offenbar nicht hinreichend, Aspekte im Arbeits-, Leistungs- und Sozialverhalten der Kinder zu erkennen, die über die Intelligenzwerte und Zensuren hinausgehenden Informationsgehalt für die spätere Leistungsentwicklung repräsentieren. Auch ist nicht ausgeschlossen, dass die gewählten Methoden zur Dokumentation des im Probeunterricht gezeigten Verhaltens als grobe Ratings inhaltlich zu wenig spezifiziert waren, um bessere

Prognosen zu erlauben. Hierbei stand die Ökonomie der Methode im Vordergrund, da die Lehrkräfte diese Angaben zu jedem vorgestellten Kind vornehmen mussten.

Erneut soll daran erinnert werden, dass für die hier berichteten Ergebnisse ausschließlich die in die Begabtenklassen aufgenommenen Schülerinnen und Schüler berücksichtigt werden konnten. Die Übertragbarkeit der Befunde auf die Entscheidungen gegen die Aufnahme ist daher nicht gegeben.

Die vorgestellten Ergebnisse weisen indirekt auf einen Aspekt der gymnasialen Begabtenklassen hin, der bereits früher kritisch festgestellt wurde. Denn die Intelligenzwerte der Schülerinnen und Schüler verdeutlichen, dass die für die Aufnahme in die Begabtenklassen gesetzten Intelligenzschwellen teilweise deutlich unterschritten werden: Während die Schulen Intelligenzschwellen von $IQ \geq 120$ kommunizieren, beträgt der niedrigste Intelligenzwert der untersuchten Stichprobe tatsächlich nur $IQ = 102$. Dies kann darauf zurückgeführt werden, dass den Schulen zu einigen Kindern mehrere Intelligenztestergebnisse vorliegen und jeweils das bessere Ergebnis als für die Auswahlentscheidung relevant betrachtet wird (vgl. Stumpf & Schneider, 2008), wohingegen in vorliegender Studie einheitlich die Ergebnisse des KFT einbezogen wurden. Wie die Darstellung der KFT-Ergebnisse zeigt, weist die hier untersuchte Stichprobe trotz dieser Abweichung von den gesetzten Intelligenzschwellen keine größere, sondern im Gegenteil eine geringere Streuung der Intelligenzwerte auf als üblich ($SD = 7.16$). Diese Tatsache könnte teilweise erklären, warum die KFT-Ergebnisse nur vergleichsweise geringe Anteile der Kriteriumsvarianzen aufzuklären erlauben.

Im Rahmen der vorliegenden Studie wurde das Ziel verfolgt, die in der Praxis etablierten Auswahlverfahren für gymnasiale Begabtenklassen zu validieren. Daher waren die Analysen auf diejenigen Prädiktoren eingeschränkt, die von den meisten Gymnasien erhoben werden. Aus Gründen der Vergleichbarkeit wurden zudem nur diejenigen Schülerinnen und Schüler in die Analysen einbezogen, deren Intelligenz im Zuge des Auswahlverfahrens anhand des KFT gemessen worden war und die an sämtlichen relevanten Erhebungen (z. B. Probeunterricht, Leistungsmessungen in der vierten Jahrgangsstufe) teilgenommen hatten. Inwiefern die Ergebnisse auf andere Intelligenztests übertragbar sind, kann momentan noch nicht beantwortet werden. Erste Analysen lassen beträchtliche Abweichungen hinsichtlich der prognostischen Validität der verschiedenen Intelligenztestverfahren erwarten (Stumpf & Trotter, 2012). Die eingangs referierten

Studien von Lehmann und Jüling (1999) bzw. Jüling und Lehmann (1999) weisen zudem darauf hin, dass durch die zusätzliche Berücksichtigung standardisierter Leistungstests (hier im Fach Mathematik) im Auswahlverfahren der Anteil aufgeklärter Varianz des später erzielten Notendurchschnitts erhöht werden kann. Darüber hinaus lassen die korrelativen Befunde von Hany (2002, 2004) eine Verbesserung der prognostischen Validität der Auswahlverfahren durch die Berücksichtigung nichtkognitiver Merkmale erwarten. Bislang werden allerdings weder standardisierte Leistungstests noch Daten zu nichtkognitiven Merkmalen der Schülerinnen und Schüler systematisch für die Auswahlentscheidungen an den Gymnasien berücksichtigt. Zukünftige Studien sollten klären, inwiefern das Einbeziehen solcher Daten zu einer weiteren Erhöhung der Varianzaufklärung beitragen und damit gewinnbringend für die Optimierung der Auswahlverfahren operationalisiert werden kann.

Langfristig wäre es wünschenswert, die Auswahlentscheidungen evidenzbasiert treffen zu können, indem verschiedene Messverfahren anhand ihrer prädiktiven Validität für die Erfolgskriterien gewichtet und kombiniert berücksichtigt werden (vgl. Feldhusen & Jarwan, 2002). Eine solche auf Basis empirischer Befunde basierende Vorgehensweise würde es erlauben, die Treffsicherheit der Auswahlentscheidungen zu erhöhen und das Potenzial der Begabtenklassen noch besser auszuschöpfen. Bislang sind die vorliegenden Befunde unserer Überzeugung nach noch zu eingeschränkt und heterogen, um konkrete Schlussfolgerungen zur Gewichtung von Einzelelementen und deren Kombination für die Auswahlentscheidungen zuzulassen. Ungeachtet dessen scheint die Berücksichtigung der Zeugnisnoten der vierten Jahrgangsstufe zusätzlich zu den Intelligenztestwerten überzeugend genug indiziert zu sein. Zukünftig sollten zudem auch die abgelehnten Kinder in den Fokus wissenschaftlicher Untersuchungen gerückt werden, da für diese Gruppe bislang unserer Kenntnis nach keinerlei Aussagen über deren weitere schulische Entwicklung getroffen werden kann.

Die weitere Validierung der Auswahlverfahren für gymnasiale Begabtenklassen erscheint u.a. wichtig und lohnenswert, da bundesweit jährlich etwa 6300 Schülerinnen und Schüler diese Auswahlverfahren durchlaufen. Die Konsequenzen der Entscheidung für wie gegen die Aufnahme sind – anders als bei niederschwelligen Begabtenfördermaßnahmen wie beispielsweise dem Frühstudium – weitreichend, weshalb eine wissenschaftlich anspruchsvolle Fundierung dieser Entscheidungen dringend notwendig ist.

Literatur

- Bäuerlein, K., Lenhard, W. & Schneider, W. (2012). *Lesetestbatterie für die Klassenstufen 8-9 (LESEN 8-9)*. Göttingen: Hogrefe.
- Ewing, M., Camara, W. J. & Willsap, R. E. (2006). *The relationship between PSAT/NMSQT scores and AP examination grades: A follow-up study (College Board Research Report No. 2006-1)*. New York City: The College Board.
- Feldhusen, J. F. & Jarwan, F. A. (2002). Identification of gifted and talented youth for educational programs. In K. A. Heller, F. J. Mönks, R. J. Sternberg & R. F. Subotnik (Eds.), *International handbook of giftedness and talent* (2nd ed., pp. 271-282). Oxford: Elsevier.
- Ferriman-Robertson, K., Smeets, S., Lubinski, D. & Benbow, C. P. (2010). Beyond the threshold hypothesis: Even among the gifted and top math/science graduate students, cognitive abilities, vocational interests, and lifestyle preferences matter for career choice, performance, and persistence. *Current Directions in Psychological Science*, 19, 346-351.
- Griesel, H., Postel, H., & Suhr, F. (Hrsg.). (2005). *Elemente der Mathematik 7: Baden-Württemberg Band 3*. Braunschweig: Schroedel.
- Hany E. A. (2002). Mathematisch-naturwissenschaftliche Spezialklassen: Wie findet man geeignete Schüler? In H. Wagner (Hrsg.), *Begabtenförderung und Lehrerbildung* (S. 261-272). Bad Honnef: K. H. Bock.
- Hany, E. A. (2004). Prognostische Validität von Aufnahmeverfahren in Spezialschulen. *Psychologie in Erziehung und Unterricht*, 51, 24-39.
- Harder, B. & Ziegler, A. (2009). *Englisch Leistungstest für 5. -7. Jahrgangsstufe*. Unveröffentlichte Testbatterie der Universität Erlangen-Nürnberg.
- Hartas, D., Lindsay, G. & Muijs, D. (2008). Identifying and selecting able students for the NAGTY Summer School: Emerging issues and future considerations. *High Ability Studies*, 19(1), 5-18.
- Heller, K.A. (2000). (Hrsg.). *Begabungsdagnostik in der Schul- und Erziehungsberatung* (2. Aufl.). Göttingen: Verlag Hans Huber.
- Heller, K. A. (2001). *Hochbegabungsd Diagnose (Identifikation)*. In K. A. Heller (Hrsg.), *Begabungsd diagnostik* (2., vollst. überarb. Aufl., S. 241-257). Bern: Hans Huber.
- Heller, K. A., Gaedike, A. K. & Weinländer, H. (1985). *Kognitiver Fähigkeitstest für 4. bis 13. Klassen (KFT 4-13+)*. Weinheim: Beltz.

- Heller, K. A. & Perleth, C. (2000). *Kognitiver Fähigkeitstest für 4. bis 12. Klassen, Revision*. Weinheim: Beltz.
- Heller, K. A. & Perleth, C. (2007). Talentförderung und Hochbegabtenberatung in Deutschland. In K. A. Heller & A. Ziegler (Hrsg.), *Begabt sein in Deutschland* (S. 139-170). Berlin: LIT.
- Horn, W. (1983). *Leistungsprüfsystem (LPS)*. Göttingen: Hogrefe.
- Jüling, I. & Lehmann, W. (1996). Zur Auswahl von Schülern für ein Gymnasium mit mathematisch-naturwissenschaftlich-technischem Profil. In E. Witruk & G. Friedrich. (Hrsg.), *Pädagogische Psychologie im Streit um ein neues Selbstverständnis* (S. 460-466). Landau: Verlag Empirische Pädagogik.
- Jüling, I. & Lehmann, W. (1999). Schüler an einem mathematisch-naturwissenschaftlich orientierten Gymnasium: Ihre Auswahl und Entwicklung. In C. Enders, C. Hanckel & S. Möley (Hrsg.), *Lebensraum - Lebenstraum - Lebenstrauma Schule* (S. 340-346). Bonn: Deutscher Psychologen Verlag.
- Kaiser, A. (1997). *Entwicklung und Erprobung von Modellen der Begabtenförderung am Gymnasium mit Verkürzung der Schulzeit. Abschlussbericht*. Mainz: v. Hase & Koehler Verlag.
- Lehmann, W. & Jüling, I. (1999). Zur Validierung eines Aufnahmeverfahrens für Schüler eines mathematisch-naturwissenschaftlich-technischen Gymnasiums. *Psychologie in Erziehung und Unterricht, 1*, 41-54.
- Middendorff, E., Isserstedt, W. & Kandulla, M. (2009). *Das soziale Profil in der Begabtenförderung. Ergebnisse einer Online-Befragung unter allen Geförderten der elf Begabtenförderungswerke im Oktober 2008*. Zugriff am 25.01.2013 http://www.bmbf.de/pubRD/soziale_profil_begabtenfoerderung.pdf
- Mönks, F. J., Peters, W. A. M. & Pflüger, R. (2003). *Schulische Begabtenförderung in Europa. Bestandsaufnahme und Ausblick. Band II Thematische Darstellung*. Nijmegen: Zentrum für Begabungsforschung, Universität Nijmegen.
- Perleth, C. & Collier, A. (2009). Zur Validität des Auswahlverfahrens. In K. A. Heller (Hrsg.), *Das Hector-Seminar: Ein wissenschaftlich evaluiertes Modell der Begabtenförderung im MINT-Bereich* (S. 123-132). Berlin: LIT.
- Renzulli, J. S. (1993). Ein praktisches System zur Identifizierung hochbegabter und talentierter Schüler. *Psychologie in Erziehung und Unterricht, 40*, 217-224.

- Rost, D. H. (1991). Identifizierung von „Hochbegabung“. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 23, 197-231.
- Rost, D. H. (2000). *Hochbegabte und hochleistende Jugendliche*. Münster: Waxmann.
- Rost, D. H. (2009). *Intelligenz. Mythen und Fakten*. Weinheim: Beltz.
- Schätz, U. & Eisentraut, F. (2007). *Delta 7 – Mathematik 7 für Gymnasien* (2. Aufl.). Bamberg: C.C. Buchner.
- Schlagheck, W. & Petermann, F. (2006). Hochbegabtendiagnostik mit dem HAWIK-III und AID 2. *Kindheit und Entwicklung*, 2, 93-99.
- Schneider, W. (2008). Expertiseerwerb. In W. Schneider & M. Hasselhorn (Hrsg.), *Handbuch Pädagogische Psychologie* (S. 136-144). Göttingen: Hogrefe.
- Stumpf, E. (2011). *Begabtenförderung für Gymnasiasten – Längsschnittstudien zu homogenen Begabtenklassen und Frühstudium*. Münster: LIT.
- Stumpf, E. (2012a). Auswahlverfahren in der Förderung begabter Schülerinnen und Schüler. In Landesinstitut für Lehrerbildung und Schulentwicklung (Hrsg.), *15 Jahre BbB: Besondere Begabungen entdecken und fördern - Impulse für Unterricht und Schule. Dokumentation der Tagung am 25.11. und 26.11.2011 am Landesinstitut für Lehrerbildung und Schulentwicklung* (S. 17-24). Hamburg: ABA Druck/Copy GmbH.
- Stumpf, E. (2012b). *Förderung bei Hochbegabung*. Stuttgart: Kohlhammer.
- Stumpf, E. & Schneider, W. (2008). Schulleistungen in homogenen Begabtenklassen und gymnasialen Regelklassen der Sekundarstufe 1. *Diskurs Kindheits- und Jugendforschung*, 1, 67-81.
- Stumpf, E. & Schneider, W. (2009). Homogene Begabtenklassen am Gymnasium - Zielgruppe und Entwicklung der Schülerinnen und Schüler. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 41, 51-62.
- Stumpf, E. & Schneider, W. (2013). Diagnostische Herausforderungen bei der Auswahl von Frühstudierenden. *Diagnostica*, 59, 61-72.
- Stumpf, E. & Trotter, S. (2012, Oktober). *Ergebnisse zu den Auswahlverfahren*. Vortrag auf der Abschlusstagung des PULSS der Kultusministerien Bayern und Baden-Württemberg, Dillingen.
- Trotter, S. (2011). *Evaluation von Auswahlverfahren für Begabtenklassen an Gymnasien*. Unveröffentlichte Diplomarbeit, Julius-Maximilians-Universität Würzburg.
- VanTassel-Baska, J., Feng, A. X. & de Bruin, E. (2007). A study of identification and achievement profiles of performance task-identified gifted students over 6 years.

[Empirical Study Quantitative Study]. *Journal for the Education of the Gifted*, 31(1), 7-34.

VanTassel-Baska, J., Feng, A. X. & Evans, B. L. (2007). Patterns of identification and performance among gifted students identified through performance tasks: A three-year analysis. [Empirical Study Quantitative Study]. *Gifted Child Quarterly*, 51(3), 218-231.

Vock, M., Preckel, F. & Holling, H. (2007). *Förderung Hochbegabter in der Schule. Evaluationsbefunde und Wirksamkeit von Maßnahmen*. Göttingen: Hogrefe.

Weiß, C. & Schneider, W. (2009). *Mathematik Leistungstest für 5.-7. Jahrgangsstufe*. Unveröffentlichte Testbatterie der Universität Würzburg.

Tabelle 1
 Blockweise Regressionsanalysen (Kriterien: Zensuren)

Prädiktoren	Zensuren Mitte 7. Jahrgangsstufe ($n = 132$)											
	Durchschnitt			Deutsch			Englisch			Mathematik		
	R^2_{korr}	ΔR^2_{korr}	β	R^2_{korr}	ΔR^2_{korr}	β	R^2_{korr}	ΔR^2_{korr}	β	R^2_{korr}	ΔR^2_{korr}	β
Block 1: KFT	.04*	-	-.21*	-.01	-	-	.01	-	-	.10**	-	-.32**
Block 2:	.11**	.07**		.06**	.07**		.05*	.04**		.12**	.02*	
KFT			-.17*			0			-0.10			-.30**
Notendurchschnitt 4. Kl.			.29**			.27**			.23**			.19*
Block 3:	.11**	0		.07*	.01		.05*	0		.13**	.01	
KFT			-.17*			-.01			-.12			-.27**
Notendurchschnitt 4. Kl.			.26**			.23*			.20*			.18*
Leistungsverhalten PU			-.01			-.03			.04			-.03
Arbeitsverhalten PU			-.13			-.11			-.05			-.15
Sozialverhalten PU			-.03			-.08			-.12			.14

Anmerkung. PU = Probeunterricht. * $p < .05$. ** $p < .01$. Der Zensuredurchschnitt entspricht dem Mittelwert der klassenweise z-standardisierten Einzelnoten in Deutsch, Mathematik und Englisch. Da bei Noten ein hoher Wert eine schlechte Leistung bedeutet, resultieren negative Beta-Gewichte für den KFT und die Probeunterrichtsmaße.

Tabelle 2
 Blockweise Regressionsanalysen (Kriterien: Leistungstests)

Prädiktoren	Leistungstests Mitte 7. Klasse											
	Durchschnitt (n = 130)			Leseverständnis (n = 115)			Englisch (n = 69)			Mathematik (n = 109)		
	R^2_{korr}	ΔR^2_{korr}	β	R^2_{korr}	ΔR^2_{korr}	β	R^2_{korr}	ΔR^2_{korr}	β	R^2_{korr}	ΔR^2_{korr}	β
Block 1: KFT	.10**	-	.32**	.14**	-	.38**	-.01	-	-	0	-	-
Block 2: KFT & Notendurchschnitt	.21**	.11**		.21**	.07**		-.01	-		.14**	.14**	
KFT			.28**			.34**			-			.05
Zensuredurchschnitt 4. Kl.			-.35**			-.29**			-			-.39**
Block 3:	.20**	-.01		.19**	-.02		.03	.04		.13**	-.01	
KFT			.29**			.33**			-			.08
Zensuredurchschnitt 4. Kl.			-.31**			-.28**			-			-.39**
Leistungsverhalten PU			.04			.06			-			-.13
Arbeitsverhalten PU			.02			-.02			-			.11
Sozialverhalten PU			.08			.02			-			0

Anmerkung. PU = Probeunterricht * $p < .05$. ** $p < .01$. Der Durchschnittswert entspricht dem Mittelwert der z-standardisierten Leistungstests im Leseverständnis, Mathematik und Englisch. Da bei Noten ein hoher Wert eine schlechte Leistung bedeutet, resultieren negative Beta-Gewichte für den Zensuredurchschnitt der 4. Klasse.